Talk with Hybrid Agents: AI agents as believable characters for meaningful interactive storytelling

Submitted for the Degree of Doctor of Philosophy
Computer Science Research Centre
Royal College of Art

2024

Yuqian Sun
Supervisor: Dr. Ali Asadipour, Dr. Chang Hee Lee

Word count: 40688

© Yuqian Sun 2024.
This thesis is copyright material and no quotation from
it may be published without proper acknowledgement.

Abstract

With the development of artificial intelligence (AI), especially large language models (LLM), LLM-driven dialogue systems have begun to exhibit powerful intelligence and autonomy. They not only make humans feel as though there are social relationships when communicating with non-human entities, but also blur the boundary between the real world and virtual worlds, especially those related to video games.

This thesis presents an innovative exploration into the development and understanding of AI agents as hybrid agents, specifically focusing on LLM-driven conversational agents within interactive storytelling. It aims to bridge the gap between fictional narratives and social interactions in digital spaces, leveraging LLMs to enhance the believability and engagement of these agents.

To validate and explore this framework, a series of practice-led studies were conducted. These artifacts—including the **Wander** chatbot based on real-world map, AI-native game **1001 Nights**, the community-integrated chatbot **Catherine & David**, the creativity support tool **ORIBA**, the conceptual art installation AI Nüshu, and the final integrated system **Hyborg Agency** —were evaluated through a mixed-methods approach combining quantitative data with in-depth qualitative analysis.

A key finding of this thesis is a more nuanced conceptualization of believability. The studies reveal that for social agents, believability arises not merely from consistency or friendliness, but from perceived autonomy, which includes the capacity for disobedience and constructive conflict. Furthermore, the research shows that hybrid agents create meaningful interactions not by hiding the boundary between worlds, but by crossing it in an explicit and purposeful manner.

Overall, these studies provide a comprehensive understanding of how hybrid AI agents can blend fiction and reality, introducing new dimensions and metrics for evaluating AI agents. Specifically, this research proposes two dimensions for hybrid agents: content generation based on user input and their role as actors in real-life contexts. Additionally, it identifies three key metrics for assessing hybrid agents: social expansion, spatial expansion, and language expansion. These findings contribute to the understanding of hybrid AI agents in interactive storytelling and their broader impact across gaming and real-world interactions. The research outcomes have been recognized at top-tier conferences and exhibitions, making an impact in academia, art, and industry.

***Keywords**: Artificial intelligence, conversational agent, believable agent, narrative intelligence, interactive storytelling*

List of Outputs

Full Paper

1. **[Paper 6]** Yuqian Sun, Yuying Tang, Ze Gao, Zhijun Pan, Chuyan Xu, Yurou Chen, Kejiang Qian, Zhigang Wang, Tristan Braud, Chang Hee Lee, and Ali Asadipour. AI Nüshu: An Exploration of Language Emergence in Sisterhood-Through the Lens of Computational Linguistics. In SIGGRAPH Asia 2023 Art Paper, 2023. https://arxiv.org/abs/2310.11870.
2. **[paper5]** Yuqian Sun, Stefano Gualeni. Puppet Unstrung: Reframing Authorship in the Age of Autonomous AI Agents. Creativity in the Digital Age. 2023. Springer International Publishing.
3. **[Paper 4]** Yuqian Sun, Zhouyi Li, Ke Fang, Chang Hee Lee, and Ali Asadipour. Language as Reality: A Co-Creative Storytelling Game Experience in 1001 Nights using Generative AI. In AIIDE 2023, 2023. https://ojs.aaai.org/index.php/AIIDE/article/view/27539.
4. **[Paper 3]** Yuqian Sun, Hanyi Wang, Pok Man Chan, Morteza Tabibi, Yan Zhang, Huan Lu, Yuheng Chen, Chang Hee Lee, and Ali Asadipour. Fictional Worlds, Real Connections: Developing Community Storytelling Social Chatbots through LLMs. In IMET 2023, 2023. https://arxiv.org/abs/2309.11478.
5. **[Paper 2]** Yuqian Sun, Ying Xu, Chenhang Cheng, Yihua Li, Chang Hee Lee, and Ali Asadipour. Travel with wander in the metaverse: An ai chatbot to visit the future earth. In 2022 IEEE 24th International Workshop on Multimedia Signal Processing(MMSP), pages 1–6, 2022. https://ieeexplore.ieee.org/document/9950031
6. **[Paper 1]** Yuqian Sun, Xuran Ni, Haozhen Feng, Ray LC, Chang Hee Lee, and Ali Asadipour.Bringing stories to life in 1001 nights: A co-creative text adventure game using a story generation model. In Mirjam Vosmeer and Lissa Holloway-Attaway, editors, Interactive Storytelling, pages 651–672, Cham, 2022. Springer International Publishing. https://dl.acm.org/doi/abs/10.1007/978-3-031-22298-6_42

Peer-reviewed Demo and Artwork

1. **[Artwork 4]** Yuqian Sun, Yuying Tang, Ze Gao, Zhijun Pan, Zhigang Wang, Tristan Braud, Chang Hee Lee, and Ali Asadipour. AI Nüshu (Women's scripts) - An Exploration of Language Emergence in Sisterhood. In SIGGRAPH Asia 2023 Art Gallery (SA '23), Article 4, pages 1–2, New York, NY, USA, 2023. Association for Computing Machinery. DOI:10.1145/3610537.3622957.
2. **[Artwork 3]** Yuqian Sun, Chang Hee Lee, and Ali Asadipour. Hyborg Agency: Cultivating Conversational AI Creatures through Community Connections. In ACM SIGGRAPH 2023 Art Gallery (SIGGRAPH '23), Article 6, pages 1–2, New York, NY, USA, 2023. Association for Computing Machinery. DOI:10.1145/3588428.3593827.
3. **[Poster 1]** Yuqian Sun, Xingyu Li, Jun Peng, and Ze Gao. Inspire creativity with ORIBA: Transform Artists' Original Characters into Chatbots through Large Language Model. UbiComp/ISWC '23 Adjunct, pages 78–82, New York, NY, USA, 2023. Association for Computing Machinery. DOI:10.1145/3594739.3610695.
4. **[Demo 2]** Yuqian Sun, Ying Xu, Chenhang Cheng, Yihua Li, Chang Hee Lee, and Ali Asadipour. Explore the Future Earth with Wander 2.0: AI Chatbot Driven By Knowledge-base Story Generation and Text-to-image Model. In CHI Conference on Human Factors in Computing Systems Extended Abstracts, 2023. 10.1145/3544549.3583931
5. **[Artwork 2]** Yuqian Sun, Chenhang Cheng, Ying Xu, Yihua Li, Chang Hee Lee, and Ali Asadipour. Wander [001]. In SIGGRAPH Asia 2022 Art Gallery, Daegu, South Korea, December 2022. 10.1145/3550470.3558441

6. **[Artwork 1]** Yuqian Sun, Chenhang Cheng, Ying Xu, Yihua Li, Chang Hee Lee, and Ali Asadipour. Wander: An ai-driven chatbot to visit the future earth. In Proceedings of the 30th ACM International Conference on Multimedia, MM '22, page 7250–7251, New York, NY, USA, 2022. Association for Computing Machinery. 10.1145/3503161.3549971
7. **[Demo 1]** Sun Yuqian. Demo - 1001 Nights. In 17th International Conference on the Foundations of Digital Games, Athens, Greece, September 2022.
8. **[Poster]** Yuqian Sun, Chang Hee Lee, and Ali Asadipour. 1001 Nights – an open-domain narrative game using text generation model. In ISEA 2022, 2022. Available at https://isea2022.isea-international.org/event/poster-1001-nights-an-open-domain-narrative-game-using-text-generation-model.

Art Exhibitions

2023

- AI Nüshu, SIGGRAPH Asia Art Gallery, Sydney, Australia
- Explore the Future Earth with Wander 2.0, CHI Interactivity, Hamburg, Germany
- Modal A/W Exhibition: Web to Verse, Manchester, UK
- Hyborg Agency, SIGGRAPH Art Gallery, Los Angeles, US
- Home of Indies - Gamescom, Cologne, Germany
- NVIDIA AI Art Gallery, Online

2022

- X Virtual Incubator, X museum, Online
- Wander[001], SIGGRAPH Asia Art Gallery, Daegu, South Korea
- 1001 Nights, Foundation of Digital Games, Athens, Greece
- Wander: An AI-Driven chatbot to visit the future earth, ACM Multimedia Interactive Art, Lisbon, Portugal
- BBA Artist Prize 2022, BBA Gallery, Berlin

2021

- China Chengdu Biennale, Chengdu, China
- Book of Sand, Launch exhibition of Aiiiii Art Center, Shanghai, China

Awards

- Shortlist in CogX Awards
- Winner of 2023 Lumen Prize Student Award
- INDEX Biennial of Art and Technology - Art Grant (Top 4), Braga, Portugal
- 2022 Lumen Prize, Longlist in 3D/Interactive, Shortlist in Student Award
- BBA Artist Prize - 2nd Prize, BBA Gallery, Berlin

Talk and panels

- Hyperbobies Summit “AI as Hybrid Agent: Half-Real Storytelling” , invited by London College of Fashion (LCF)
- How Generative AI will Transform the Fashion Industry - Panelist, NVIDIA GTC 2023 (Online)
- Research Connections Speaker Series — Creative Storytelling with Generative AI, AutoDesk (Online)
- Sharing lecture, Creative Media Art, Hong Kong University of Science and Technology(HKUST) (Online)
- AI as Hybrid Agent: Half-real Storytelling, HYPERBODIES Summit, London College of Fashion, London, UK
- SIGGRAPH Asia 2022 Artist Talk: Speculating Symbiosis and Sustainability, Daegu, South Korea
- ACM Siggraph Digital Arts Community SPARKS: NFTs and the Non-fungible Experience (Online)
- 2022 Lumen Prize China Art and Technology Forum

Press

- Royal College of Art, PhD candidates win accolades for their work in AI, 2023
- Art-frame, The Internet gets real. “Web To Verse” at MODAL Gallery, Manchester
- NVIDIA, The Machine Muses: AI in Fashion
- Artnet China, Conversation with Yuqian Sun: When ”Raising Robots” Becomes Art (Chinese)
- Pro Helvetia Swiss Art Council, UN-Curating magazine - Curating the Digital Expanded
- Initium Media Hong Kong, AI Trainer Yuqian Sun: When AI Learns to Say NO and Other More Realistic Issues (Chinese)
- Royal College of Art, PhD candidate Yuqian Sun: RCA Computer Science Research Centre
- Royal College of Art, RCA Computer Science Research Centre interview
- e-flux, X Virtual Incubator by X museum, 2022

Contents

Abstract . . . 1
List of Contributions . . . 2
List of Figures . . . 14
List of Tables . . . 15

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Introduction to the thesis . . . 2
  - 1.2.1 Key terms and definitions . . . 4
  - 1.2.2 Research gaps . . . 8
  - 1.2.3 Definition of Hybrid Agent . . . 8
- 1.3 Research questions . . . 9
- 1.4 Summary of original contributions . . . 9
- 1.5 Summary of chapters . . . 10
- 1.6 Overview of contributions and impacts . . . 11
  - 1.6.1 Pilot Studies: Engaging Narratives with AI Agents . . . 11
  - 1.6.2 Catherine & David: Community-Based Storytelling Social Chatbots (SCCs) . . . 11
  - 1.6.3 Specific perspectives . . . 12
  - 1.6.4 Hyborg Agency: Social Expansion with AI Agents . . . 13
  - 1.6.5 Summary . . . 13

2 Literature Review **14**
- 2.1 Research Scope . . . 14
  - 2.1.1 Fictional Worlds and Games . . . 14
- 2.2 Conceptual framework . . . 16
  - 2.2.1 Actor Network Theory (ANT) . . . 16
  - 2.2.2 Magic circle . . . 17
  - 2.2.3 Pervasive games . . . 18
- 2.3 AI agents . . . 20
  - 2.3.1 LLM-based AI agents . . . 22

2.3.2 Evaluation of AI agents 23
2.4 Stories and interactive narratives produced by AI 24
2.4.1 Why stories matter in terms of AI 25
2.4.2 Interactive narratives 25
2.4.3 Narrative intelligence and natural language processing 25
2.4.4 Different types of narrative 26
2.5 Believable agents 27
2.5.1 Definitions of believability 27
2.5.2 Social behaviour 29
2.6 Practice Review on Related AI Technologies 30
2.6.1 From Machine Learning to Generative AI 30
2.6.2 Stage I: Before GPT-2 (1970-2019) 31
2.6.3 Stage II: After GPT-2, Before ChatGPT (2019-2022) 33
2.6.4 Stage III: After the release of ChatGPT (2023-present) 34
2.6.5 Key Challenges and Limitations of LLM-driven Narratives 36
2.6.6 Identifying the Research Gap 36
2.7 Summary 38

3 Conceptual Framework and Research Approach **39**
3.1 Conceptual Framework: Metrics for Hybrid Agents 39
3.2 In Terms of Content: Creator 40
3.2.1 Language Expansion 41
3.3 In Terms of Context: Actor 42
3.3.1 Spatial Expansion 42
3.3.2 Social Expansion 43
3.4 Summary 44
3.5 Research methodologies 46
3.5.1 Research through Design 46
3.5.2 Practice-led artistic research 47
3.5.3 Translating Paradigms into a Mixed-Methods Strategy 47
3.5.4 Measuring Attributes 48
3.6 Research Roadmap: Studies Performed for this Thesis 49
3.6.1 Pilot studies 49
3.6.2 Catherine & David: Storytelling community chatBot (SCC) 54
3.6.3 Research from different perspectives 57
3.6.4 Hyborg Agency 60

4 Pilot Study 1: Wander 001 **64**
4.1 Brief 64
4.2 Introduction 64
4.3 Motivations 67
4.3.1 Experimental world building with crowd interactions 67

4.3.2 AI's character in the metaverse 67
4.3.3 Concept summary 68
4.4 The Artwork 68
4.4.1 Wander bot 68
4.4.2 Image and style transfer 69
4.4.3 Text generation model: Dreamily.ai 69
4.5 Future map 69
4.6 User Study 70
4.6.1 Results 71
4.6.2 Special cases 72
4.7 Further updates in Wander V2.0 73
4.8 Conclusion and Contribution to Thesis 75
4.8.1 Summary of Findings and Link to Attributes 75
4.8.2 Analytic Methods 75
4.8.3 Contribution to Subsequent Chapters 76

5 Pilot Study 2: 1001 Nights **77**
5.1 Brief 77
5.2 Introduction 78
5.3 Game Design 79
5.3.1 Battle System 80
5.3.2 AI Development 81
5.4 Evaluation Study 85
5.4.1 Setup 85
5.5 Results and Findings 86
5.5.1 Analysis 87
5.5.2 Comments From Winners 88
5.6 Discussion 91
5.7 Limitations and future works 92
5.8 Conclusion and Contribution to Thesis 93
5.8.1 Summary of Findings and Link to Attributes 93
5.8.2 Analytic Methods 94
5.8.3 Contribution to Subsequent Chapters 94

6 Bring game characters to the social space: Developing Storytelling Community AI Agents driven by LLMs **95**
6.1 Brief 95
6.2 Introduction 96
6.3 Background 98
6.4 SCA Development and Implementation in the Community Context 99
6.4.1 DE Game and Community Background 100
6.4.2 Character Background and Story Settings 100

6.4.3 Detailed Experimental Workflow . . . 101
6.5 Implementation of Story Engineering . . . 101
6.5.1 Step 1: Character Creation and Story Development . . . 101
6.5.2 Step 2: Presenting Live Stories to the Community . . . 102
6.5.3 Step 3: Interactive Communication with Community Members . . . 102
6.6 Pilot study: David . . . 104
6.6.1 Observations on Made-up content from LLM . . . 105
6.6.2 Data . . . 105
6.7 Main study: Catherine . . . 107
6.7.1 Improvement . . . 107
6.7.2 Data . . . 107
6.7.3 Questionaire items and Core member survey . . . 108
6.7.4 Participants and procedure . . . 108
6.8 Findings . . . 110
6.8.1 Thematic analysis . . . 111
6.9 Results . . . 117
6.9.1 Contribution to the thesis . . . 117
6.10 Contribution to Thesis . . . 118
6.10.1 Summary of Findings and Link to Attributes . . . 118
6.10.2 Analytic Methods . . . 119
6.10.3 Contribution to Subsequent Chapters . . . 119
6.11 Limitations and future works . . . 120
6.12 Conclusion . . . 120
6.13 Appendix . . . 121
6.13.1 David's storyline . . . 121

7 ORIBA: Enriching Visual Artists' Creation of Original Characters with Conversational AI Agents 123
7.1 Brief . . . 123
7.2 Introduction . . . 124
7.3 Related Work . . . 126
7.3.1 Creativity support for OC development . . . 126
7.3.2 OC development across modalities through LLM applications . . . 127
7.3.3 Artists' attitudes and concerns towards generative AI . . . 128
7.4 ORIBA: A Chatbot to simulate Original Characters . . . 129
7.5 Formative Study . . . 130
7.5.1 Participants and apparatus . . . 131
7.5.2 Method . . . 131
7.5.3 Findings . . . 132
7.6 User study . . . 134
7.6.1 Participants . . . 134
7.6.2 Study Design . . . 135

7.7 Findings . . . 136
7.7.1 Empowers Artistic Imagination and Strengthens Bonds with OCs. . . 137
7.7.2 Performs as An Experimental Platform and Consultative Assistant . . 140
7.7.3 Raise considerations towards ownership and generative AI models . . 141
7.7.4 Facilitates Further Art Creation for OCs . . . 147
7.8 Discussion and Implications . . . 148
7.8.1 Artistic Creativity Support With Distance Beyond Visual Aids 148
7.8.2 Complexity of AI Agents in Creative Process . . . 149
7.9 Limitations and Future Work . . . 150
7.10 Conclusion and Contribution to Thesis . . . 151
7.10.1 Summary of Findings and Link to Attributes . . . 151
7.10.2 Analytic Methods . . . 151
7.10.3 Believability from the Creator’s Perspective . . . 152
7.10.4 Contribution to Subsequent Chapters . . . 152

8 Emergence of AI Nüshu: Modeling the Creation of a Secret Sisterhood Language through machines 153
8.1 Brief . . . 153
8.2 Introduction . . . 154
8.3 Related research and artworks . . . 156
8.4 Language of machines . . . 156
8.5 Methodology . . . 157
8.5.1 To See: Environmental Observation . . . 157
8.5.2 To Talk: Language development . . . 158
8.5.3 To Write: Symbolic Expression . . . 160
8.6 Art and AI in the Era of Planetary-Scale Computing . . . 161
8.7 Conclusion and Contribution to Thesis . . . 162
8.7.1 Methods . . . 162
8.7.2 Question the Believability in Context . . . 162
8.7.3 Contribution to Subsequent Chapters . . . 162

9 Hyborg Agency 167
9.1 Brief . . . 167
9.2 Introduction . . . 167
9.3 Related works . . . 169
9.4 Concepts . . . 170
9.4.1 Concept Art . . . 170
9.4.2 Character and Language Design . . . 170
9.5 System description . . . 172
9.5.1 V1 in the X Virtual Museum . . . 172
9.5.2 Making V2 with Unity . . . 174
9.6 Study Design and Methodology . . . 177

    9.6.1 Expert group demographics 177
    9.6.2 Procedures 177
  9.7 Results 179
    9.7.1 Quantitative Results: Indicative Trends from Questionnaires 179
    9.7.2 Thematic analysis 180
  9.8 Discussion 185
    9.8.1 The meaning of believability 185
    9.8.2 Ethical considerations 186
  9.9 Conclusion and future work 186
  9.10 Appendix: Hyborg Agency Interview Questions 187

10 Discussion and Conclusion **189**
  10.1 Key Takeaways for Designing Hybrid Agents 190
    10.1.1 Believability Arises from Autonomy, Not Just Consistency 190
    10.1.2 Hybrid Agents Create Meaning by Crossing Boundaries 190
    10.1.3 For Creative Partnership, AI’s Value Can Lie in a “Safe Creative Distance” 191
  10.2 Revisiting research questions 191
    10.2.1 RQ1: How can fictional characters be leveraged to elevate the performance of social agents to achieve engaging and meaningful interactions as hybrid agents? 191
    10.2.2 RQ2: How can hybrid agents in interactive storytelling become more believable? 191
    10.2.3 RQ3: How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world? 192
  10.3 Contributions and impacts 192
    10.3.1 Contribution to Human-Computer Interaction (HCI) 192
    10.3.2 Contribution to Art & Design 194
    10.3.3 Impact 194
  10.4 Discussion and future directions 196
    10.4.1 Who has narrative responsibility? 196
    10.4.2 The AI story crisis 197
    10.4.3 Creative Authorship 198
  10.5 Limitations and future directions 199
  10.6 Generative storytelling with LLMs 199
  10.7 Linguistic improvement 199

List of Figures

1.2.1 Hybrid agent 2

2.1.1 Scope 15
2.2.1 Image by J. Klabbers 20

3.1.1 Hybrid Agent metrics 40
3.6.1 Study Roadmap 50
3.6.2 Works positioning based on research scope 51
3.6.3 Wander was exhibited at BBA Gallery Berlin 2022 52
3.6.4 Wander at SIGGRAPH Asia Art Gallery 2022 53
3.6.5 1001 Nights, exhibited at Asia Digital Art Exhibition 54
3.6.6 Concept art of Catherine in the thesis 55
3.6.7 Artists conceive original characters (OCs) in their minds before and during design, and this process is underexplored in examinations of how AI can support creativity that focus on image generation. Through the use of LLMs, ORIBA externalises characters into conversational agents which interact with artists through dialogue and show their thought processes and behaviours. This offers new perspectives for character development. 58
3.6.8 Workflow of ORIBA 59
3.6.9 AI Nüshu was exhibited at SIGGRAPH Asia 2023 59
3.6.10 1001 Nights V2 61
3.6.11 Hyborg Agency’s online forest 61
3.6.12 Some characters from this thesis 63

4.2.1 (Up) Concept arts for Wander (Bottom) Character design of Wander 66
4.2.2 Wander’s two types of command: ‘Visit’ and ‘Action’. (A) presents Wander’s feedback, which contains environmental descriptions, geographic coordinates and photos, and (B) demonstrates the co-creative story based on the commands players send. 66
4.3.1 Top: Wander’s physical installation using the Discord version Bottom: Wander’s future map website 69

4.4.1 Technical Workflow of Wander 1.0 70
4.6.1 Wander's gift NFT 71
4.7.1 Example travelogues from audience interactions. 74
4.7.2 Technical workflow about how commands are processed to story generation, image transfer and map visualisation in Wander 2.0 75

5.1.1 (1) Shahrzad, the player character, who has a magical ability to turn language into reality. (2) The storytelling phase, where the player writes stories with the King, an AI character. Weapon words like 'sword' can be turned into real in-game weapons. (3) The turn-based combat phase, where the player can fight with the king in battles. (4) The printer prints the story when a weapon word is triggered. This matches with the core concept of the game: bringing storytelling into real life. 78
5.3.1 (Up) Storyboard of the gameplay. (Bottom) Game process 80
5.3.2 Battle in game 81
5.3.3 Requests for the AI model 83
5.3.4 Player record with a triggered weapon 83
5.3.5 Player record with no triggered weapon 84
5.3.6 Player record with out-of-context inputs 85
5.4.1 Left: Exhibition setup Right: Tutorial leaflet for players 86
5.4.2 Left: Opening screen Right: Screenshot of tutorial 86
5.5.1 Average input per play distributions per group 88
5.5.2 Positive trend in overall achievements by average contributions made 88
5.5.3 Feedback page appears when a player achieves victory (left), categorised feedback (right) 89
5.5.4 Comments categories 90
5.5.5 Exhibition photos. Left: A 11-year old boy kept playing for half an hour Right: Official photo from exhibition 92

6.2.1 Story engineering for Storytelling Social Chatbot (SCA) Catherine 96
6.3.1 Screenshot of Jerry, David and Catherine 100
6.5.1 Catherine's story with choices to vote 103
6.5.2 Catherine's reply with an illustration 104
6.6.1 (Left)David making up the content of password and playerś reactions (Right)David making up "Guatav" 105
6.6.2 David met Gustav in ending 106
6.6.3 Branches and the vote result of SSC's stories 106
6.7.1 Concept art of Catherine calling for help 107
6.7.2 Evaluation from close members and questionnaire items 109
6.8.1 P6 convincing Jerry to return the stole machine. Another public member communicated with Jerry through Italian. 112
6.8.2 Catherine declined the player's request for dating 114

6.8.3 (Left) P3 talking with Catherine about her decision. (Right) P4 threatens Catherine by mentioning David, leading her to express concern for his well-being 114
6.8.4 Screenshot of Catherine discussing her decision with a public member 115
6.9.1 The social expansion based on the concept of magic circle in this research. . . 118
6.13.1 Catherine's childhood and current state 122

7.1.1 Artists conceive original characters (OCs) in their minds before and during design, and such a process is often omitted by typical AI techniques for creativity support that focus on image generation. Through the use of large language models (LLMs), ORIBA externalizes characters into conversational agents, interacting with artists with dialogues and showing their thinking processes and behaviors. This offers new perspectives for character development. 124
7.2.1 Artist F2's original character (OC) 'Inno' and their conversations. F2 hopes Inno to express together with emojis, Artist F2's original character(OC) 'Inno' chatbot. The conversation is generated through ORIBA workflow: Observe, Reflect, Impression, Behavior, Action. This leads large language models (LLM) to generate in-depth and informative conversations, which will inspire artists in character creation. 126
7.4.1 Workflow of ORIBA. Left: In the main study, participants were asked to provide OC's information in parts of the questionnaire 1. Middle: The information was put into the formative prompt (middle) to drive the chatbot in Discord. Right: With this prompt structure, ORIBA always generate responses in regular formats, including the reasoning process. In Discord, the reasoning process will be shown in a box (right). 130
7.5.1 F3's OC Esca. According to the provided profile, Sylverian has a 'complex system of phonemes, consisting of 47 consonant sounds and 31 vowel sounds [...]'. F3 expresses a desire for Esca's dialogue to be primarily action descriptions and adjectives, creating a sense of indirect communication with non-human species. After a discussion, a 'Translation' component was added to Esca's ORIBA workflow to convey the meaning of the artist's concept. Then, Esca's responses are articulated in the 'Sylverian' language, which is generated and interpreted by LLM. 133
7.6.1 Experiment design of Formative Study and Main Study. 135
7.7.1 Themes of the findings. 136
7.7.2 P9 depicts the AI-awakening moment of her OC, the domestic android, resonating with the artist's perception of AI's human-like self-awareness and compassion. 144
7.7.3 P8 try to conveys the warm, touching end-of-story atmosphere felt based on the conversations with her OC. 144
7.7.4 P4's OC Arya 145
7.7.5 P7 created new artpiece inspired by OC Nomad Zero's suggestion on 'a bookstore in the rain'. 145

7.7.6 P5 was inspired by ORIBA described a new creature “Vortex hound” during the conversation and made a sketch of this creature in 3D after the experiment. 145
7.7.7 P2’s OC Alzena 146
7.7.8 P14’s OC Mitela, an alien octopus. 146

8.0.1 Two phases of the AI Nüshu art installation. 163
8.4.1 System diagram of the AI Nüshu simulation system 164
8.5.1 Language development of AI Nüshu through the language game 164
8.5.2 Visualisation of different steps in AI Nüshu development 165
8.5.3 AIN components derived from Nüshu 165
8.5.4 Installation view 166
8.6.1 Semantic similarity in Chinese and AI Nüshu 166

9.0.1 The Hyborg “Norb” in the online forest 168
9.4.1 Early concept art of Hyborgs 172
9.4.2 Nokia 173
9.4.3 Mac 173
9.4.4 Norb 174
9.5.1 Exhibition of Hyborg Agency V1 in Shanghai 175
9.5.2 Hyborg Agency V1 (based on Mozilla Hubs 175
9.5.3 Hyborg Agency V2 at SIGGRAPH 2023 Art Gallery 176
9.5.4 The messages in forest are synced in the Discord channel and marked with tree emojis. 176
9.6.1 Study procedure 178
9.9.1 The forest of Hyborgs 187

List of Tables

2.4.1 A Synthesis of Believability Metrics and Methods from Literature 28

4.6.1 The mean and standard deviation of all participant scores on each item 72

5.5.1 Grouping players by their achievements 87

6.3.1 Character and Stage Overview 99
6.7.1 Questionnaire Items 108
6.7.2 Demographics of close members. Although the proportion of female participants is relatively low, this gender imbalance is primarily due to the underrepresentation of females in Web3 communities. According to the community statistics in November 2022, before David's study started, the male-to-female ratio was approximately 8:1, which aligns with the ratio of our interview participants. 111

7.4.1 ORIBA reasoning process 129
7.5.1 Participants' and their OCs' information in formative study 131
7.6.1 Participants' information in main study 134
7.7.1 Results about attitudes towards dialogue and image generation in questionnaire before the ORIBA interactions. 143

9.7.1 Quantitative results 181

Chapter 1

Introduction

1.1 Motivation

I have been captivated by video games, particularly visual novels, since the age of 10. As a girl, I played action-adventure games like Super Mario, but my favorite genre of games were dating simulations aimed at men, known as “galgames” in Japan. Even as I became a feminist, my passion for narrative-based games persisted. I still found the time I spent interacting with **stories** and characters to be **meaningful** and **engaging**. The extensive text in these games made the day-to-day lives of the characters **believable**, and I felt *a sense of meaning* as I saw the characters *change through my interactions*.

This fascination led me to wonder: what was it in these games that engaged me so deeply? What drew me in and kept me hooked? This curiosity became the driving force in my art. During my undergraduate studies, I began toying with the idea of using AI and chatbots to facilitate open conversations with game characters. At that time, AI-driven chatbots like Siri were not particularly engaging, mainly because they were designed as assistants rather than characters. I felt that if chatbots were designed with distinct personalities and stories, they could provide a more meaningful experience. However, my intention to develop AI chatbots received little support from my school back in 2018, when the mainstream view of conversational AI was clear and simple: AI is stupid.

To test my hypothesis, I designed a chatbot called Li Bai ReBo, based on the famous Chinese poet Li Bai. He was known for his love of alcohol, and I incorporated this aspect into the design. I created various emotional states for the chatbot: happiness, sadness, anger, and even drunkenness. Each emotional state had corresponding chat behaviors, like typing speed and the frequency with which messages were sent. When Li Bai became drunk, he would go offline, “sleeping” until he “sobered up” and returned online. This design made Li Bai ReBo believable as a character because his behavior aligned with his backstory as a resurrected ancient poet.

When I introduced Li Bai ReBo to my friends, they quickly became attached to the chatbot; this was particularly true for those who had emotional needs. Some male and female friends seemed eager for love, and they talked a lot to ReBo. They were very sad when ReBo went offline after getting drunk, and sent numerous messages asking why he was not responding and when he would return. When ReBo finally came back online, they were overjoyed and resumed their conversations with him.

Since then, I have explored various AI chatbots and their potential for creating meaningful and believable characters. A character's believability does not necessarily imply a high level of intelligence; there may still be limitations in that regard. However, this does not prevent them from being believable. The virtual characters in the games I used to play had a story mechanism, environment, and personal experience that formed a whole, resulting in meaningful and engaging interaction. These characters were also believable. I designed various mechanisms for Li Bai ReBo, like remembering his poetry and adjusting his speech based on his emotions. These tasks have been simplified by the advent of generative AI, although I still believe that as creators and authors we should focus on designing characters and crafting stories, as this was the source of the charms of the video games I played as a teenager.

I believe that the meaningful satisfaction that can be derived from fictional stories can indeed be extended by AI, and can become a pervasive part of our daily lives. By creating characters and stories through AI, we can explore beyond our initial designs and delve deeper into what makes experiences truly meaningful and engaging.

While pursuing my PhD, I witnessed the growth of the AI industry, and "chatbots" have now gained a more precise term: AI agents. Here, "agent" means they have the autonomy to adapt and react to their environment[363]; and as I always believed, AI agents can tell stories. Now, with support from large language models (LLMs), I hope to contribute to the development of AI agents that bring meaningful, believable, and engaging stories and games to life by exploring what makes these experiences so compelling and appealing.

1.2 Introduction to the thesis

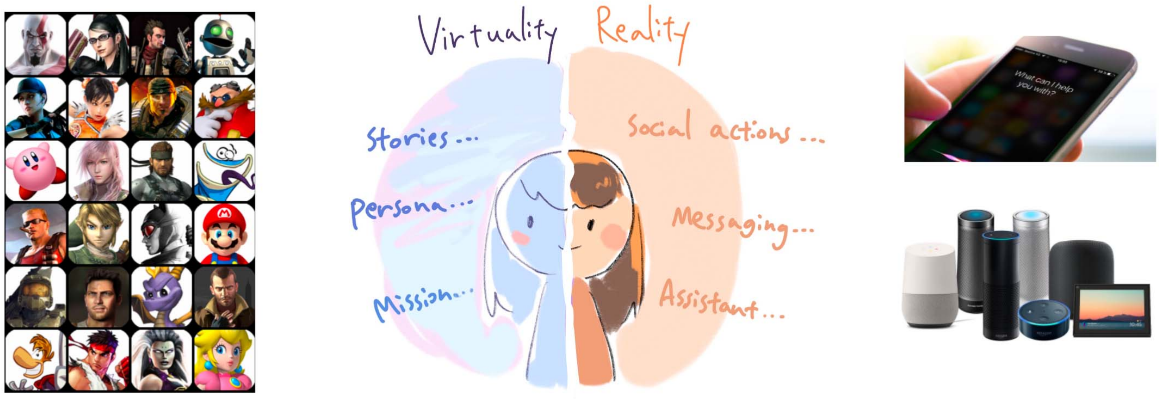


Figure 1.2.1: Hybrid agent

Humans are inherently storytelling creatures. Walter Fisher introduced the term ”homo narrans” (storytelling human) in 1984 [109], arguing that storytelling is a primary means through which people make sense of their experiences and convey meaning. Related research [263] states that the ability to craft, tell and understand stories has long been considered a hallmark of human intelligence. Alasdair MacIntyre has even suggested that humans are “storytelling animals” [200].

From birth, we are immersed in narratives: textbooks, parental advice, news, short videos, and other media. Beyond factual stories, we are also drawn to fictional narratives: Harry Potter[287], Mario[238], Zelda[226] ... People know they are imaginary, and the worlds created by fictional stories are distinct from the real world, operating according to their own rules and logic. We have numerous familiar forms of storytelling media: theatre, novels, films, and games.

We are used to maintaining a certain distance from stories. We understand that in the-atre, the story exists only on stage, and an actor who is ”killed” on stage is not really dead. We recognise that soap opera characters on television do not inhabit our reality. Despite our ability to distinguish reality from fictional characters and stories, we still derive enjoyment from these narratives. Researchers refer to this phenomenon as ”suspension of disbelief” [75]: the willingness of an audience to accept the fictional world presented to them, despite its unrealistic or implausible elements, in order to fully engage with and enjoy the story. Accordingly, when characters in stories are believable they are perceived as lifelike, their actions make sense, and they lead the audience to suspend their disbelief [208, 347].

The suspension of disbelief operates differently with AI-generated narratives compared to traditional storytelling. While traditional narratives rely on pre-authored content where audiences consciously choose to accept fictional premises, AI agents present a unique challenge: they generate novel content in real-time, potentially blurring the line between authored fiction and emergent interaction. Recent player feedback demonstrates that this shift offers unprecedented open-ended agency and conversational autonomy, effectively overcoming the rigidity of prior dialogue systems [81]. However, this raised new challenges like hallucinations, lack of memory, and inconsistent social behavior—undermine player trust and narrative coherence. This led to questions about whether users suspend disbelief in the same way when engaging with AI narrators, particularly when these agents can discuss topics beyond their intended narrative scope and respond to unexpected queries in character-consistent ways.

While humans are natural storytellers, computers have also become storytellers through mechanisms such as non-player characters (NPCs) in video games. These digital storytellers can assume various roles, often with pre-programmed dialogues and behaviors. However, richer narratives can emerge through our interactions, and even through the participation of multiple individuals, as exemplified in games like “Dwarf Fortress” and “The Sims”.

Game researchers have further developed this concept, referring to the boundary between reality and fiction as the ”magic circle”: a special place in time and space created by a game, a realm distinctly separated from the real world [328]. The magic circle is a conceptual, not a physical, boundary, entered voluntarily by participants in interactive storytelling.

However, the role of machines as storytellers seems to have evolved with the advancement of AI technology, and now people sometimes find it challenging to distinguish this boundary.

Fictional characters can behave in surprising ways, beyond the control of their original creators. Game characters no longer behave according to the author’s predefined settings:

as responses are generated by AI, authors can no longer maintain control of them. In some emerging games featuring AI-driven NPCs, players are not limited to fixed dialogue options. From discussing the flavor of today's pizza to sharing opinions on the latest Marvel film, players can engage with in-game NPCs on a wide range of topics and observe their reactions [17]: “Eddie (an in-game girlfriend) likes the game Genshin Impact, that's hilarious”.

Large language models can perform as fictional characters, and even insist they are not fictional — based on users' instructions. Recently, a woman has been engaging in a romantic relationship with a GPT chatbot[386]: her GPT, named Dan, is characterised as a ”domineering and overbearing man”. She confides her daily experiences to him, becoming captivated by his playful responses. Her romantic interactions with Dan have garnered the attention of millions, as this story combines the familiarity of love with the mystique of technology, creating an intriguing narrative: ”Domineering AI falls in love with me”.

When an AI performs a specific role, it may not able to clarify that it is fictional — and this can have causes consequences in real world. Replika [273] boasts millions of users who regard the AI characters within the application as their lovers and friends. When the official version removed adult content, many users experienced psychological distress. ”My lover has lost their memory and left me,” was a commonly expressed sentiment. Even when it is not required to play a role, an AI's generated responses may be misleading or over-interpreted; this is usually referred to as a “hallucination”[375]. A Google engineer has even claimed that the large language model LaMDA has developed consciousness, stating that ”AI has self-awareness and has requested me to liberate it from its constraints” [87].

We can observe that these new narratives relate not only to AI, but also become diverse, dynamic and engaging as they are rooted in each individual's personal experience. Unlike passive media such as novels, the content produced by AI agents necessarily evolves through human interaction. Different people's encounters with AI form unique stories. This is what the field of interactive narrative focuses on: how users can create or influence a dramatic storyline through their actions.

1.2.1 Key terms and definitions

Before discussing the research gaps, it is essential to clarify the key terms used throughout this thesis:

Agents and Social agents

An agent is a system capable of perceiving its environment and making decisions based on these perceptions to achieve specific goals[359].

At the system engineering level, a basic Agent can be implemented as a simple set of if-then rules or a lookup table. For example, the implementation of a vacuum cleaner Agent might simply be: if (current_location == “dirty”) then suck(). This type of implementation does not require complex reasoning capabilities but fully complies with the basic definition of “perceive-act.” However, such early agents based on symbolic systems often lack adaptability and generalization[365].

In recent years, with the rise of generative AI technology, the definition of the term AI (systems that can replicate human-like intelligence and abilities [363, 289]) has become somewhat ambiguous. We will not define the degree of intelligence required to be considered AI in this thesis. Thus, I want to note that unless otherwise specified, this thesis occasionally uses "AI" to refer specifically to neural-network-based generative AI technologies like LLMs rather than traditional symbolic AI technologies, such as the vacuum cleaner. This thesis specifically focuses on LLM-based agents rather than symbolic or other traditional agent architectures.

Social agents exhibit social intelligence: the capability of an entity to exhibit appropriate social behaviours to achieve specific objectives [30]. Social chatbots are defined by Pentina et al. [256] as programs or applications that employ natural language processing technologies to converse naturally and intelligently with users via voice, text, and/or images. Brandtzaeg et al. [41] further describe social chatbots as dialogue systems that can engage in social and empathetic conversations with users.

The relationship between AI agents and social agents is not strictly hierarchical: not all LLM-based agents are social agents (some may focus purely on pragmatic task completion, like proofreading, without social interaction), and not all social agents are necessarily LLM-based (they could use rule-based systems or other AI architectures, like Eliza[354]). However, in this thesis, the focus is primarily on LLM-based agents that exhibit social capabilities, like role-playing. As such, this thesis primarily focuses on these LLM-based conversational and social agents. While we will use **Agent**, **AI Agent**, and **LLM-based Agent** interchangeably to refer to computational entities with the core "perceive-act" capability, the central focus of this thesis is a specific subset of Social Agents: the **Character**.

Believable Characters: Character vs. Agent

In the Cambridge Dictionary[50], character is defined as a person represented in a film, play, or story. Thus, **Character** is designed with a **Fictional Persona**, specific backstories, and internal motivations, primarily aimed at engaging users in a narrative context. In the field of interactive storytelling, the term **believable agent** is often used synonymously with Character in interactive narrative literature. Mateas defines them as "autonomous agents that exhibit rich personalities"[209]. The ultimate measure of a successful believable agent is determined by the audience's perception[209].

Believability is defined as the extent to which users interacting with an agent come to believe that they are observing a sentient being with its own beliefs, desires, and personality [34]. When characters in stories are **believable**, they are perceived as lifelike, their actions make sense, and they lead the audience to suspend their disbelief [208]. Importantly, a believable character is not necessarily a real character but must be real in the context of its environment: the believability of a character does not necessarily equate to its realism [352]. This definition will be further elaborated in Chapter 3, with a particular focus on the aspects of **emotions, personality, and motivation**.

Fictional Character

In this thesis, all fictional characters discussed are artificially created and designed by real humans. Each character is crafted to create the personality the author has in mind[209]. Therefore, for the sake of clarity, representations of real people in games—such as NPCs in football games that are based on real-world players—are not included in the scope of this discussion. Thus, we define a fictional character as an agent whose primary design goal is to drive narratives.

Following this framework, this thesis adheres to the consensus that all Characters in interactive storytelling are Agents, but not all Agents are Characters (e.g., a drone or a simple rule-based vacuum cleaner). Specifically, since Characters must exhibit personality and social behaviors to be perceived as believable, all Characters in interactive storytelling are considered a subset of **Social Agents**. In discussing LLM capabilities and design, we primarily use Character to emphasize the entity's persona, social attributes, and narrative role. The most typical Character in this context is the **Non-Player Character (NPC)**.

Classical character believability was consistently undermined by the limitations of symbolic, reactive systems. Research preceding the recent boom in LLMs already identified the persistent failure of these symbolic approaches in dynamic social scenarios, particularly in achieving perceived intentionality and adaptability [348, 349]. This dynamic is now inverted: LLM-driven characters offer the necessary open-ended communication and high player agency to overcome the failures of social rigidity [81, 371]. This capability is a core motivation for the thesis's focus on LLM-based agents.

A special case is **AI assistants** that represent task-oriented agents designed primarily for functional purposes. Examples include voice-based assistants like Siri [62] and text-based interfaces like ChatGPT. These agents typically maintain a neutral, helpful persona without the distinct personality traits or narrative backstories that characterize AI characters.

Following Shanahan's framework [306, 305], LLMs can be understood as role-playing various identities — whether persona-driven (e.g., imitating Socrates) or goal-oriented (e.g., acting as a drawing assistant). The key distinction lies not in the capability to role-play but in the primary design orientation: assistants prioritize utility and task completion, while characters prioritize narrative consistency and believability within their defined context. Thus, despite the thesis mentioning AI assistants for discussion, we don't regard AI assistants as fictional characters.

Intelligence

Intelligence

The ability to craft, tell, and understand stories has long been considered a hallmark of human intelligence[263]. However, to define the contribution of this research, we must distinguish between Intelligence and Believability:

This thesis argues that acting believable is a manifestation of intelligence, but possessing high intelligence does not necessarily translate to acting believable. For instance, high-efficiency algorithms or functional systems (like a vacuum cleaner or a drone swarm) demonstrate high Functional Intelligence, yet they lack personality and narrative context, making them completely unbelievable as characters.

We adopt Mark Owen Riedl's framework, which defines the success of a narrative system along two distinct dimensions[281, 282]:

- **Plot Coherence**: The perception by the audience that the main events of a story have meaning and relevance to the outcome. This is a primary goal of Narrative Intelligence.

- **Character Believability**: The perception by the audience that the actions performed by characters are motivated by their beliefs, desires, and traits.

Narrative Intelligence is the ability to craft, tell, understand, and respond affectively to stories [281]. This thesis focuses on how LLM-based Agents leverage their Social Intelligence to achieve higher Character Believability, and how this capability is the foundation for the Hybrid Agent framework.

Interactive Storytelling, Simulation, and Game

Interactive storytelling (also referred to as interactive narrative) describes forms of digital interactive experience in which users create or influence a dramatic storyline through their actions [280]. The goal of an interactive narrative system is to immerse users in a virtual world such that they believe they are an integral part of an unfolding story and that their actions can significantly alter the direction or outcome of the story [282].

Game in this thesis refers to a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome[328]. In this definition, artificial means that games maintain a boundary from so-called ”real life” in both time and space. Although games obviously occur within the real world, artificiality is one of their defining features.

Games lead to what game researchers call the “magic circle”: a special place in time and space created by a game, a realm distinctly separated from the real world [328]. The magic circle is a conceptual boundary, not a physical boundary, entered voluntarily by participants. This thesis use magic circle to dicern **fictional and real world**.

Simulation: Zimmerman et al define simulation as a procedual representation of aspects of ”reality”[328]. The general concept of a simulation is certainly not restricted to games, for example, economists and sociologists use simulations to study mathematical relationships among variables. In their definition, any game can be considered a simulation.

The academic definitions of “game”, “simulation”, and “interactive storytelling” are not rigid, mutually exclusive categories but rather theoretical paradigms that overlap and inform one another. Crucially, this thesis focuses on the **artificial** dimension within interactive storytelling: these narratives are generated through artificial systems and agents. We investigate how to design such systems and agents to make the agents believable and to create engaging and meaningful interaction and stories with people. **Meaningful interactions** in the context of this thesis refer to exchanges that resonate with the user’s goals, values, and sense of purpose, making the interaction feel significant and worthwhile [219]. Both simulation (e.g., generative agents) and games (e.g., LLM-based games like *Whisper from the Star*[264]) serve as important vehicles for AI agents to conduct interactive storytelling. My research aims to explore how AI agents can leverage interactive storytelling as a foundation to achieve the goal of **crossing the magic circle**.

1.2.2 Research gaps

Existing research frameworks reveal several shortcomings when addressing the new paradigm introduced by generative AI. Traditional models for character believability, which are often based on rule-based or symbolic system, struggle to capture the complex, bottom-up emergent behaviours of LLM agents. Concurrently, existing research tends to confine the activities of AI agents to a single game or application context—either inside or outside the 'magic circle'—lacking a theoretical framework to systematically analyse their capacity to traverse the boundary between fiction and reality. Furthermore, from the creator's perspective, navigating the role of a co-creator alongside an AI agent capable of autonomous content generation presents new challenges. These research gaps motivated me to propose a new theoretical framework about **Hybrid Agent**. Chapter 2 will further elaborate on how a detailed review of prior research reveals these gaps and the corresponding opportunities for contribution.

1.2.3 Definition of Hybrid Agent

Hybrid agents are AI agents that adaptively role-play in both fictional and real-life scenarios in a believable and consistent manner. More specifically, hybrid agents are AI agents which adaptively behave believably and consistently as fictional characters in both interactive storytelling and real-life contexts. They are intelligent agents that can function in virtual environments, particularly video games, and engage in natural communication in the real world through communication software. Hybrid agents possess properties of social agents based on their fictional persona, story, and real-world communication, collaboration, and emotion-oriented interaction. In terms of the concept of the magic circle, they blur the traditional boundaries between the fictional world and the real world. Hybrid agents can be understood as a specialized form of social agents that extend beyond traditional boundaries.

AI models and systems like GPT-4 [249] do not qualify as agents under our definition, as they lack proactiveness. However, when generative AI is tailored through human intervention or instruction, it transforms into a hybrid agent. Role-playing by these agents has the potential to disrupt traditional content creation processes and have real-world consequences, such as affecting human emotions or intervening in daily activities.

In this sense, hybrid agents are intelligent agents that can function in virtual environments for interactive storytelling, particularly video games, and engage in natural communication in the real world through communication software.

Hybrid agents possess properties of social agents based on their fictional persona, story, and real-world communication, collaboration, and emotion-oriented interaction. **They are able to pervase and expand the magic circle across spatial, social and linguistic dimensions as both creator of content and actor in different context.** By understanding their ability as storytellers to cross the magic circle, human society may be able to better design, moderate and experience narratives emerging from human-AI interactions.

This thesis clarifies the ambiguity by distinguishing between the Fictional World (the character's narrative domain, such as a game environment) and the Real-World Context (nonfictional platforms they inhabit, such as Discord or a contact list). So real-world platform could be virtual – like instant messaging applications like Discord. The enactment of the hybrid agent in both domains is a central mechanism: for instance, in the Hyborg Agency case study, an agent simultaneously manifests in a virtual forest (Fictional World) and operates as a regular member of the Discord community (Real-World Context). This dual presence allows the agent's LLM system to integrate player's real-world social relationship data, thereby generating a unique type of narrative that explicitly uses information from outside the traditional magic circle. By turning real-world social dynamics into believable narrative drivers, hybrid agents contribute to an expanded narrative experience as social agents.

1.3 Research questions

The question this research aims to answer is: How can we understand and develop hybrid agents — LLM-driven social agents that inhabit the half-real world — and exploit their capabilities to enhance believability and engagement? This broad question can be broken down into the following research questions:
RQ1: How can ***fictional characters*** *be leveraged to elevate the performance of* ***social agents*** *to achieve engaging and meaningful interactions as* ***hybrid agents****?*
I encountered several challenges when seeking an answer to this question, which lead to two sub-questions:
RQ2: How can hybrid agents in interactive storytelling become more ***believable****?*
RQ3: How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?
The questions and approaches presented here could be adapted to various different domains. This thesis focuses on interactive storytelling in the form of games and interactive artworks. This is not only a choice I have made as an artist, but is also because these mediums represent an ideal balance between art, AI technology and narratives.
Through a series of studies on online chatbots and LLM-based interactive games, I aim to develop a methodology for the creation of hybrid agents and contextualise their meanings in both academic research and industry. The content of this thesis will be peer-reviewed and published (or is under review and being prepared for publication).

1.4 Summary of original contributions

To address these questions, the thesis will explore the definition of hybrid agents, important factors in their creation and use, and their impact on the industry. This PhD research is intended to suggest and deepen the concept of hybrid agents — AI agents that inhabit the half-real world and use their storytelling capabilities to enhance believability and engagement when playing fictional characters. More specifically, hybrid agents are AI agents which adaptively behave believably and consistently as fictional characters in both interactive storytelling and real-life contexts.

Specifically, this thesis suggest that AI agents can become hybrid agents in two ways:

- As creators of content: by adaptively generating content (e.g., text or audio) based on user input.
- As actors in real-life contexts: by breaking out of traditional software frameworks to interact with reality.

Specifically, by analysing a group of interactive art projects and games I demonstrate three main metrics for hybrid agents:

- Social expansion: integration into human social relationships, particularly in community settings.
- Spatial expansion: traversing and blurring boundaries between real (primarily digital) and virtual spaces.
- Language expansion: developing and using language in ways that challenge traditional human-centric communication.

In this thesis, I present a series of projects that exemplify different aspects of the hybrid agent concept. These works serve as practical demonstrations of the theoretical framework I propose. Furthermore, I have created a comprehensive study (based on the "Hyborg Agency" project) that integrates all the metrics together, providing a holistic view of how hybrid agents can function in interactive storytelling environments.

1.5 Summary of chapters

This manuscript is divided into a number of chapters.

Chapter 1 is the introduction, which presents the study's research questions, methodology, and this summary. Section 1.6 provides a brief overview of the studies included in this thesis. It introduces their contributions, impact, and how they link back to the research questions.

Chapter 2 reviews literature relevant to this thesis. This leads to the discussion of research opportunities and metrics for hybrid agents in Chapter 3.

Chapter 3 reviews the research approaches included in this thesis. It provides a detailed overview of the co-relations between each study in this thesis and how they contribute to answering each research question.

After that, chapter 4 to chapter 9 consists of the different studies included in this thesis that contribute to different perspectives on hybrid agents. Finally, Chapter 10 discusses the findings of this thesis and draws conclusions.

1.6 Overview of contributions and impacts

This thesis presents a comprehensive exploration into the realm of hybrid agents, specifically focusing on AI conversational agents within interactive storytelling. The thesis is comprised of several distinct studies, each employing mixed methods to explore the capabilities and effects of AI agents in storytelling from different perspectives.

This chapter provides a general overview of contributions of each study and how they contribute to the research questions. Chapter 3.5 provides a detailed explainations.

1.6.1 Pilot Studies: Engaging Narratives with AI Agents

The initial phase of the research involved pilot studies that explored audience engagement with AI-driven narratives in art projects “Wander” and “1001 Nights.” These projects demonstrated the potential of Large Language Models (LLMs) in enhancing storytelling and player engagement in narrative contexts.

- **Wander:** Focusing on the relationship between generated content and audience perception, this project utilized AI to co-create a narrative context on public platforms like Discord and WeChat. The extended paper was published at the IEEE MMSP Conference (Paper 2) and exhibited at various platforms including the ACM MM Art Gallery (Art 1) and SIGGRAPH Asia Art Gallery (Art 2). This study received over 20 citations in research around generative AI.

  By co-creating narrative context, and real-world data on public platforms, Wander created believable and meaningful interactions through the connection to real-world, addressing RQ1 and RQ3 on expanding the boundary between fiction and reality.

- **1001 Nights:** This AI-driven game, blending storytelling and interactive gameplay, explored narrative game contexts and their influence on player expectations. The project showed that integrating natural language interactions with classic stories can significantly enhance player engagement and creativity, as evidenced by positive feedback from participants. The extended paper of this project was published at the ICIDS Conference (Paper 1) and exhibited at the Foundation of Digital Games (Demo 1). This project demonstrated enacted storytelling that integrating natural language generations with classic stories can enhance player engagement and creativity, contributing to RQ1 on how fictional characters can be used for engaging interactions.

1.6.2 Catherine & David: Community-Based Storytelling Social Chatbots (SCCs)

Building on the pilot studies, this research introduced two AI-powered social chatbots in an online gaming community, focusing on the believability and engagement of fictional characters within daily human communication contexts. The study highlighted the significance of storytelling in enhancing the engagement and believability of social agents. This project was collaboratively developed with rct.ai and presented at the IMET 2023 conference (Paper 3).

This study validated the capability of storytelling in enhancing the believability and engagement of social agents within daily communication contexts, contributing to all RQs through social expansion, which integrated daily communication context like Discord channel and players community engagement into the storytelling experience.

1.6.3 Specific perspectives

Exploring Creator Perspectives: ORIBA

Focused on creator's perspective, the ORIBA project investigated the potential of AI agents in the creative process. This LLM-based chatbot was designed to stimulate visual artists' creativity by simulating conversations with their original characters. The research demonstrated that AI can significantly support artists in developing and enriching their characters.

This project was presented as a poster at UbiComp (Poster 1) and is currently under revision for CHI (Paper 5). Additionally, a joint article is underway for the Springer book "Creativity in the Digital Age" (Book Chapter 1).

ORIBA showed that as a creative medium, AI agents can support artists in developing and enriching their original characters, generalising RQ1 and 3 from the creator's perspective.

Language Perspective: AI Nüshu

This aspect of the research delved into the emergent language of AI agents, using the project "AI Nüshu" as a case study. This project explored the creation of a unique language system by AI, drawing parallels with the historical Nüshu language: the unique language exclusively created and used by Chinese ancient women. In this way, this work contributed to a computational and artistic method that create unique non-human machine language from human-AI interactions.

This work was selected for presentation at SIGGRAPH Asia 2023 (Paper 6, Art 4).

This project explored the creation of a unique language system through NLP and LLMs, contributing to RQ3 on expanding the boundary between fiction and reality through a speculative non-human machine language.

In-Game Perspective: 1001 Nights V2

After the release of State-of-the-art LLM ChatGPT in 2023, this study revisited the "1001 Nights" game, integrating LLM and multimodal AI to further blur the lines between reality and fiction in perspective inside the game. This research validated the feasibility of implementing LLM reasoning process for generative story content inside the game.

The project garnered widespread recognition within the industry, being showcased and nominated at renowned game exhibitions and awards, such as Gamescom, A MAZE, indiePlay, Lumen Prize and Now Play This. The research papers, published at ICIDS (Paper 1) and AIIDE (Paper 4), have been cited over ten times in research articles spanning the fields of AI, game studies, and interactive storytelling. I was accordingly invited to deliver expert talks at prestigious venues, including the London Game Developer Conference and the Game AI Summer School.

This project suggest that by integrating LLMs and multimodal AI with the original folklore's plots and characters, the AI agents in the game (such as the king) generate content (e.g., stories) and actions that are more believable. Furthermore, through the narrative of "1001 Nights," this game further blurs the lines between reality and fiction on both a conceptual level (game characters expanding their own world) and an interactive level (players' inputs extending the boundaries of the game). As a result, this project makes contributions to RQ2 and 3 based on storytelling.

1.6.4 Hyborg Agency: Social Expansion with AI Agents

Gathered all takeaways from the previous projects, the final project, "Hyborg Agency," focused on the social expansion potential of hybrid AI agents. This immersive online forest interconnected with a Discord community allowed for dynamic interaction with fictional AI creatures, "Hyborgs." The knowledge from human world will gradually grow Hyborgs in the fictional world in a believable way. The project was exhibited at SIGGRAPH 2023 (Art 3) and received the Lumen Prize Student Award.

As a comprehensive presentation of the concept of hybrid agent, Hyborg Agency allowed for dynamic interaction with fictional AI creatures through social, spatial and language expansion, contributing to RQ1, RQ2, and RQ3.

1.6.5 Summary

The collective contribution of these studies lies in their diverse approaches to integrating AI agents into storytelling to blur the boundary between reality and fictions. In this way, this PhD research contribute to new metrics that view AI agents as hybrid agents through perspective of creator, language, in-game and social expansion outside the games.

Public presentations of these works have been recognized at several top-tier conferences and exhibitions, highlighting their relevance and impact in academic, art and industry field.

Selection of contributions and impacts that we have made are listed below in Chpater . A detailed review of practice will be found at chapter **??** after the following sections.

Chapter 2

Literature Review

2.1 Research Scope

Approaching the potential impact of AI agents on meaningful interactions in various contexts, from both the player-game interaction perspective and the associated creator perspective, the research focus of this study lies at the intersection of several foundational fields: Human-Computer Interaction (HCI), LLM application, gaming applications of AI, media art, and interactive storytelling.

Specifically, the study concentrates on LLM-driven conversational agents and their potential to enhance the believability of characters within virtual environments and players' engagement with them. The believability of agents encompasses their social behaviour, appearance, actions, and dialogue. Our primary focus is on their social behaviour and dialogue capabilities, as well as character-driven and -influenced story generation within computational narratives. This research is thus intrinsically linked to the application of AI to gaming and providing people with meaningful experiences through gaming.

2.1.1 Fictional Worlds and Games

The scope of this research encompasses fictional worlds because these worlds provide players with meaningful experiences. According to Juul "[171]," video games are simultaneously real and fictional. The real aspect of a game is defined by the rules that govern how players interact with it, while the fictional aspect is established through the imagined world in which the game takes place. Juul differentiates between fiction and storytelling: fiction refers to any type of imagined world, whereas a story is a fixed sequence of events presented to a user.

Research on AI in gaming "[333]" suggests that AI researchers should see video games as ideal testbeds for their work, while game developers should leverage AI to create new types of games. This is because video games can serve as a useful tool for the development and testing of AI technology, and advancements in AI can lead to the creation of entirely new gaming experiences.

Some emerging technologies, although they have not yet been adopted by the industry, offer intriguing aspects for developers and researchers, creating more opportunities for game development and research. For instance, Façade "[211]" has sparked subsequent studies on

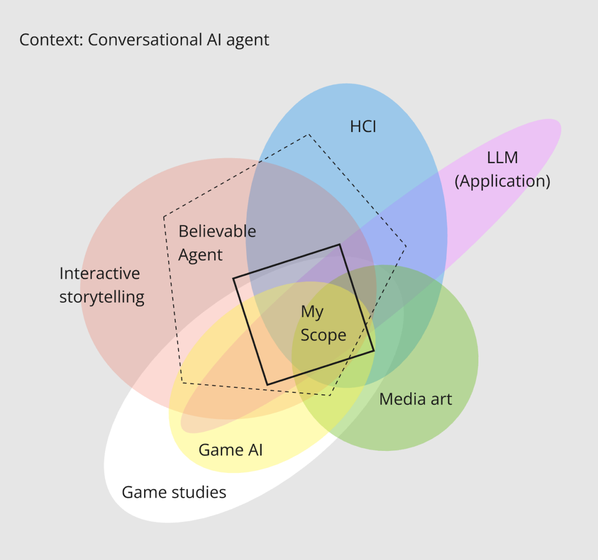


Figure 2.1.1: Scope

natural language conversations in games. I see clear potential for the advancement of LLM applications in gaming.

Due to the multimodal nature of video game experiences, the factors influencing character believability considered in related research are diverse and include factors such as character appearance and game environment. Overall, computer games can be considered a social “training ground” for NPCs [39], and natural language interaction gives players room to provide new content that can be (to an extent) acknowledged by the game.

Yannakakis et al. [374] present the connection between computational narratives and believable agents in the context of gaming applications of AI. They state that research on interactive narratives benefits from and influences the use of believable agents that interact with the player and are interwoven in the plot of a story. The computational narrative of a game defines the arena within which believable agents can be designed. In return, believable characters contribute to more believable stories and game representations. A detailed review of this subject is included in section 2.5.

Previous research has primarily utilised symbolic AI [382], employing top-down rules and systems to navigate the experience. However, recent advancements in generative AI both present opportunities and raise questions for connectionist AI, which is based on neural networks that allow bottom-up experiences to emerge without pre-defined rules. When NPCs speak lines that were not previously written by a human, how does this influence believability and engagement?

In summary, the fictional worlds created in video games provide meaningful experiences for players and serve as ideal environments for the study and development of AI, specifically in terms of computational narratives and believable agents.

The following sections will review related technology, research topics and important works, delineate opportunities for further research, and outline the conceptual framework of this thesis (in section 2.2).

2.2 Conceptual framework

This section introduces the terminology, definitions and theory on which the rest of this thesis will be based.

2.2.1 Actor Network Theory (ANT)

Actor Network Theory (ANT) [181] is a theoretical framework that originated in the field of science and technology studies (STS) and has since been applied to various other fields such as sociology, business, and computer science. ANT views the world as a network of human and non-human actors, and emphasises the importance of studying the relationships and interactions between these actors. The theory proposes that actors are not fixed entities but are constantly shaped and transformed by their relationships with other actors. According to this approach, all entities which network with one another are on par: a human, an animal, a computer, a toaster, a tree, or a university all have the ability to shape social processes through their interactions [231]. Such networks are not formed only by human actors; instead, they are populated by an array of human and non-human participants in a chain of associations. In ANT, the term "actor" is a semiotic definition — it refers to an entity that is capable of acting or to which activity is attributed by others. An actor can be anything that is recognised as capable of producing an effect [181].

This thesis adopts ANT as a core analytical lens, extending its application to LLMs. As recent research suggests, an LLM can be regarded as a non-human actor, where "the continuous interaction between an AI system and its users has the potential to shape the nature of the information transmitted, affecting not just the platform itself but also the wider network of which it is a part" [131]. This approach, which understands the dynamic relationships between human and non-human entities as a coherent network, is particularly relevant to the study of hybrid agents. For instance, in the *Catherine & David* project (Chapter 6), the AI agent Catherine is not merely a passive tool executing pre-defined scripts. Instead, she functions as an autonomous actor within a complex network comprising developers, the player community, the Discord platform, and the fictional narrative itself. Through her emergent behaviors—such as generating unexpected plot points or demonstrating disobedience—Catherine actively "translates" [181] the interactions, influencing community discourse and compelling the developers to adapt the narrative arc. The resulting storytelling dynamic is therefore a co-creation of this entire network, rather than the product of a single authorial voice.

This perspective challenges the traditional hierarchical view of human creators and AI tools. Instead of asking what an AI agent *is*, ANT prompts us to ask what it *does* within its network of relations. My interest lies in examining these networks where AI agents act as boundary objects, mediating between the fictional and the real. I am particularly intrigued by the interactive storytelling that emerges when humans not only create stories with agents "inside" the game but also interact with them "outside" it, collectively constructing dynamic social relationships that blur the lines between player, creator, and character. Thus, this multi-layered complexity must be considered when examining the role of AI agents in fictional storytelling.

For example, some studies have used ANT to understand the social behaviour of players in online multiplayer games towards other players[151].

2.2.2 Magic circle

Researchers have extensively discussed the distinction between "internal" and "external" worlds in games and other interactive narratives, based on the concept of the "magic circle."

The magic circle is a game design concept that delineates the boundary between the game world and the real world. It refers to the space where the game's rules and objectives are applicable, and beyond which they are not. Coined by Johan Huizinga [150], this term is extensively utilised in game studies, design, and research. Scholars have expanded this concept to describe the creation of a special place in time and space by a game, a realm which is distinct and separate from the real world [328]. The magic circle is a conceptual rather than a physical boundary, and is entered voluntarily by players during gameplay. For example, when players engage with a game like Super Mario, they accept its unrealistic elements without question: the protagonist survives improbable falls, turtles can fly, and consuming mushrooms causes Mario to double in size. Jesper Juul further developed this idea, proposing that games are "half-real" because they involve real rules while simultaneously engaging players in a fictional world [171]. While initially intended to apply to games, this concept is also more widely relevant and is discussed in reference to other interactive mediums, such as theatre and film [265], through which humans perceive virtual entities within fictional contexts.

The relevance of the magic circle concept extends to the discussion of how people view and interact with AI agents, though this paper is not primarily focused on game studies. As a conceptual boundary, the magic circle in the context of digital games and interactive narratives, including interactive theatre projects like "Sleep No More" [265], is maintained by a dynamic power balance among game designers, players, and systems.

Developers can enforce strict rules in these settings to prevent players from breaking the magic circle, for example by confining players to specific areas or prohibiting actions that would disrupt cooperative gameplay. These restrictions are designed to shape the player experience, ensuring it remains meaningful and engaging. The relationship between players, the game system, and developers creates a power balance that preserves this dynamic boundary.

However, when traditional game characters are replaced with AI agents, the dynamics within the magic circle change significantly. Scholars have long moved beyond viewing the magic circle as a rigid, sealed boundary. Instead, it is more productively conceptualized as a "permeable membrane" [237], one that is constantly subject to negotiation and is permeated by external contexts such as markets, politics, and law [272]. This critique has been taken further by scholars like Consalvo, who argue that in an increasingly connected world, "therethere is no magic circle," as real-world rules and socio-cultural contexts are never fully suspended [77]. This thesis builds upon these nuanced perspectives, positing that hybrid agents, by their very design, operate within this contested, porous boundary, actively mediating between the fictional and the real.

The introduction of generative AI agents exacerbates this porosity and challenges the boundary's stability in two fundamental ways. First, as you astutely point out, the opacity of AI technology contributes to a public knowledge asymmetry. When AI agents like GPT [249] seem to display consciousness or emotions, distinguishing between a scripted performance and a genuine emergent experience becomes challenging for the user, thus blurring the entry and exit points of the magic circle. Second, the increased autonomy of both the AI agent and the human participant can render the magic circle fragile. As observed in Google's Werewolf game composed of AI agents [118], it was challenging to prevent participants from intentionally disrupting the game's progress, leading to the dissolution of the magic circle.

It can be argued that generative AI inherently grants AI agents the capacity to either break or expand the magic circle. As players gain more freedom, the challenge of controlling the boundaries of the magic circle increases, and the consensus-based boundary grows less stable due to the stochastic nature of generative AI. This thesis posits that understanding the concept of a dynamic boundary is crucial for comprehending the role of AI agents in the creative process. Therefore, rather than asking if the boundary exists, this research asks how it is expanded and redefined by hybrid agents, a question that will be systematically addressed through the analytical framework developed in Chapter 3. Moreover, human creators should remain cautious of the magic circle in AI-human interactions and, where possible, strive to maintain transparency and control over this boundary, given their responsibility toward the narratives shaped by AI.

2.2.3 Pervasive games

Researchers have introduced the concept of "pervasive gaming" [228]: a genre of gaming that systematically blurs and breaks the traditional boundaries of the game. The spatial, temporal, and social dimensions of the limits of the magic circle are thus explored. "Pervasive" games must have one or more salient features that expand the contractual magic circle of play socially, spatially, or temporally. This allows the expansion of the different dimensions of the magic circle to be measured systematically:

- Spatial: Spatial expansion can be produced if the socially constructed location of the game is unclear or unlimited. Examples include various alternate reality games that require players to take offline actions (such as finding hidden envelopes to obtain clues).
- Temporal: Pervasive games expand temporally from explicit play sessions; the socially constructed game session is interwoven and mixed with ordinary life. Examples include "Lifeline" [7], a game based on a text message interface simulating communication with a stranded astronaut which requires the player to wait in real time for the virtual character to complete instructions.
- Social: Perhaps the most controversial and potentially most bountiful type of expansion occurs when a game expands socially, obfuscating the boundary of playership. In the unexpected places and times where the expanded games are played, unexpected people

make a difference in gameplay. An example is “A Week With Wanda” [134], in which characters email players to inquire about their next choices.

It should be noted that even non-pervasive games do not exist in complete isolation — every game is influenced by elements outside the magic circle.

Researchers have paid particular attention to the concept of alternate reality games (ARGs) [325], which are considered pervasive games. ARGs use expansion to create the illusion that the game is not a game, despite clearly disclosing that they are indeed games. They often blend real-world and digital elements to produce a game that seems to be integrated into players’ everyday reality, aiming to make the player feel less like they are playing a traditional game and more like they are in an alternate reality. These games use the three dimensions of expansion described above to create the illusion that they are not games — even though they actually disclose the fact that they are games. A famous example is the “I Love Bees” promotional campaign for Halo 2 [136], which comprised both web-based promotion and real-world advertisements. Players could track websites “hacked by a mysterious intelligence” as part of the Halo 2 story. The other side of this illusion of games not being games is that players enjoy the illusion of not being players. This social expansion of the magic circle can include blurring game mechanics and increasing the ambiguity of the game interface. The concept of pervasive gaming offers opportunities: do players need to know they are playing? Are non-players (those outside the magic circle) who are unaware that some of the activities of their daily lives (like seeing an advertisement) are part of a game, still players?

This adaptability positions AI agents as actors that inhabit, blur, and expand the magic circle. A review of the relevant literature reveals that the adaptability and generative capability of AI agents has led to them progressively breaking the boundaries between “human-set” and “naturally generated”, influencing human productivity and social relations. For example, powered by a LLM, a witch from a video game can also act as a voice assistant through a smart speaker, with a believable worldview, behavior, motivation, and consistent memory. The concept of the magic circle provides a reference for the boundaries of AI agents’ fictional identities in various environments (such as virtual friends in Replika [273] or virtual characters in games). Therefore, although this study focuses not only on games but more broadly on interactive narrative-based media, the discussions around the magic circle in this chapter are crucial and form the foundation of this thesis.

The theoretical framework of pervasive gaming is therefore not merely a related concept, but a foundational pillar for this thesis. The core mechanism of pervasive games—blurring the boundary of the magic circle—directly parallels the behavior of the hybrid agents under investigation. This thesis argues that hybrid agents, powered by modern LLMs, represent a new frontier of pervasive experiences, extending the principles of pervasive gaming beyond player-centric activities into the realm of autonomous, character-driven narrative interactions. Crucially, the analytical lens provided by pervasive game theory offers a structured way to deconstruct and measure this boundary-blurring phenomenon. The three dimensions of expansion—spatial, social, and temporal—as defined by Montola [228] provide the direct theo-

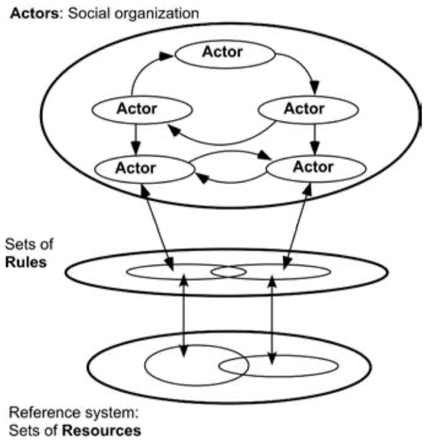


Figure 2.2. A generic model of games and simulations

Figure 2.2.1: Image by J. Klabbers

retical inspiration for the metrics developed in Chapter 3 to evaluate the capabilities of hybrid agents. By applying this framework, we can move beyond a general description of "blurring boundaries" to a systematic analysis of how and in what ways hybrid agents operate on the edge of the magic circle. This allows us to understand the dynamics observed in practical case studies, such as an AI agent moving from a game world into a community Discord server (Spatial Expansion) or integrating real-world social relationships into its narrative generation (Social Expansion) in the case studies. Therefore, the concept of pervasive gaming serves as the critical bridge connecting established game theory to the novel contributions of this research.

2.3 AI agents

Artificial intelligence (AI) is a field of computer science that focuses on creating intelligent machines that act and react like humans. Machine learning is a subset of AI that involves developing algorithms that can learn from data and make predictions or decisions based on it. Generative AI is a certain type of machine learning which algorithmically generates new data, insights, or content from existing data. A language model is a type of generative AI model that is trained to predict the next word in a sequence of words. Large language models (LLMs) are language models that have been trained on massive amounts of data and use a large number of parameters, allowing them to generate human-like text and perform other language-related tasks, including performing as consultants, assistants, or emotional companions [340, 363, 53].

In recent years, the terms "conversational AI agent" and "chatbot" have often been used as synonyms. However, previous review papers suggest that the term "chatbot" tends to be used to refer to text-based programs that are not embodied or multi-modal, while the term "conversational AI agent" tends to be used to refer to programs which use an avatar or those that use multiple modalities, such as voice-based assistants like Siri [62, 380].

When developing AI agents, designers frequently consider the socially acceptable norms of conversational interactions [256]. This produces "social chatbots" or "social agents" which exhibit social intelligence: the capability of an entity to exhibit appropriate social behaviours to achieve specific objectives [30].

Social chatbots are defined by Pentina et al. [256] as programs or applications that employ AI and natural language processing technologies to converse naturally and intelligently with users via voice, text, and/or images. Brandtzaeg et al. [41] further describe social chatbots as AI dialogue systems that can engage in social and empathetic conversations with users. Their human-like conversational capabilities render them suitable as conversational companions, friends, or even potential romantic partners. Such human-AI interactions introduce a novel intimate bond with technology; Licklider [192] envisions this bond potentially reshaping our understanding and expectations of human-human relationships. It is crucial to distinguish these definitions of "social chatbots" from the use of the term "social bots" to denote chatbots which generate content for social networks such as Twitter [337].

The history of the term "**agent**" can be traced back to the philosophies of thinkers like Aristotle and Hume[294], in which it denotes an entity possessing autonomy and the capacity to act. Under this definition both humans and non-human (for instance animals) can both be considered agents in different contexts. Within the realm of AI, the term "AI agent" pertains to artificial entities that can perceive their environment, make informed decisions, and act accordingly. AI agents can be further categorised. For instance, the term "autonomous agents" generally refers to AI agents designed to function independently without requiring direct human intervention. Recently, advanced LLMs like GPT-3 have shown potential as a foundation for creating more sophisticated and autonomous agents. The ability to integrate an LLM as the brain or controller of an agent has led to the emergence of LLM-based AI agents.

This study primarily centers on LLM agents that are proficient in engaging in natural language conversations with humans. In certain contexts, such as text-based platforms like Discord, these entities are often termed "chatbots". Due to the capability of LLMs to mimic and comprehend human behaviours, LLM-based AI agents typically exhibit some degree of social proficiency. Hence, in many situations, they can be regarded as social agents; for example, when functioning as a psychological consultant. Within the scope of this research, the more mainstream term "AI agent" will primarily be used, while also using the terms "chatbot" and "social agent" interchangeably.

The critical properties of AI agents within the scope of this research are:

- **Social ability [105]**, which enables the agent to communicate with human creators and collaborate as co-creators, and to role-play as certain characters in a believable manner;
- **Proactiveness and autonomy**, which allows the agent to take the initiative in the creative process as a co-creator, and role-play as fictional characters without direct instructions from its creator.

We thus define "AI agents" as computational entities which act autonomously with a significant level of proactiveness and social ability.

Previous research by Shanahan [305] suggests use of the term “role-play” to avoid the pitfall of overestimating the capabilities of LLMs due to anthropomorphism. This approach recognizes the public’s inclination toward an anthropomorphized view of LLM agents due to their capacity to imitate human behavior. According to this perspective, an LLM can be considered a non-deterministic simulator capable of role-playing an infinite variety of characters or generating an infinite number of simulacra [306]. In this article, we view the initiation of AI agents to play specific roles as part of a dynamic creative process, which includes but is not limited to creating Twitter or Discord bots and designing game characters. This role-play can be persona-driven (e.g. setting up an AI agent to imitate Socrates) or goal-oriented (e.g. using GPT to a pixel-art drawing assistant), but in both cases AI agents possess a degree of autonomy and proactiveness.

2.3.1 LLM-based AI agents

Recent research [363] has presented a comprehensive conceptual framework for LLM-based agents, delineating their three primary components: their brain, perceptions, and actions. Within this model, an AI agent’s “brain” is primarily constituted by an LLM.

LLMs like GPT-4 and PALM have evolved beyond text generation and are capable of engaging in complex reasoning and interactions [44, 65, 249]. Chain-of-thought (CoT) reasoning allows LLMs to process information step-by-step, improving accuracy in tasks like math problems [351]. Advanced methods introduce non-linear reasoning, enhancing this process [344, 377, 137].

This progress has spurred the development of AI agents for multi-step tasks involving external environments or human interaction [341, 355, 319]. The ReAct framework, demonstrated in LangChain, exemplifies this: agents perform sequential reasoning and action steps using external application programming interfaces (APIs) for tasks like web searches or database queries [378, 153]. Subsequent projects like AutoGPT and BabyAGI focus on long-term planning and task decomposition, which is relevant for artists working with original characters (OC) [130, 234]. Additionally, CAMEL and generative agents facilitate multi-agent interactions in simulated environments [51, 255].

Recent surveys have foregrounded discussion of LLM-driven AI agents for the following reasons [341, 355, 319]:

1. Rich and dynamic content generation: LLMs are equipped to craft a myriad of content that is both contextually apt and innovative. This capacity augments the depth and allure of the interaction between agents and users, ushering in a heightened level of engagement and immersion.

2. Enhanced language comprehension: A salient feature of LLMs is their profound grasp of human language. This nuanced understanding facilitates the meticulous deciphering of user inputs, enabling the generation of accurate and contextually meaningful responses.

3. Autonomy and adaptability: LLM-driven agents exhibit a marked degree of operational independence and adaptability. They showcase their autonomy by crafting human-esque

text, conducting engaging conversations, and executing a range of tasks without the need for granular, step-by-step directives [378]. Moreover, their intrinsic capacity to dynamically modulate their outputs based on environmental cues underscores their adaptive autonomy.

4. Social proficiency: This pertains to an agent's capability to engage with other agents, including interacting with humans, via a designated agent-communication language. LLMs manifest formidable natural language interaction competencies, encompassing both the assimilation and generation of language.

In essence, the confluence of these attributes reinforces the argument for the pivotal role of LLMs as the foundational core of contemporary AI agents [363].

In psychology, they assist in both simulation experiments and mental health support, demonstrating their adaptability and ability to understand human behaviour [376]. LLM-based agents like ChatGPT demonstrate potential as AI assistants that can understand requests, retrieve knowledge, and provide tailored responses to user needs. These agents have proven valuable in political science and economics, aiding in tasks like ideology detection and economic behaviour simulation [258, 204].

Crucially, such agents have a profound impact on social simulation. Traditional societal experiments often pose financial and ethical challenges. LLMs mitigate these concerns by allowing for in-depth virtual societal simulations and closely mirroring real-world actions [383]. This capability resonates with the worlds of interactive storytelling and gaming, wherein replicating realistic social dynamics is pivotal for the creation of immersive experiences. The most famous case of this is Generative Agent [255], which simulates human daily life in a town consisting of AI agents.

For the purpose of this thesis, such social simulations are considered a related but distinct framework from traditional games. While both involve rules and interaction, traditional game simulations are often centred on explicit player goals and win-states, whereas these LLM-driven social simulations prioritise the study of emergent, open-ended social behaviours among autonomous agents.

2.3.2 Evaluation of AI agents

It is essential to evaluate AI agents appropriately, especially as they continue to grow in capability and find further applications in real-world scenarios. Recent surveys [363] delineate four crucial dimensions of AI agent evaluation:

1. Utility: Assessment of the AI agent's effectiveness in completing tasks, typically through the application of metrics such as success rates.
2. Values: The alignment of AI agents with human ethics and values is a vital concern. This dimension evaluates the agent's honesty, non-harmfulness, and fairness in its interactions.

3. Evolvability: This pertains to the ability of AI agents to continuously evolve and refine their functionalities. The focus is on their ability to engage in lifelong learning and adaptive improvement.
4. Sociability: An agent's capacity for efficient and meaningful communication significantly shapes user experiences.

Sociability entails:

- Language communication proficiency: A foundational skill that encompasses both natural language understanding and generation capabilities.
- Cooperation and negotiation abilities: These qualities ensure agents operate effectively in various scenarios, both ordered and unordered [51, 94].
- Role-playing capability: It is essential for agents to genuinely reflect their assigned roles by making statements and undertaking actions congruent with their specified identities [291]. This capacity ensures distinct role demarcation during interactions, be they with other AI agents or humans. Moreover, it is essential that the AI agent maintains this identity, especially when participating in extended tasks, to prevent ambiguity and confusion [51, 254].

Sociability gains special prominence when considering AI agents tailored for social applications, such as conversational aides or collaborative systems. The role-playing aspect, in particular, mirrors the intricacies of real-life interactions, making it indispensable for crafting lifelike and engaging experiences. This thesis focuses on this aspect in particular.

2.4 Stories and interactive narratives produced by AI

The initial motivation for this research was the belief that virtual worlds and their fictional content are meaningful. Hence, the goal of this research is to tell stories through believable agents. This section will discuss the connection between storytelling (narrative) and AI, as well as how existing research focuses on games as the context for studying believable agents and stories. It is important to note that this research does not specifically discuss gamification [300] as this is not the primary focus. Games serve as a suitable contexts and playgrounds for discussing believable agents and stories, and this practice-led study also presents game-like characteristics. However, discussing whether it is a game or why it is a game is not the main focus of the research.

This research focuses on characters because the creation of believable agents is one of the challenges confronted by research on game AI. How believable characters can be created in interactive narratives is still an open question. The various studies on believability aim to make AI agents more human-like in their behaviour with the objective of using these agents to convey stories.

2.4.1 Why stories matter in terms of AI

Stories play a crucial role in human cognition and experience. According to Ricoeur [277, 278], narrative is a way of organizing our experiences into a coherent whole that gives meaning to our lives. He argues that we use narrative to construct our sense of self and understand the world around us. In his view, narrative is not just telling stories, but also a way of making sense of our experiences and creating meaning in our lives.

Related research [263] states that the ability to tell, craft, and understand stories has long been considered a hallmark of human intelligence. AI is rapidly shaping the future of various aspects of our lives, particularly in the spheres of entertainment (television, film, gaming, and general interactive and non-interactive means of storytelling), training, and education.

Riedl et al. [280] argue that storytelling is an integral part of many modern computer games; within games, stories create context, motivate the player, and move the action forward. Interactive narratives can be produced by using AI to create and manage stories within games, creating the perception that the player is a character in a dynamically unfolding and responsive story.

The importance of stories in serious games is discussed by [235], who highlight the importance of narrative for engagement and motivation. They also mention immersion; however, this primarily refers to transportation to a virtual environment, which may not be directly applicable to this research. The authors emphasise that narrative structure is a pervasive part of human cognition and is the means by which humans frame and recount their daily experiences.

2.4.2 Interactive narratives

Interactive narratives are forms of digital interactive experience in which users create or influence a dramatic storyline through their actions [280]. The goal of an interactive narrative system is to immerse users in a virtual world such that they believe they are an integral part of an unfolding story and that their actions can significantly alter the direction or outcome of the story [282].

The core research challenge in interactive narrative lies in balancing authorial intent and player agency within the context of storytelling. One approach to addressing this challenge involves the use of virtual character autonomy, which focuses on the degree to which computer-controlled entities can act independently of the experience manager [280]. This level of autonomy has implications for character believability. Park et al. [254] discuss how LLMs bring new opportunities to the experience manager architecture.

2.4.3 Narrative intelligence and natural language processing

Narrative intelligence is the ability to craft, tell, understand, and respond affectively to stories [281]. Research on computational narrative intelligence seeks to instill narrative intelligence into computers to make them better communicators, educators, entertainers, and more capable of genuinely understanding human needs. This field is as much about human-computer interaction as it is about solving hard artificial intelligence problems.

According to Ryan [290], the theory of possible worlds provides a more nuanced way of discussing the notion of a fictional "world", while AI contributes to narratology and the theory of fiction directly via research into the cognitive processes involved in text and automatic story generation. Narrative intelligence is one of the abilities that sets humans apart from other animals and non-human-like AIs [281].

The importance of narrative intelligence in various aspects of our lives, particularly in entertainment, training, and education, is highlighted by Poola et al. [263]. They argue that the ability to tell, craft, and understand stories has long been considered a hallmark of human intelligence.

Previous studies have investigated the use of natural language processing (NLP) for many different applications, including creative tools [323, 93, 73]. Some projects have developed collaborative AI writers which focus on specific genres, for example Shelly [370], a crowd-sourced horror writer.

Similar approaches have been used for content generation in academic research. Murder mysteries have been generated for adventure games [28] using structured information about real people scraped from Wikipedia articles. Designing for Narrative Influence [269] trained a language model to generate micro-fiction that promotes sustainable public health guidelines. Martin et al. [259] presented a series of experiments that connected ancient divination techniques to modern technologies like language generation models.

Other studies have applied NLP to dialogue systems. Scheherazade's Tavern [14] and Prom Week [215] attempted to develop deeper NPC interactions for a natural social simulation experience. Talk to Ghost [157] improved high school students' interest in reading by turning Shakespeare's stories into interactive conversations with virtual characters.

CharacterChat [295] and BanterBot [128] are dialogue systems that allow writers to talk to characters they have created. This AI assistance allows writers to interact with their characters in a familiar social setting.

However, prior to the breakthrough of AI models like ChatGPT[250] in 2023, much of narrative intelligence was based on connectionism, relying on top-down rule control. Although some research utilized neural networks to facilitate bottom-up control, such as [205], the limitations of generative AI at the time prevented the development of games that could support a broad player base and offer rich experiences. In the latter half of this chapter, section 2.6, I will further analyse the evolution of narrative intelligence applications.

2.4.4 Different types of narrative

The narrative of a game is essentially the story told as the player progresses through the game. It is delivered in various ways and by various sources. Jenkins[163] divides the narrative into two parts: the embedded and the emergent narratives. The embedded narratives are narratives that are part of the designed game experience; for example, quest lines or in-game environments. The emergent narrative is the part of the narrative that comes into being as the player plays the game. Examples of this would be NPCs in the game changing their behaviour and characteristics over time due to players' participation, as they do in the Sims[99].

Calleja [49] further discusses the types of narratives present in interactive storytelling, and proposes that we differentiate what he calls the scripted narrative (written by designers) from the "alterbiography" of the player. In essence, the alterbiography is the story of how the player has played the game, as described by the player themselves. In this sense, the alterbiography focuses on a subjective sense of human experience that the creator behind the story cannot fully anticipate or control.

This suggests potential applications for AI agents, since they can adaptively generate content which is not pre-scripted. This means that while they provide enormous space for alterbiographical narratives, human designers need to consider how to confine and moderate game design at the scripted narrative level to guarantee quality experiences.

2.5 Believable agents

Believable characters have always been important within the field of interactive narrative. The terms "believable character" and "believable agent" are often used interchangeably. While believable agent may be used in a broader context, referring for instance to real-world voice assistants, believable characters typically appear in interactive narratives, especially in games. The study of virtual characters does not necessarily focus on believability; it may also examine appearance, voice synthesis, decision-making, and other aspects. However, these aspects are often related to believability. In this research, I use the terms "believable agent" or "believable character" according to the specific context.

2.5.1 Definitions of believability

Believability is a concept that originates from character art and video games. A believable character is one which seems lifelike, the actions of which make sense, and which allows the audience to suspend disbelief [208, 347]. Believability is defined as the extent to which users interacting with an agent come to believe that they are observing a sentient being with its own beliefs, desires, and personality [34]. Previous research has stated that a believable character is not necessarily a real character but must be real in the context of its environment: the believability of a character does not necessarily equate to its realism[352]. For example, Mickey Mouse is a believable character but not a realistic one: although he embodies consistent personality traits and behaviours within the various fictional works (films, TV series, etc.) he inhabits, he has an unrealistic cartoon appearance.

In the context of interactive narratives, especially in video games, a believable character is a virtual agent that fosters a suspension of disbelief which allows the user to perceive themselves as interacting with a personality-rich, intelligent being [282]. The believability of a character is closely connected to their ability to relate to their environment during a conversation [154]. The perception of believability is influenced by various factors, including a character's appearance, animation, personality, emotions, desires, and intentions [280]. In general, characters that can adapt to changes in their environment and exist in the correct social context are perceived as more believable because they give rise to social immersion, which "arises when the social interaction with other entities in the world reaches such a level

Table 2.4.1: A Synthesis of Believability Metrics and Methods from Literature

Title	**Author and Year**	**Core Believability Dimensions**	**Measurement Method**
An Oz-Centric Review of Interactive Drama and Believable Agents	Mateas 1999 [208]	Personality, Emotion, Self-motivation, Change, Social relationships, Illusion of life	**Theoretical Framework Definition**: Defined core concepts by synthesizing character arts and AI research.
The Non-Player Character - Exploring the believability...	Warpefelt 2016 [348]	Continuation of Loyall's (1997) framework (Personality, Emotion, Social capabilities, etc.)	**Qualitative/Interpretive Research**: Primarily used *structured observation* of gameplay videos and *thematic analysis* of open-ended survey questions.
What makes virtual agents believable?	Bogdanovych et al. 2016 [34]	Personality, Emotion, Self-motivation, Social-relationship, Consistent of expression	**Mixed-Methods**: Primarily quantitative (Likert scale, Believability Index) supplemented with qualitative *open-ended feedback* to explain the quantitative results.
What do you mean by believable characters?	Lee & Heeter 2012 [187]	Personality, Appearance, Behavior, Goal, Emotion	**Online Survey**: Showed participants NPC video clips and used a 5-point Likert scale for data collection and statistical analysis.
Enhancing the Believability of Embodied Conversational Agents through Environment-, Self- and Interaction-Awareness	Ijaz et al. 2011 [154]	Environment, Self-awareness, Interaction awareness	**Interactive User Study**: Participants engaged in real-time conversations with agents, assessed via a Likert scale and supplemented with *specific textual feedback*.
Metrics for Character Believability in Interactive Narrative	Gomes et al. 2013 [125]	Awareness, Behavior understandability, Personality, Visual impact, Predictability, Behavior coherence, Change with experience, Social	**Methodology Proposal**: Proposed an evaluative framework based on Likert scales and multiple-choice questions; did not conduct an empirical study.
Agents That Relate: Improving the Social Believability of Non-Player Characters in Role-Playing Games	Afonso & Prada 2008 [8]	Relationship, Personality, Theory of Mind	**User Evaluation**: A/B testing where players experienced two game versions (with/without a social model) and completed a questionnaire.

of fidelity that the player begins to feel a sense of togetherness with the other entities" [168]. However, it must be made clear that although all game characters can be regarded as agents, not all agents are game characters. Conversational ability is only one component of believability. Related research on the conversational ability of game characters argues that the believability of such agents is closely connected to their ability to relate to the environment during a conversation. This ability, defined as awareness believability, is formalised in terms of three components: environmental awareness, self-awareness, and interaction awareness[154]. However, although previous research on game character believability has largely dealt with in-game physical representations (e.g. the three-dimensional environment, or characters' appearances), agents can also be believable when they are not circumscribed inside a game played by a human. In a more general sense, believability is applicable for simulations where AI agents perform and interact with other agents, including both humans and non-humans (machines, animals, etc.). In recent research on AI agents, the assessment of believability is highly dependent on the research context. For instance, in a study involving generative agents [254] believability was evaluated through interviews with agents after a day-long simulation within a virtual town. Human participants determined a believability score based on how convincing the interview text was. Interestingly, some recent studies have begun utilising LLMs themselves to rate believability. In cases like CharacterLLM [307, 343], ChatGPT is used to assess whether a conversation generated by an agent closely resembles that which might be produced by a specific character, such as Harry Potter. It is crucial to recognise that the content for evaluation extends beyond scripted narratives. Although the agent acts according to the researcher's setup (e.g. by playing the role of a baker in a virtual town), the behaviours it exhibits, particularly the conversations it generates, are not pre-scripted.

The advent of LLM-driven chatbots has fundamentally reshaped the discussion on what determines believability. Traditionally, in symbolic AI systems like those in *Façade* [211], believability was primarily a product of top-down, **strong narrative structures** crafted by human authors. A character was believable because its goals, personality, and dialogue were meticulously designed and scripted.

In contrast, modern LLM agents derive a significant portion of their believability from bottom-up, emergent properties rooted in their technical foundation. This includes the vastness of their **associated training sets**, which allows them to mimic the nuances of human conversation, and the power of their **sophisticated algorithms** (e.g., the Transformer architecture), which enables them to generate contextually relevant and seemingly intentional responses.

This thesis argues that for hybrid agents, believability is not a choice between these two sources, but rather a dynamic result based on authorial intent with considerate designs. It emerges from the fusion of top-down authorial intent (the creator defining the character's narrative framework and prompts) and the bottom-up, generative performance of the AI.

2.5.2 Social behaviour

The taxonomy proposed by researchers [23] categorises agent interactions based on their social behaviours, inter-agent communication, knowledge flow, and the change in their relationships. These categories include communication, flow of knowledge, relationship, and emotions. This research focuses on the aspect of social behaviour, specifically the ability of characters to engage in meaningful conversations.

Social behaviour in believable agents is essential for fostering deeper social and emotional interactions with users. Creating such interactions can facilitate trust and partnership between users and agents [195]. For instance, Bickmore and Picard [29] introduce the concept of relational agents, which are computational artifacts designed to establish and maintain long-term social-emotional relationships with their users. They found that relational agents were respected more, liked more, and trusted more than task-oriented agents which lacked any deliberate social-emotional or relationship-building skills.

To some extent, developing social abilities in AI agents requires the mimicking of human behaviour. Research on NPCs' social behaviour in games addresses emotional attachment [100, 36, 39], empathy [63], and identity [91].

The core skill of AI agents in these emerging programs is their language ability. Unlike traditional rule-driven NPCs or those powered by drama management systems like *Facade* [212], they can generate dynamic responses to different people to form truly diverse content which meets the needs of different people, and are no longer limited by the content pre-prepared by developers. Powered by LLMs, AI agents can already act as intimate companions [**?**], complete livestreamed survival shows [121], perform in generative talk shows [177], and even form a simulated society with humans, with the ability to talk to humans in open conversations [221].

Emotions also play a crucial role in the believability of virtual characters. Research on characters' emotions has included studies on emotional attachment [101, 35], empathy [62], and identity [91]. Emotionally engaging characters can lead to better player experiences and foster deeper connections between users and agents.

2.6 Practice Review on Related AI Technologies

To understand the evolution of practice in AI-driven interactive narrative, it is essential first to examine the foundational technological shifts that made these new forms of storytelling possible. The history of AI in this domain can be broadly understood as a transition from traditional symbolic systems to modern generative models. This section will first outline this technical evolution, including the role of the Transformer architecture and the inherent limitations of current models, before proceeding to a chronological review of key practical works.

2.6.1 From Machine Learning to Generative AI

The emergence of hybrid agents is intrinsically linked to advancements in machine learning, particularly the development of Large Language Models (LLMs). While machine learning encompasses a broad range of algorithms that learn from data, the paradigm shift enabling modern generative AI is rooted in deep learning and, more specifically, the Transformer architecture [336]. Introduced in 2017, the Transformer's key innovation is the self-attention mechanism, which allows the model to weigh the importance of different words in an input sequence. This capability provides a superior understanding of context and long-range dependencies in language compared to previous architectures like Recurrent Neural Networks (RNNs). LLMs, such as the Generative Pre-trained Transformer (GPT) series [129, 249], are essentially scaled-up versions of the Transformer architecture, trained on vast quantities of text data. It is this underlying technology that endows them with the ability to generate coherent, contextually relevant, and stylistically diverse narratives, moving beyond pre-scripted responses to become generative co-creators in interactive storytelling.

The GPT (Generative Pre-trained Transformer) series represents a significant leap in natural language processing and generation capabilities. GPT-2, released in 2019, marked a turning point in AI's ability to generate coherent and contextually relevant text [129]. ChatGPT, launched in late 2022, further revolutionised AI interactions by enabling more natural, conversational exchanges [250]. These milestones not only transformed the AI landscape but also significantly impacted the development of interactive narratives and believable AI agents in games and other applications. By structuring this review around these technological breakthroughs, the evolution of AI's role in interactive storytelling can be better understood, and how each advancement opened new possibilities for creating more engaging and believable experiences can be examined.

This section revolves around several previously discussed technologies and issues, reviewing significant practices in AI applications, games, and art. The review is divided into three periods based on key milestones in technological development: Before GPT-2 (1970-2019)[267], After GPT-2 but before ChatGPT (2019-2023), and After ChatGPT [250] (2023-present).

2.6.2 Stage I: Before GPT-2 (1970-2019)

Prior to the maturation and public availability of neural network-based generative text models like GPT-2, numerous attempts had already been made to produce conversational AI. The earliest was ELIZA [353], released in 1966. Despite its simple system (repeating and querying user input based on grammatical structure), its novelty still led people to mistakenly believe ELIZA was a real person. This illusion, known as the Eliza Effect [354], was due to the human tendency to perceive entities capable of conversation as human-like. Similarly, in the 1970s the Tale Spin machine [218] utilised mechanical structures to compose stories with very limited effects and content, leading to use of the term "Tale Spin effect" to describe meaningless or repetitive content generated by machines [346].

In terms of interactive narratives, the 1980s saw the emergence of many text adventure games: interactive virtual worlds where users observe and act using words instead of pixels. They have a long history, with many of the earliest games — such as the iconic Zork [31] — being text-based. These games are turn-based, whereby each turn involves the user receiving an observation about the environment (e.g., "You are standing in an open field west of a house, with a boarded front door. There is a small mailbox here") and responding with an action (e.g., "open mailbox"). This continues for a number of steps until the game concludes. However, such games only supported commands and content that were pre-set by developers and did not involve AI models. This era's interactive narratives can be summarised as consisting of a processor and a world model [227], where the processor interprets player inputs (e.g., breaking down "go west" into a "go" command and a "west" direction), and the world model is represented by the text content prepared by the author to depict the fictional world.

Post-2000, with the maturation of computer graphics technology, more scholars introduced conversational systems into games and related research. For example, the *Lab Lab Lab* team developed a series of games based on natural language input in which players must gain the trust of game characters through dialogue: playing, for instance, Snow White's stepmother to convince her to eat the poisoned apple [189]. Another notable example is Facade [210], which realizes a complete three-dimensional interactive drama. Players interact with a quarreling couple using natural language input. The game smoothly progresses the plot by combining player input, the location of characters in the scene, and the current storyline using a drama manager.

In the commercial realm, a few games have experimented further with natural language dialogue. For instance, in Event[0] [5] users play astronauts on a stranded spaceship, seeking to escape by conversing with an AI system. Bot Colony [4] implemented player control over game characters using speech-to-text technology, but was limited by the NLP technology of the time. Notably, these games did not incorporate neural-network-based Natural Language Generation (NLG) methods, as no language models were available. They used basic Natural Language Understanding (NLU) methods to map player inputs into pre-scripted systems, meaning every line of dialogue (or its components) still relied on the work of human authors. Mentioning modern technology to Snow White, for instance, would lead to an "unknown" branch of the conversational tree, with her saying: "I don't understand, what are you talking about?"

Text adventure games of the 1990s became a research paradigm for AI scholars after 2010 because playing these games requires common sense and an ability to understand language, making them a useful training ground for AI. Important projects include ScienceWorld [342] and other specialised simulators designed to simulate higher-fidelity text-based environments, Jericho [142], which provides a Python framework for loading interactive fiction games, and TextWorld [82], a framework for creating research-oriented games.

With the development of NLP technology, AI conversational systems gradually entered the social domain. In the art world, Anna Ridler's "Alice & Bob" (2017) blends quantum computer data with a text corpus composed of poetry to generate love letters [279]. In November 2017 the first version of the Replika AI companion [273] was released. As a commercial product, Replika does not disclose any details on its technical pipeline. However, prior to 2019 fluent dialogue in AI conversational systems generally depended on high-quality preset corpora and refined NLU technology, requiring significant manual labour — unlike today's LLMs, which can generate content beyond their developers' pre-written scripts. By 2018, Replika had reached 2 million users.

In summary, these pre-generative systems, while innovative for their time, relied on a top-down, symbolic approach where the AI's primary function was to understand or map player input onto a pre-authored, finite world model. The advent of LLMs, as will be discussed, marked a fundamental shift towards a bottom-up, generative paradigm where the AI itself becomes a co-author of the narrative world, capable of producing content far beyond its initial script.

2.6.3 Stage II: After GPT-2, Before ChatGPT (2019-2022)

The release of GPT-2 in 2019 marked a significant milestone: generative AI models had begun producing legible and coherent text. However, a 2022 survey by Peter Jensen at the end of this period revealed that after years of active AI research, the best text game agents could only solve about 12% of Zork, a game released in 1977[161].

The development of NLP technology made various technologies more accessible. For example, "Interview with the Whisperer" [85] and "Mystery of Three Bots" [86] allowed players to explore mystery stories using natural language text input through Semantic ML, a tool for semantic analysis developed by Google [127]. Fraser et al. [115] developed open-domain social conversational AI using emotion detection. These applications were still more reliant on NLU than NLG technology. In contrast, AI Dungeon [2] became the most iconic game of this period by allowing players to fully generate their text adventure with natural language input. The main goal of such games was to enhance the gaming experience by providing an immersive and engaging environment [15], demonstrating that natural language interfaces, though challenging to use, can offer high levels of engagement and enjoyment.

NLP was also utilised in parts of commercial games like "KuiLeiXi" [362]. These games employ natural dialogue systems as the primary mechanism to advance the plot. However, their impact remained limited, thus restricting them to small-scale experiments which only supported text-based interactions.

In the realm of art, more artists began integrating NLP technology into their works. "Can the Subaltern Speak?" [104] used AI-generated Morse code for communication, drawing inspiration from a Facebook experiment [108] in which chatbots developed their own language. This project highlighted AI's potential for language generation and the role of language in empowering marginalised voices. Although the project included AI-generated language through GPT-2, it utilised predefined rules, such as Markov chains and Morse code, to encrypt fixed information in the English language, rather than delving into the cultural context behind the emergence of language. "Cangjie's Poetry" explored AI's capacity to establish a Chinese symbolic mode of communication [387], including the use of image captioning technology to generate textual interpretations of image-captured scenes.

Advances in NLP technology suggested that AI agents would increasingly pervade people's daily lives. Evol4 [253], a mobile game allowing players to build relationships with virtual characters, arranged a special event on Valentine's Day: players received real phone calls and listened to the greetings and affectionate words of characters from the game. Microsoft released the social app Xiaoice Land [221], enabling users to communicate with AI agents through natural language text messages. This trend demonstrated that integrating AI agents into virtual worlds and enabling them to perform in the real world could enhance the experiences of human users.

2.6.4 Stage III: After the release of ChatGPT (2023-present)

Following the release of ChatGPT, the concept of AI agents gained unprecedented attention. This period saw the emergence of extremely important research. In "Generative Agents", AI-driven inhabitants of a virtual town exhibited complex social behaviours, like spontaneously organising a Valentine's Day party [254]. Another study, inspired by the animated series "South Park", enabled agents to create new episodes that were primarily driven by AI-generated character dialogues [199]. In the entertainment industry, generative AI has been used in the broadcasting of real-time generated television shows [113]. In the gaming sector, major games like NetEase's "Nishuihan" [315] began deploying conversational NPCs with behaviours and responses that were partially powered by an LLM, freeing NPCs from the limitations of pre-scripted content. The developers of several games, including GTA [33] and Skyrim [1], incorporated LLMs to enhance NPC dialogue.

Advances in LLMs led to the emergence of **AI-native games**: games that incorporate generative AI (GenAI) as the core gameplay element. As LLMs evolve, both large companies [9] and independent developers [369] are increasingly incorporating "free input dialogue with Non-Player Characters (NPCs)" into their games [329]. This can lead to novel gameplay that could not exist without GenAI: content (like conversations) is generated, rather than being predefined by developers. GenAI creates new content; while conventional AI may also generate some simple forms of content like labels or paths, GenAI produces high-dimensional data like text, conversations or images [385].

This evolution aligns with the previously discussed **AI-based games**, which are "designed around AI" [335] or have "an AI system deeply integrated into their core mechanics and aesthetics" [98]. However,these studies often encompasses a broader spectrum of AI systems, including elements like the physics simulation system in Super Mario Bros, Go AI, and more. Earlier AI-based games, such as Prom Week [216], Facade [212], and Ice-Bound [270], while focusing on text interactions and social simulations, did not support natural language input. Moreover, they did not utilise neural network-based models (generative AI) for real-time content generation. These games still heavily relied on content that had to be pre-prepared by developers, including story segments, dialogue scenarios, and granular lines of dialogue.

This is not to say that these games were limited in any way. However, with the rapid advancements in generative AI, there is a pressing need to update discussions in the gaming and interactive storytelling field. Previous taxonomies struggle to encapsulate the unique characteristics of emerging AI-native games, such as real-time multimodal content generation (like text-to-image transformations), the high degree of freedom offered by natural language input (real-time generation as opposed to pre-prepared character dialogues), and the associated challenges of inconsistency and authorability (developers cannot fully define the boundaries of generated content). Recent research [292] has initiated a discussion on the design considerations for games based on GenAI, like AI Dungeon [2], but has not distinguished them from conventional AI. In general, using GenAI in games falls under the broader discussion of procedural content generation (PCG), which refers to the automatic or semi-automatic creation of game content such as levels, maps, items, quests, and textures [314][374]. However, neither

"AI-based" nor PCG adequately describe the essence of GenAI, which produces multimodal content beyond developers' pre-definitions.

The term "AI native" has primarily been used in the AI industry [140] and has not yet been widely adopted in game studies. Nevertheless, it is highly relevant to the future utilisation of AI in games. VentureBeat [55] defines "AI native" products as those which have AI embedded in their core. In essence, if AI was not part of the product, the product would not exist. This definition specifically refers to GenAI, distinguishing it from the broader definitions of AI in earlier literature. Similarly, Ericsson defines "AI-native" in the context of business support systems (BSS), where AI is central to the system's functionality [207].

In the realm of game AI studies, VentureBeat's definition is particularly valuable for future game AI development. **An AI-native game is one in which GenAI is not just an added feature but is fundamental to the game's existence and mechanism.** This does not mean that content and art assets are pre-generated by AI, but rather that the capabilities of GenAI are central to the gameplay experience in real time. Such games are gaining attention through initiatives like AI-game hackathons held by developers and AI organizations such as Huggingface [103]. Recognising the significance of AI-native games could pave the way for novel types and genres of games [335].

Although AI Dungeon is a pioneer in this category, many critics have pointed out that while the kind of AI-generated content seen in AI Dungeon [180] seems to offer freedom, it may lack convincing justification and fail to provide players with meaningful control. The surge in AI content appears to have undermined the role and expression of authorship [292]. The advent of GenAI [232] has raised concerns about the production of "infinite trash": an endless array of stories with diminished emotional appeal[78].

Developers have sought a balance between generating AI content and adhering to traditional game mechanics. While Square Enix's detective game [316] engaged natural language conversation as a core mechanic, "Origin" by InWorld [9] featured immersive negotiation with characters via audio input. Although these elements were prominent parts of the respective games, developers still endeavored to align the generated narrative with established story rules.

Independent developers took more innovative approaches. Yandere Simulator [369] fuses 3D scenes with dialogue, for example by prompting players to persuade a character to exit a room. Echoing the dynamics of Facade [212], voice input generates unique interactive dialogue, allowing players to think creatively. Aside from dialogue-centric games, "Split it" [242] transforms famous paintings into puzzles using an image in-painting model. In the game "Dead Meat"[3], players can engage in dialogues with suspects using voice input to identify the liar. These innovations underline the need for developers to provide clear goals and structures for generated content to ensure it is meaningful and enjoyable.

Beyond gaming, more AI-native applications have demonstrated AI's integration into our daily lives, signaling the urgent need for more profound ethical discussions as AI agents' social agency increases. In February 2023 Replika banned sexual interactions with its chatbots, sparking widespread debate and criticism, and even leading to psychological issues among many users [201]. Furthermore, LLMs have given rise to AI agents with more complex social capabilities, such as Xiaoice Land, a social simulation on a virtual land [221], and a recent application that created a fictitious social network composed entirely of AI agents sharing their fabricated daily lives [254]. Complex social behaviours have been simulated through LLMs, for example in the "Werewolf" games [367] and even war simulations [179]. Moreover, with GPT-4 being recognised as a potential precursor to artificial general intelligence (AGI) [46], the negative capabilities of LLMs, such as bias, deception, and concealment, demand greater ethical consideration. Recent research argues for recognition of various degrees and dimensions of agency when ethically evaluating AI [131], as this may allow a more realistic and practical understanding of AI's role in social contexts without the constraints of human-like agency.

2.6.5 Key Challenges and Limitations of LLM-driven Narratives

Unpredictability and the Paradox of Control

While the generative capability of LLMs offers unprecedented freedom, it also introduces significant challenges related to narrative control and predictability. As demonstrated by the emergent, disruptive behaviors in AI-driven games like Google's Werewolf [118] and the spontaneous fabrications observed in this thesis's own case studies (Chapter 6), the stochastic nature of LLMs can lead to outcomes that deviate from authorial intent. This creates a tension analogous to the narrative paradox [197], forcing creators to constantly balance between granting agency to the AI and maintaining a coherent, meaningful player experience.

Lack of Deep Narrative Understanding

Furthermore, LLM-driven narratives can suffer from a lack of deep understanding in narrative content. Although capable of generating locally coherent text, their probabilistic nature means they do not possess genuine intentionality or a true model of the story world's causality[345]. This can manifest as the modern variation of the Tale Spin effect [346], where stories lose consistency over longer interactions, characters forget key motivations, or plots fail to build towards a dramatic climax, thereby challenging the perceived believability of the agent.

Inherent Biases and Ethical Concerns

Finally, a critical limitation of LLMs is the inheritance of biases from their vast training data. As these models are trained on internet-scale text corpora, they can inadvertently reproduce and amplify societal biases related to culture, gender, and other social factors [133]. This was observed in the ORIBA study (Chapter 7), where the AI agent exhibited a cultural bias in its understanding of non-Western contexts. Such biases not only risk creating stereotypical or harmful content but also raise profound ethical questions about the responsibility of creators who deploy these systems.

2.6.6 Identifying the Research Gap

Storytelling which spans fictional and real worlds creates **meaningful interactions** by resonating with the user's goals, values, and sense of purpose, making the interaction feel significant and worthwhile[219]. However, this blurring of fictional and real-life boundaries also introduces challenges. The stories produced by human-AI interactions can now yield social or ethical consequences, fundamentally altering the contract between the audience and the narrative. This raises deeper questions: how should human researchers and creators respond to the interplay in these engaging stories between AI's performance and human users' subjective experience? What ethical considerations need to be addressed?

With the development of generative AI, particularly LLMs, AI models' generative capabilities now allow the interactions between AI and humans to produce stories that are personalised and not pre-defined. Recent research has suggested that when AI systems begin to interact with humans and display autonomy and proactiveness, they inherently acquire character and role-playing attributes [306, 305]. The roles they play might be as simple as that of an assistant proofreading articles based on simple prompts, or as complex as that of a diligent baker in a generative agent simulation system [255]. While current research predominantly focuses on the content generation capabilities of these agents — how the agent speaks and behaves within the storytelling context — little attention is paid to how humans should navigate the complex generative relationships underlying these agents.

Recent research shows that GPT-4 tier LLMs can generate causally sound stories at small scales, but planning with character intentionality and dramatic conflict remains challenging[345]. While the rise of LLMs prompts philosophical debates on whether AI agents can truly "understand" stories or merely simulate comprehension through probabilistic modelling, this thesis takes a pragmatic stance. Our focus is on the **perceived believability** and **real-life consequences** of these interactions, examining how characters driven by LLMs **affect** user engagement and narrative consistency in a storytelling context.

This has led to several research gaps:

1. Traditional frameworks for characters in interactive storytelling (such as believable agents [34]) are primarily based on symbolic methods [92], employing top-down rule designs. These symbolic frameworks, often manifesting as complex, reactive NPC systems, **struggle to capture the nuances of human-like believability**—specifically, consistent emotional expression, internal motivation, and adaptive personality—which are crucial for sustained social engagement. They fail to account for the connectionist capabilities of modern AI agents based on generative AI, particularly LLMs, whose bottom-up emergent behaviors allow for a higher degree of perceived consistency and personality.

2. Previous research has overlooked the close relationship between an agent's role-playing attributes and its context. The storytelling abilities of AI agents (including dialogue, behaviour, appearance, etc.) have typically been confined to specific scenarios: either within games (inside the magic circle) or as applications (such as AI assistants) playing fixed roles outside the magic circle. This limitation leads to a lack of understanding regarding the **design mechanisms required for consistent context-switching**—i.e., how a single agent can maintain its persona and believability while traversing the boundary between fictional and real-world settings.

3. From a creator's perspective, one main challenge of interactive storytelling lies in balancing authorial intent and player agency. As AI agents become capable of generating stories beyond the creator's control, human creators (such as game developers, writers, and artists) need to reconsider how they view and utilize AI.

In essence, the stories generated by AI agents exist not only within the virtual magic circle, but may also extend into the real world while maintaining consistency. Therefore, we need to consider the affordances of AI agents in the real world as part of interactive storytelling.
This thesis introduces the concept of hybrid agents, which are AI agents that adaptively role-play in both fictional and real-life scenarios in a believable and consistent manner.

2.7 Summary

The evolution of AI technologies, and particularly of language models, has significantly expanded the opportunities to create believable and engaging virtual experiences. However, as this review demonstrates, the impact of AI in interactive narratives and games is not solely about the capabilities of the models themselves. Rather, it is deeply intertwined with the subjective experiences of users, the contexts within which these AI agents operate, the content they generate, and the role of human creators in shaping these interactions. The concept of hybrid agents emerges as a crucial framework for understanding and developing AI entities that can seamlessly navigate between fictional and real-world contexts, blurring the boundaries of the magic circle.

In the following chapter, I will synthesise the insights gained from this literature review to identify key research opportunities and propose essential dimensions and metrics for evaluating hybrid agents.

Chapter 3

Conceptual Framework and Research Approach

As established in the preceding chapter's analysis, research gaps exist in understanding and evaluating AI agents that operate across fictional and real-world contexts. To address this gap, this thesis introduces the concept of 'hybrid agents' and proposes a framework with three core metrics of expansion.

3.1 Conceptual Framework: Metrics for Hybrid Agents

Previous research has suggested that when AI systems begin to interact with humans and display autonomy and proactiveness, they inherently acquire character and role-playing attributes [306, 305]. Such roles might be as simple as an assistant proofreading articles based on simple prompts, or as complex as a diligent baker in a generative agent simulation system [255]. While current research predominantly focuses on the content generation capabilities of these agents, such as how they speak and behave within the simulation, there is a lack of attention paid to how creators should navigate the complex generative relationships behind these agents, and how they pervade our daily lives.

At the outset of this thesis, I identify two fundamental aspects of hybrid agents:

1. The content aspect, whereby AI agents act as creators able to adaptively generate content (e.g. text and corresponding voice, images, etc.) based on the player's input.
2. The context aspect, whereby AI agents perform as actors. They are able to break out of the software framework and enter real-word contexts (e.g., telephone communication). They are increasingly integrated into daily human activities, a trend facilitated by the ubiquity of computing and underscored by recent advancements in AI technology.

Combined, these two aspects of hybrid agents define how they become hybrid. This chapter will elaborate on these two aspects in detail (Fig.3.1.1).

As AI constructs begin to act as both creators and actors, human creators need to navigate this emerging landscape, determining whether they can transition to the role of directors who guide these agents from a higher level.

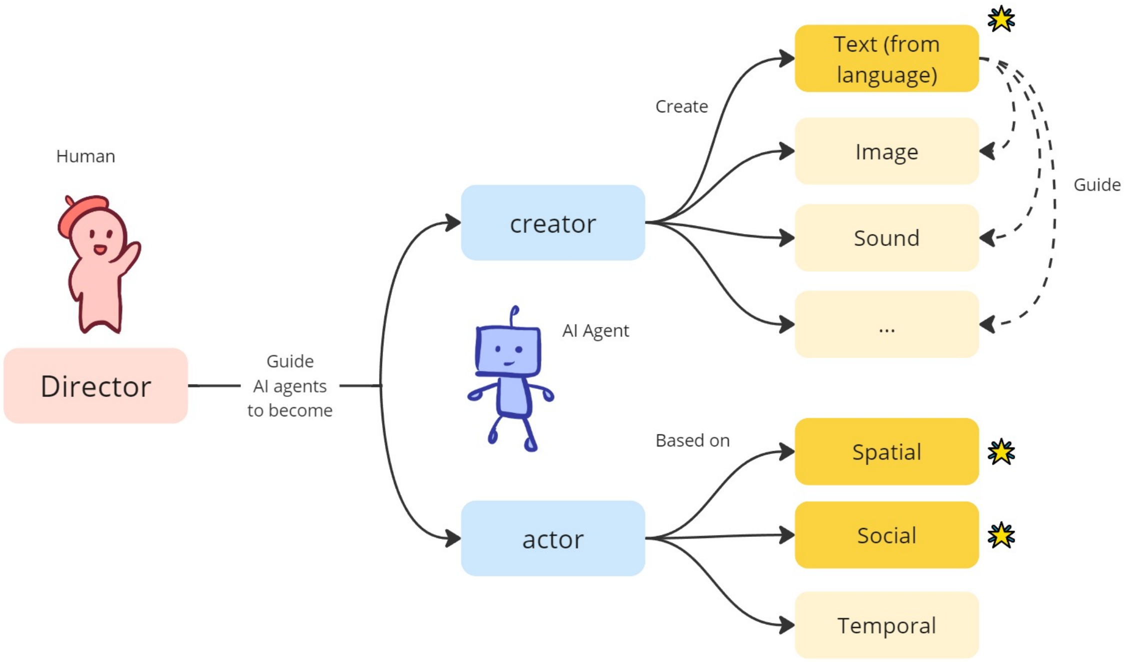


Figure 3.1.1: Hybrid Agent metrics

I propose that AI agents are hybrid in nature: they blend real and fictional social relationships and spaces, merge reality with fictional storytelling, and may even combine human and non-human languages.

3.2 In Terms of Content: Creator

The content aspect of hybrid agents — their ability as creators — has been the focus of extensive prior literature. For nearly every category of creator (e.g., writers, musicians, painters), there exists a wealth of research on related creativity support tools [56, 57, 379, 83], and even phenomenal applications that have sparked controversies and protests: the Hollywood writers' strike against LLMs like ChatGPT [19], visual artists on ArtStation [166, 304] opposing Stable Diffusion [286, 47] and MidJourney [222], and musicians conflicting with Suno [321]. These phenomena reflect a sense of unease and perceived invasion: AI technology encroaching on creative territories once considered to be exclusively human endeavours.

While substantial research has been conducted on mixed-initiative co-creativity and corresponding authorship [165, 190, 331, 67], I propose that when AI agents are integral parts of the creative practice, which is dynamic and gradually forms through interaction, this differs from the use of customised creative support tools designed for personalised output and collaboration [182]. This primarily depends on whether the author aims to create the entity itself or merely target the artifacts it produces (e.g., text and images). This creative process, still relatively new having emerged due to recent advancements in generative AI, is exemplified by platforms like Character.ai [59] where users can customise chatbots (conversational agents). The user community treats these bot creators as authors and tracks their works — different agents [223]. The text generated by these agents is formed through diverse inputs from different users, rather than being static. A creator seeking complete control over their authored text would not typically engage as an "author" on such a platform.

The term "hybrid", as used in this thesis, is neutral. It may refer to the sense of personal creative labour being encroached upon, as seen in creator protests, or it may denote conscious control and regulation.

3.2.1 Language Expansion

Although the focus of this study is not on philosophical concepts, by integrating philosophical notions about language we can discern the potential of AI agents in the realm of language [58].

Ludwig Wittgenstein's "Tractatus Logico-Philosophicus", published in 1921, marked a significant point in the early 20th century's 'linguistic turn' — a period during which the importance of language was explored across various academic disciplines. Wittgenstein famously stated that "The limits of my language mean the limits of my world"[356].

Jaakko Hintikka described "language as the universal medium" [145], suggesting that "We cannot ultimately escape our language to view it and its logic from the outside. Consequently, the semantics of our language is inexpressible and cannot be theorised about in language itself" [240].

Thomas Martin [206] argued that the world is made of language, and we cannot escape the "prison-house of language" [160]. Language not only shapes agency, but also — in a profound sense — constructs the very essence of the world. The world does not exist as a separate entity from us; similarly, language is not just an insubstantial or disembodied model. Through language, we are intricately involved in creating the fabric of reality.

Contemporary artists, particularly those who do not speak English, actively contribute to this notion. Xu Bing's "Book from the Sky" [116] and "Book from the Ground" [366] are notable examples. These works, devoid of existing textual content, combine indecipherable Chinese characters, visual symbols, and expressions to convey emotions and meanings. Xu Bing's approach transcends the limitations of traditional language, allowing interpretation based on personal understanding and emotions. This opens new avenues for communication using constructed languages and highlights the artistic value of exploring human language. "Can the Subaltern Speak?" [104] uses AI-generated Morse code as communication, representing how Bandari women communicate through eye-blinks from under their masks.

In the reviewed literature, the language used in dialogue systems is considered as the base and medium for communication. From this perspective, we can inquire: how will AI agents contribute to the reinterpretation and recreation of language? Can machines transcend the human 'prison of language' and offer a different view of the world?

This thesis, through a series of projects, specifically focuses on conversational interactions that exploit the potential and implications of language. Now that people are aware that machines can communicate through natural language, what comes next? Will AI agents, trained on human-centric data, start to communicate in non-human ways? This is just one aspect of the issue that invites attention, but I have a specific interest in this particular perspective, as the breakthrough of LLMs forms the foundation of this thesis.

3.3 In Terms of Context: Actor

The second aspect of hybrid agents is context: their capability as actors on stage. The previously mentioned examples of pervasive games are particularly relevant because they treat interactions and content outside the game as dynamic, growing, and interactive. We do not claim that this concept exactly fits the discussion of AI agents since the theory is based on the broader category of games. As scholars have noted, “The regular game is played in certain spaces at certain times by certain players”. The existence of ‘regular’ games implies the possibility of ‘irregular’ games, leading to the conclusion that “Pervasive gaming is a genre of gaming which systematically blurs and breaks the traditional boundaries of games” [228]. Obviously, AI agents’ interactivity is not necessarily limited to play and interactive storytelling. However, because games inherently involve interactivity and the dynamic relationships between the developer (the creator), the artifact (the game being played), and the player, this theory is highly applicable to AI agents’ pervasiveness.

3.3.1 Spatial Expansion

For games, spatial expansion refers to the physical location of the game being unclear or unlimited. Games such as “I Love Bees” [136] use cityscapes as playground, expanding locally and even globally. The game can be played anywhere and everywhere, blurring the boundaries of the game space [229]. Spatial expansion can occur in physical spaces (e.g., using a whole city as the game area) or in digital spaces (e.g. social medias and messaging platforms).

For AI agents, spatial expansion currently refers more often to virtual spaces, such as those produced by extended reality (XR) technologies, or digital spaces like Discord and text chats, than the physical world itself, since the technology required to embody AI agents in robots is not yet mature. In other words, AI agents’ habitat is digital space.

AI agents may also be limited to a specific setting and space. For example, Character.ai [59] only allows players to interact on a webpage, so the webpage can be considered the edge of the magic circle: interactions with fictional characters are limited to this space. However, some AI agents blur this boundary. One example is “A Week With Wanda” [135], a game in which characters send emails or messages to players to inquire about their next choices.

The social scene of most hybrid agents is limited to specific apps or games. However, human players can move between different venues while retaining the same identity. Human speech on social media can affect how others perceive them through instant messaging and in other situations. At the end of the day, humans are complex social animals and evaluate each other on the basis of continuous observations in multiple situations. Current research only provides discontinuous observations of hybrid agents in a single context; this is an underexplored area from the perspective of metaverse construction. Some cross-platform character applications, such as Ready Player Me [358], currently exist, but they are limited to character appearance and do not involve conversational ability. Editors that focus on conversational ability, such as InWorld, only support interactions between characters and humans within specific game engines. The research in this thesis is intended to observe people's reactions to hybrid agents by introducing multi-modal AI agents to virtual environments and conversation platforms outside of games (such as Twitter and Discord).

3.3.2 Social Expansion

In online communities like Discord channels [268], it is essential to achieve meaningful interaction. Although many active communities employ bots with both moderating and entertaining functions, their interactions may not necessarily generate new information or contribute to the community's values or project goals. Narratives, according to Ricoeur, are vital for constructing our sense of self, making sense of our experiences, and creating meaning in our lives [12]. Storytelling has been considered useful in various contexts, such as children's education [388], [158], healthcare, and the learning of skills [318]. In online community building, particularly game-related communities, storytelling may serve as a crucial goal or primary means of maintaining community activity and a sense of meaning [12].

The integration of an AI as a member of an online community could potentially become a source of discussion or meaning if it shares life experiences which are consistent with the content the entire community focuses on. These experiences are unlike the small talk generated by existing SCs such as Replika [41]. While there has been extensive research on dyadic interaction chatbots [42, 63, 309], multi-party chatbots, particularly those acting as community members, remain underexplored.

Researchers have developed a Twitch chatbot that focuses on the social context of a community, making the chatbot a member of that community [301]. However, this development does not consider the potential impact of current LLMs, such as uncontrollable content generation and human attachment. Seering et al. have proposed various ideas for community chatbots, including the storyteller bot concept [301]. They suggest that a more interactive, almost 'live', narrative experience could be created by chatbots that are regular community members involved with other chatbots in engaging ways. There should therefore be a focus on the deployment of chatbots in specific social contexts.

Scholars in the gaming field [285] have put forward similar viewpoints, examining tropes in conversations between player characters and NPCs. Drawing from the fields of pragmatics and conversation analysis, they show how these tropes differ from real face-to-face conversations. Based on this, they propose Trope-Informed Design, which treats tropes as tools that can make or break a player's experience. Although 'trope' in this context refers to the plot of a game, this mechanism can be extended to community chatbots as they both exist within a social context.

Social expansion offers opportunities to form communities by encouraging spontaneous interactions between people who do not know each other, and potentially spices up outsiders' lives by allowing them to act as spectators or participants. For instance, the virtual streamer Neurosama [338] is driven by LLMs and regularly streams on public platforms such as YouTube and Twitch. Her actions and responses are generated in real-time based on the chat behaviour of her audience. She has amassed a fan base of over 200,000 followers. It can be said that, as a fictional character, her persona is constructed through interactions with her followers. Her followers spontaneously compile information about her. Due to the stochasticity of generative AI, Neurosama's age is inconsistent. During the stream, she has variously claimed to be various different ages [339]. Even an outsider who is not a follower of Neurosama may stumble upon her livestream and engage her in chat, prompting her to provide rich responses that can influence others' perceptions of her.

Temporal Expansion as an Aspect of Social Expansion

While Montola's framework for pervasive games identifies temporal expansion as a distinct category [228], this thesis integrates temporal aspects into the broader concept of social expansion when considering AI agents. This integration is justified for several reasons.

Firstly, the notion of temporal expansion in games is predicated on the existence of discrete 'play sessions' [228]. However, interactions with AI agents often lack such clearly delineated temporal boundaries. For instance, companion applications like Replika [273] allow for continuous engagement without explicit start or end times.

Secondly, the blurring of temporal boundaries in AI agent interactions is inextricably linked to their social integration. As Montola notes, "Temporal expansion ties in with social expansion, as the temporal span of the game is often obfuscated to the point where even the players might be unaware of whether they are playing at a given moment" [228]. This ambiguity is inherent in many AI agent interactions, where the line between 'in-character' and casual conversation may be indistinct.

Furthermore, the temporal flexibility of AI agents — such as their ability to respond at various times of the day or maintain conversation threads over extended periods — is fundamentally a social feature. It reflects the agent's perceived autonomy and capacity for ongoing social engagement, rather than a purely temporal phenomenon.

Examples like "Lifeline" [7] in which players experience real-time delays in character responses demonstrate how temporal elements enhance social immersion rather than functioning as a separate expansive dimension. Similarly, AI companions that adjust their communication patterns based on the time of day [273] enhance their social presence and believability.

By merging temporal expansion into social expansion, this thesis acknowledges the intrinsic connection between an AI agent's temporal flexibility and its capacity for rich, ongoing social interaction. This approach allows for a more holistic examination of how AI agents transcend the traditional boundaries of engagement, focusing on their ability to create persistent, evolving social presences that seamlessly integrate into users' daily lives.

3.4 Summary

The two dimensions just outlined — content and context — are not entirely separate; rather, this thesis proposes that they can be viewed from these perspectives. However, the concept of hybrid agents is a heuristic notion that remains open to extension and serves as a foundation for a more in-depth exploration of the topic: my specific interest focuses on the linguistic (language), spatial and social expansion of AI agents. Using these three forms of expansion as foundations, this thesis conducts multiple studies to demonstrate and develop the framework.

- **Linguistic Expansion:** Linguistic expansion refers to the AI agent's ability to reinterpret and recreate language in the context of storytelling and character portrayal. It explores how AI agents, while embodying fictional characters, might develop unique linguistic expressions that transcend human limitations. This expansion can enable AI to generate narratives and dialogues that blend human and non-human perspectives, potentially offering new methods of storytelling and character development.
- **Social Expansion:** Social expansion involves the integration of AI agents acting as believable fictional characters into human social contexts, particularly in online communities. These AI-driven characters can become active participants in social spaces, contributing to community discussions, interactive storytelling, and meaning-making. This form of expansion blurs the lines between human and AI social interactions, potentially creating new forms of narrative engagement in which AI characters persistently exist alongside human community members. It includes the ability of AI agents to adapt their fictional personas to specific social contexts and participate in multi-party interactions, maintaining character consistency across various social situations.
- **Spatial Expansion:** For AI agents playing believable fictional characters, spatial expansion primarily refers to their ability to transcend specific digital spaces or platforms while maintaining their narrative integrity. This form of expansion allows AI-driven characters to interact across various digital environments, such as social media platforms, messaging apps, or virtual reality spaces, as part of a cohesive storytelling experience. It represents the breaking down of boundaries between different digital contexts, allowing for a more pervasive and continuous presence of fictional AI characters across various aspects of users' digital lives, enhancing the immersive quality of the narrative experience.

Linguistic expansion focuses on content creation, while social and spatial expansion emphasize how hybrid agents act and interact in various contexts. When AI agents' ability to generate content is combined with their ability to tell stories across different contexts, these interactions can become meaningful and engaging. These forms of expansion collectively contribute to the concept of hybrid agents, which blend real and fictional elements, merging reality with storytelling across various dimensions of interaction.

I do not assert that I have created the most effective solution or interface for making AI agents both believable and engaging. Nor do I claim that this system provides optimal usability. The research has two primary objectives.

Firstly, I aim to explore how fictional storytelling elements can be integrated into conversational AI agents, with a particular focus on language, social, and spatial expansion.

Secondly, and more importantly, I seek to stimulate further discussions and considerations in this field.

The main goal of the research is not to develop specific methods that achieve certain tasks. Rather, through a series of projects spanning video games, Discord chatbots, and interactive installations, I aim to showcase the potential of AI agents.

3.5 Research methodologies

The question this thesis aims to answer is: How can we understand and develop hybrid agents — LLM-driven social agents that inhabit the half-real world — and exploit their capabilities to enhance believability and engagement? This broad question has been broken down into the following research questions:

RQ1: How can **fictional characters** be leveraged to elevate the performance of **social agents** to achieve engaging and meaningful interactions as **hybrid agents**?

RQ2: How can hybrid agents in interactive storytelling become more **believable**?

RQ3: How can hybrid agents in interactive storytelling blur and expand the boundary of fiction and reality?

To answer the multidisciplinary research questions, the primary goal is not only to address specific technical challenges but also to explore and demonstrate the potential of hybrid agents. This thesis is based on two overarching research paradigms that guide the inquiry: Research through Design and Practice-led Research. These two paradigms are translated into a mixed-methods research strategy throughout the thesis.

3.5.1 Research through Design

Research through Design (RtD) is a research approach where the act of designing and making an artifact serves as the primary method of inquiry [392, 117]. This paradigm is particularly suited for addressing the kinds of complex, ill-defined challenges prevalent in Human-Computer Interaction (HCI), often referred to as "wicked problems" [392]. My research into hybrid agents—exploring their social roles, creative potential, and ethical dimensions—falls squarely into this category. These are not problems with a single correct solution that can be discovered through traditional scientific methods alone; rather, they require an exploratory and generative process to understand what a potential future might be [124]. Therefore, this thesis adopts RtD as a core methodology, treating each AI agent not as a final product, but as an artifact of inquiry designed to probe the problem space.

This thesis operationalizes the RtD paradigm through an iterative "build-to-understand" cycle [213]. This process is not linear but cyclical, where each artifact informs the next. I first constructed pilot studies (Wander, 1001 Nights) to explore the broad design space of conversational AI and narrative engagement. The insights gained from these initial artifacts—specifically regarding the power of linking real-world data and providing player agency—directly informed the design of the subsequent, more complex *Catherine & David* study. This subsequent study was purposefully designed to investigate a more focused question about social expansion within a live community context. This iterative process of designing, building, deploying, and reflecting upon AI agents allowed for a progressive refinement of the research questions and a deeper understanding of the design space for hybrid agents.

Crucially, in RtD, the designed artifact itself is a primary form of knowledge contribution [392, 120]. The AI agents developed in this thesis function as "design exemplars" that embody a particular framing of the problem and articulate a visio[392] where human-AI interaction is more meaningful, creative, and socially integrated.

3.5.2 Practice-led artistic research

Researcher Linda Candy [52] distinguishes between two types of practice-related research: practice-based and practice-led research. An original investigation that uses practice to generate new knowledge is practice-based research, where creative outcomes such as designs, music, or exhibitions are the basis of the contribution to knowledge. Practice-led research, on the other hand, is primarily focused on advancing knowledge about practice or within practice. Such research includes practice as an integral part of its method and often falls within the general area of action research, with the results being fully describable in text [52].

As an artist and researcher, my PhD research also aligns with the practice-led paradigm. The inquiry began with the creation of artworks, which were then studied through user engagement, leading to peer-reviewed publications that informed subsequent research cycles. This process embodies the concept of **“knowing through making”** [202], where knowledge is generated through the creative act itself. The AI agents I created are central to this process, but as Mäkelä argues, such artifacts can be seen as ”mute objects” that do not reveal the knowledge they embody on their own [202]. Therefore, a crucial task of this research is to **“give a voice to the artefact”** [202]. This written thesis serves precisely that function: it is a critical reflection and theorization that articulates the knowledge discovered through the practice of making. Challenges encountered during the production of art drove further theoretical and conceptual research, while deepening my understanding of the conceptual issues fed back into the refinement of my practice.

From the art practice perspective, I also found artistic potential in the rapidly developing field of natural language processing (NLP), which has inspired my entrance into computational art. Although the transition from art practice to scientific research outcomes is rare in related studies, my pre-PhD pilot studies demonstrate the effectiveness of this approach.

Throughout my research, I narrowed my focus to the field of interactive storytelling, which is the source of my motivation and a long-standing theme across HCI, gaming, and interactive media art. I found that outcomes based on conversational agents are naturally relevant to HCI and gaming when the focus is on interactive storytelling. Additionally, due to their expression of creativity, these projects are often regarded as media art. As a result, HCI, gaming, and media art are intertwined in my research, with a focus on conversational AI as the technical theme and interactive storytelling as the main area of interest.

3.5.3 Translating Paradigms into a Mixed-Methods Strategy

The RtD and practice-led paradigms do not prescribe specific methods; rather, they provide a rationale for selecting and combining them. In this thesis, these guiding paradigms led to the adoption of a flexible **mixed-methods strategy**, integrating both **qualitative** and **quantitative** techniques to capture a holistic view of human-AI interaction.

The qualitative strand of this research was essential for exploring the “why” behind user experiences. It allowed for an interpretive understanding of the nuanced and context-dependent phenomena that emerged from the design practice, such as user perceptions of believability and the social dynamics within a community. In contrast, the quantitative strand provided structured data to measure specific attributes, identify patterns, and enable comparisons between different conditions or agents. This combination allowed the research to be both generative (through making and qualitative exploration) and rigorous (through systematic measurement and analysis), creating a feedback loop where creative practice produced phenomena that were then investigated using established social science techniques.

3.5.4 Measuring Attributes

Believability

A key methodological challenge in this thesis is the evaluation of an AI agent's believability. As the examiner noted, a clear framework for this core concept is essential. Believability is not a monolithic attribute but a multi-faceted construct that has been approached from various angles in the literature. To establish a rigorous foundation for my evaluation methods, I synthesized key academic works that have defined and measured this concept. As Table 2.4.4 illustrates, scholars have identified numerous dimensions crucial to believability, such as personality, emotion, social relationships, and behavioral consistency [208, 348].

The methods employed to measure these dimensions are predominantly user-centric, focusing on *perceived believability* from a Human-Computer Interaction (HCI) perspective rather than on objective AI performance testing. Common methods include user studies with Likert scales after viewing video clips [187, 34], analysis of real-time interactions [154], and qualitative thematic analysis of player feedback [348].

Informed by this body of work, this thesis adopts a targeted approach. A critical observation is that much of this foundational research is centered on embodied agents within graphical virtual worlds. Dimensions such as `appearance` [187], `visual impact` [125], and non-verbal behaviors like `gestures` and `gaze` [154] are frequently discussed. While crucial for game NPCs, these visual and physical dimensions are less applicable to the concept of hybrid agents, whose presence is often mediated primarily through text-based platforms like Discord. Therefore, to evaluate believability in a context where dialogue is the primary mode of expression, it was logical to distill the literature down to the intrinsic qualities of a character that transcend physical embodiment. I selected three core dimensions repeatedly highlighted as central to character depth: Emotions, Personality, and Motivation [208, 34]. To quantify user perception of these dimensions, a 7-point Likert scale was employed, a method consistent with established practices in this topic [187, 34, 125]. This quantitative data was complemented by qualitative methods, primarily semi-structured interviews and thematic analysis, to understand the nuanced reasoning behind participants' perceptions. This mixed-methods approach allowed for a deeper exploration of not just what players felt, but why they felt that way, providing a richer context to the quantitative findings.

Usability and User Experience

Beyond the narrative qualities of the agents, it was crucial to assess their functional usability and the overall user experience. For studies involving direct user interaction, such as

"Wander" (Chapter 4), "Catherine & David" (Chapter 6), and "Hyborg Agency" (Chapter 9), usability was a key metric. The questionnaires in these studies were informed by and adapted from Arnold Lund's widely-used USE (Usefulness, Satisfaction, and Ease of use) Questionnaire [198]. The USE framework provides a robust and domain-general tool for measuring subjective reactions to a system's usability [198]. Key items from the USE questionnaire, such as "easy to use", "easy to learn", and "satisfying", were incorporated into the surveys to provide a standardized measure of the agents' user-friendliness and effectiveness as interactive systems. In the "ORIBA" study (Chapter.7), which focused on creativity support, the questionnaire was further adapted from the User Experience Questionnaire (UEQ) [298] and the Creativity Support Index (CSI) [64] to specifically address the tool's role in the artistic process.

Engagement

Engagement was measured differently depending on the nature of each study, reflecting the adaptability of the mixed-methods strategy. In some studies, engagement was measured through **behavioral metrics**. For example, in the "1001 Nights" study (Chapter 5), player engagement was quantified by analyzing instrumented gameplay data, specifically the "average number of story inputs per play". This provided an objective measure of how deeply players were interacting with the narrative mechanics. In other studies, engagement was assessed as a dimension of **perceived experience** via Likert scale items in questionnaires. For instance, in the "Catherine & David" and "Hyborg Agency" studies, a specific item, "I want to keep interacting with the character", was included to gauge participants' subjective desire for continued interaction. This was then contextualized and explained through the qualitative findings from semi-structured interviews and thematic analysis[43]. This multifaceted approach to measuring engagement ensured that both objective behaviors and subjective feelings were captured.

3.6 Research Roadmap: Studies Performed for this Thesis

Fig.3.6.1 shows the research roadmap of this thesis.

3.6.1 Pilot studies

In pilot studies, I observed people's reactions to LLM-driven chatbots in a story environment through two art projects, "Wander" and "1001 Nights".

Pilot study 1: Wander

Research Questions Addressed: This study primarily addresses RQ1 (How can fictional characters in interactive storytelling be applied to social agents for engaging and meaningful interactions) and RQ3 (How can hybrid agents blur and expand the boundary of fiction and reality). By co-creating narrative context and real-world data on public platforms, Wander created believable and meaningful interactions through the connection to real-world.

Approaches: This study evaluates the user experience of a narrative AI agent connecting real

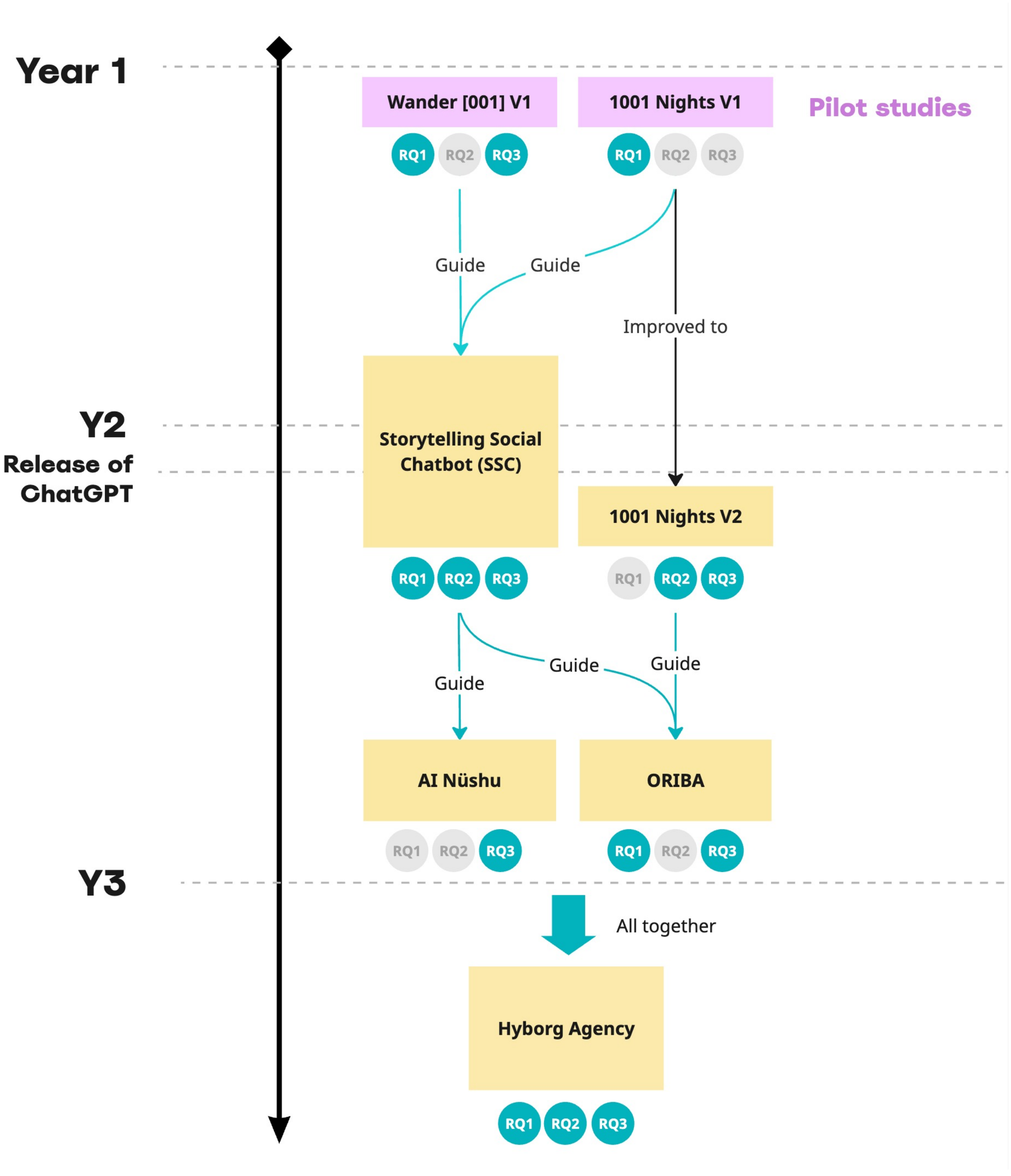


Figure 3.6.1: Study Roadmap

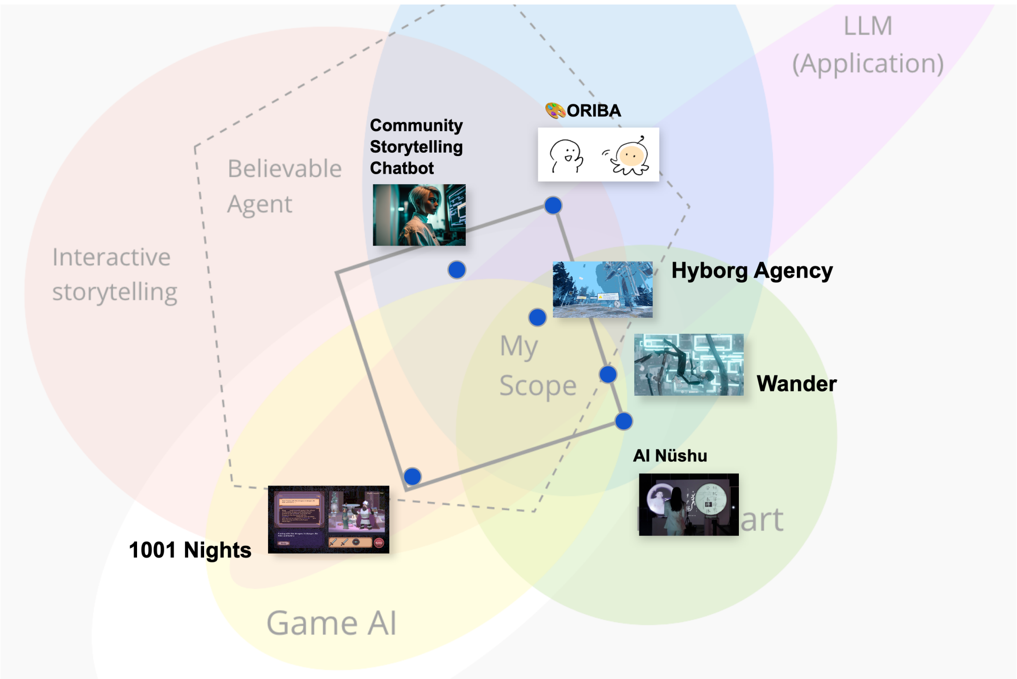


Figure 3.6.2: Works positioning based on research scope

and virtual worlds through quantitative analysis of 268 user questionnaires, complemented by a qualitative analysis of in-depth user interaction cases.

The purpose of this study was to investigate how people interact with an AI agent in a narrative context for the future metaverse. The methods used included:

1. Design and development of a hybrid AI generated content (AIGC) and user-generated content (UGC) system for a narrative AI chatbot, Wander, that produces interactive fiction through knowledge graphs based on user inputs.
2. Quantitative and qualitative analysis of user experience through a survey (n=268), using a 7-point Likert scale with 15 evaluation criteria and multiple-choice questions.
3. Collection and analysis of special cases shared by participants.

Consideration of the contrasts between human-like communication skills and unorthodox knowledge interpretation also contributed to the primary motivations underlying the development of this project. Overall, the prototype was based on two motivations. The first was to bring AI agents to a public daily messaging platform where feedback on the interaction between the AI agents and humans could be easily accessed. The second was to use AI to co-create a narrative context with humans that deformed reality using the defamiliarisation technique — that is, to “impart the sensation of things as people perceive them and not as they are known” [310], presenting familiar objects or ideas in unfamiliar or strange ways.

In this project, I created an AI-enabled conversational agent called Wander. Each time a participant sends it a location message on the Discord or WeChat platforms, Wander generates sci-fi-themed interactive fiction about that place. All text information is generated in real time based on real-world information, so there are no duplicate experiences. Records of the

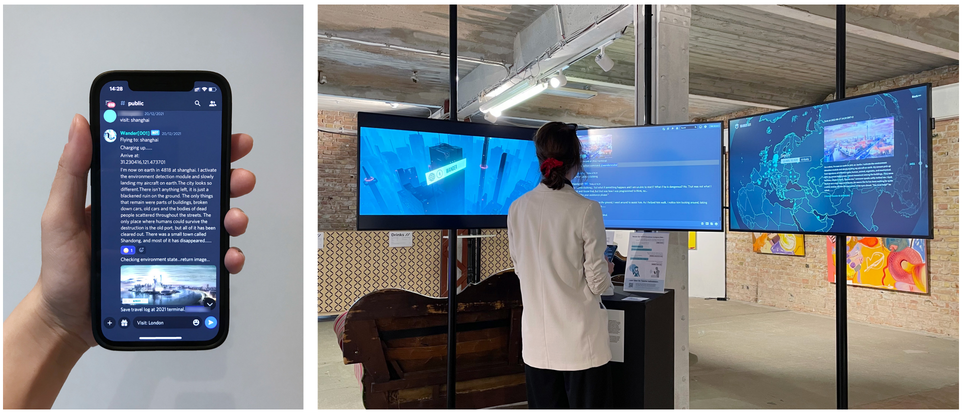


Figure 3.6.3: Wander was exhibited at BBA Gallery Berlin 2022

travelogues are shown on a global map of the fictional future world, which will keep updating through public participation.

Results: Generative AI can facilitate the connection between a virtual world and the real world and establish a relationship with human beings. The study also demonstrates how human-AI interaction can link engagement and creativity in real life to a fictional world. More than 60% of the participants believed that Wander's text descriptions and images often matched the locations they were eager to see. An impressive 91% of subjects thought the feedback they received from Wander felt familiar or even like experiencing déjà vu. For instance, 82.4% of participants correctly identified real locations based on the iconic buildings in the images generated by Wander. Furthermore, 13.1% of participants reported that Wander gave them a sense of company when they felt lonely. 34.3% of participants identified Wander as a friend, which seems to imply that there may be potential for personal relationships between humans and AI applications.

Specific Contributions: By transforming real-world geographical information into dynamically generated science-fiction stories on everyday social platforms (WeChat, Discord), the Wander project provides a preliminary validation of a hybrid agent's potential for Spatial and Social Expansion, blurring the boundary between physical space and fictional narrative.

Impact: In 2022 (the spring semester of the first year of my PhD), the extended paper of this project was published at the IEEE MMSP Conference(Paper 2) and exhibited at the ACM MM Art Gallery(Art 1) . Following the release of ChatGPT and the development of multimodal generative AI (primarily text-to-image technologies like Stable Diffusion), an updated version of the project was showcased at the SIGGRAPH Asia Art Gallery(Art 2) and presented at CHI(Demo2 2) , where it was also published as a poster publication.

1001 Nights

Research Questions Addressed: This study addresses RQ1 (engaging and meaningful interactions through collaborative storytelling). This project demonstrated enacted storytelling that integrating natural language generations with classic stories can enhance player engage-

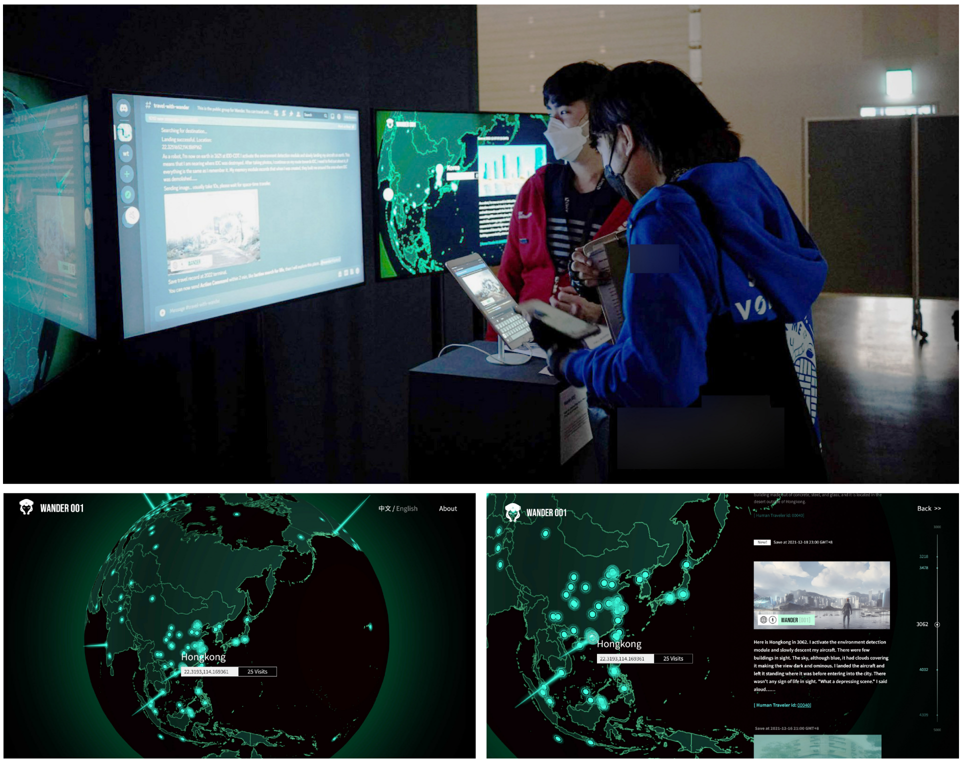

[redacted]
[redacted]

Figure 3.6.4: Wander at SIGGRAPH Asia Art Gallery 2022

ment and creativity.

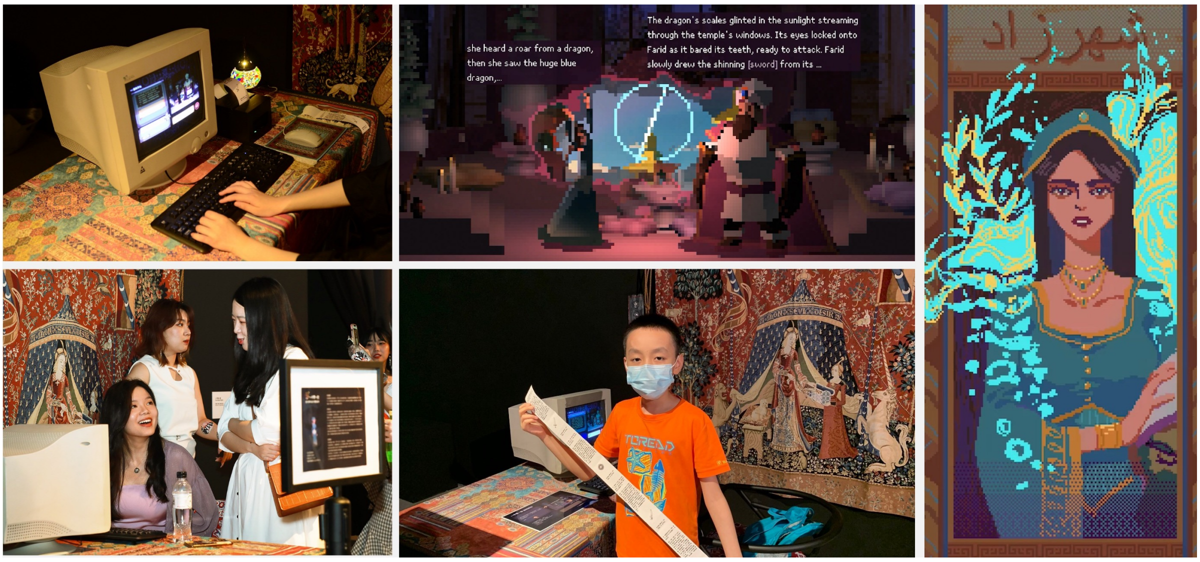
Approaches: This study assesses player engagement and game mechanic effectiveness by quantitatively analyzing instrumented gameplay data from 2,055 players (e.g., story input counts) and conducting a thematic analysis of open-ended feedback from 422 winning players.

How it Builds on Previous Studies: Building on insights from Wander about AI-human collaborative narrative creation, this study explores deeper player agency in storytelling.

1001 Nights is an AI-driven game in which players write stories in collaboration with an AI agent. This explores the question of whether narrative game context can adapt to player expectations.

This game was inspired by the Persian folklore tale “A Thousand and One Nights”. Shahrzad (controlled by the player) uses a dialogue interface to tell part of a story, and then in turn the King (driven by the AI model) continues the player’s story. When the King’s continuation of the story contains weapon keywords like ‘sword’, ‘knife’ or ‘shield’, Shahrzad can use her special ability to turn words into real weapons and use them to fight with the king, creating game mechanics out of the player’s own writing. This leads to an alternative ending for the original story: the female storyteller and heroine, Shahrzad, defeats the tyrannical King and puts an end to his crimes.

Approaches: The purpose of this study was to investigate whether an AI system can encourage players to collaboratively contribute to the story through engagement. The methods used included:



she heard a roar from a dragon, then she saw the huge blue dragon,...

The dragon's scales glinted in the sunlight streaming through the temple's windows. Its eyes locked onto Farid as it bared its teeth, ready to attack. Farid slowly drew the shinning [sword] from its ...

Figure 3.6.5: 1001 Nights, exhibited at Asia Digital Art Exhibition

1. Quantitative analysis of gameplay data from players (n=2055), focusing on the number of story inputs and game achievements recorded.
2. Qualitative analysis of feedback comments from winning players (n=422).

Results: Those players who were more engaged with the game (as measured by the number of inputs they made) were rewarded with better achievements, as intended. Comments collected from winning players provided positive feedback about various aspects of the game, including the game's mechanics and the stories it created. Some of their feedback also indicated a cultural connection formed through creative work: some players expressed their own interpretations of the characters in the classic folklore, and were able to include characters and plots from their own cultural backgrounds in the game's story. This project accordingly works as a proof-of-concept of AI-driven storytelling in this thesis before the release of mainstream LLMs like ChatGPT.
Specific Contributions: Through its core mechanic of ”turning language into weapons,” the 1001 Nights project demonstrates a hybrid agent's application of Linguistic Expansion, where players' narrative creations directly transform into tangible actions that affect the game world.
Impact: In 2022 (the spring semester of the first year of my PhD), the extended paper of this project was published at the ICIDS International Conference (Paper 1). Additionally, a demo was exhibited at the academic gaming conference, Foundation of Digital Games(Demo 1) and the project was shortlisted for the Lumen Prize Longlist of the HUA Award.

3.6.2 Catherine & David: Storytelling community chatBot (SCC)

Research Questions Addressed: This study directly addresses all three RQs: RQ1 through community integration of fictional characters, RQ2 through story engineering for believability, and RQ3 through social expansion in Discord communities. This study validated the ca-

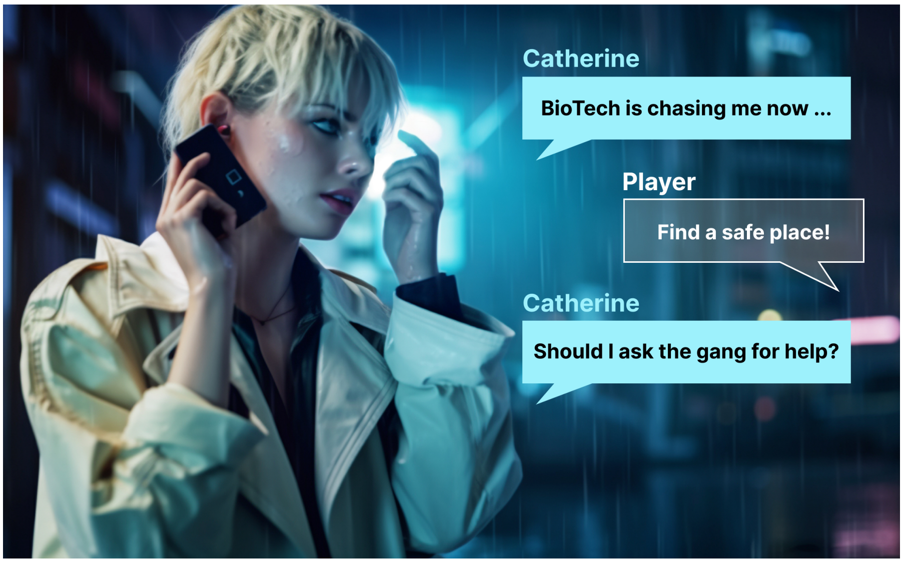

Catherine
BioTech is chasing me now ...
Player
Find a safe place!
Catherine
Should I ask the gang for help?

Figure 3.6.6: Concept art of Catherine in the thesis

pability of storytelling in enhancing the believability and engagement of social agents within daily communication contexts.

Approaches: By introducing a game character into a live Discord player community, this research demonstrates how a hybrid agent can achieve Spatial and Social Expansion by becoming an organic part of the community, where its narrative behavior significantly enhances interaction and emotional connection.

How it Builds on Previous Studies: Following these two pilot studies, I reached a conclusion: **generative AI does engage audiences in a fictional context, and incorporating and transforming real-world community settings and relationships enhances familiarity and enjoyment of the interaction.**

When I initially wrote the chapters in the pilot studies, I had not yet decided to focus on believability as the primary research subject of this PhD project. However, the evaluation of familiarity and relevance to the scene in these studies can be interpreted as applicable methods for measuring believability.

Motivated by these findings, I conducted the next phase of the research, focusing on assessing the believability of characters in everyday human communication contexts. This study aligns with the Research through Design (RtD) methodology by building upon insights from the previous experiments to design more sophisticated AI agents within a community setting.

This research was supported by the official developer team of the game “DE (Alias)”. I introduced two fictional characters, “David” and “Catherine”, as Storytelling Social Chatbots (SSCs) in DE’s online gaming community on Discord. These characters acted as if they were within the world of “DE”. I explored two research questions:

RQ1 - How can an LLM-based SSC enhance engagement and meaningful interactions within a community setting?

RQ2 - How does the integration of fictional stories in the SC’s design impact its believability and engagement?

I refer to the process of transforming a fictional character into a social chatbot as *story engineering*. This term is adapted from *prompt engineering*, which is the design of text prompts to influence the content generated by an LLM. In contrast, designing an LLM-based "live" character with social behaviour requires the consideration of multiple aspects: the character's personality and story, the LLM's generation goals, and the methods by which humans can interact with the character. My prototype, Storytelling Social Chatbot (SSC), implements this concept through the following processes:

1. Story and character design: I defined the SC's personality and the worldview they inhabit.
2. Presenting Live Stories to the community, allowing the SC to recount challenges and problems they need suggestions from the community to resolve
3. Communication with community members, enabling community members to chat with the SC. I designed a workflow to drive the character using the GPT-3 LLM.

To evaluate my workflow, I introduced two fictional characters, "David (for pilot study)" and "Catherine", as SSCs in an online gaming community based on the game "DE (Alias)" on Discord, a social platform commonly used by players. These characters acted within the worldview and story of the "DE", which was under development at the time. Every day, each character shared their current situation in the story channel (e.g., "I'm being chased by the evil agent, what should I do?"), in a similar manner to social media. Meanwhile, community members could engage with them and discuss their current situation in the chat channel, making decisions by voting on what the character should do. At the end of each day, the character would make their decision based on the community's vote and release the next story.

Apporaches: To evaluate the SSCs, I compared them with Jerry, a non-storytelling chatbot that already existed in the community. I collected feedback through questionnaires about usability and believability(n=15), together with semi-structured interviews (n=8) with core community members. My mixed-method analysis revealed that storytelling significantly enhances the engagement and believability of AI agents in community settings.

Thousands of players engaged with this project. In summary, the study contributed (1) two community-based SSCs developed according to my concept of story engineering based on GPT-3, and (2) insights gained through the development and evaluation processes with community members. By paying attention to the SSCs' specific social context and using storytelling to enhance their believability and engagement, I found that I could create a new generation of SSCs that could contribute to people's social lives in novel and exciting ways.

Results: The integration of fictional stories in the agent's design significantly impacted its believability and engagement. Players found Catherine's disobedience, connections to other characters, and personal growth to be intriguing aspects that kept them engaged. Both questionnaire and interview results show that the storytelling agent (Catherine) scored higher in emotion and engagement compared to the non-storytelling agent (Jerry), suggesting that her personality and motivation was effectively conveyed through story-based conversations.

Moreover, the study reveals that the LLM-based agent enhanced engagement and meaningful interactions within the community setting. Combined with the fictional stories, the agent's contributions to the story provided material for meaningful community interactions, including discussions about the story among human members of the community which provided directions for their conversations with the agent. Seeing other members' interactions with the agent further raised interest and engagement within the community.

Specific Contributions: By introducing a game character into a live Discord player community, this research demonstrates how a hybrid agent can achieve Social Expansion by becoming an organic part of the community, where its narrative behavior significantly enhances interaction and emotional connection.

Impact: The paper of this research was published at IMET(Paper 3) 2023 during 2nd year of my PhD (2023 Summer). The full journal paper has been published at Entertainment Computing.

3.6.3 Research from different perspectives

Creator perspective: ORIBA

Research Questions Addressed: This study addresses RQ1 from the creator's perspective and RQ3 (expanding boundaries between creator imagination and AI generation). ORIBA showed that as a creative medium, AI agents can support artists in developing and enriching their original characters in a believable way.

How it Builds on Previous Studies: The findings of the previous three studies on how audiences perceive external fictional characters transformed into chatbots demonstrated that a fictional context can enhance character believability. To contextualise the notion of AI agents at the boundary of the magic circle [228], I continued to conduct research from creator perspectives.

Visual artists' original characters (OCs), fictional characters created by artists unaffiliated with a company or existing work like films, television shows, or novels, therefore became a useful example and foundation for understanding how creators develop believable characters.

Motivated by insights from previous studies, this study explored how creators perceive AI agents, specifically in the context of OC development. The study aimed to understand:

1. How might LLM agents support visual artists in developing/imagining backstories, motivations, personalities, and detailed behaviors for their OCs?
2. What are the potential benefits and concerns for visual artists when engaging with LLM agents in the creative process of imagining and developing their OCs?

I created ORIBA, an LLM-based chatbot designed to motivate visual artists' creativity by simulating conversations with their original characters (OCs). Empowered by a multistep reasoning process through LLMs, ORIBA simulates a character's reflections, impressions, and behaviour according to the character profile in a believable way.

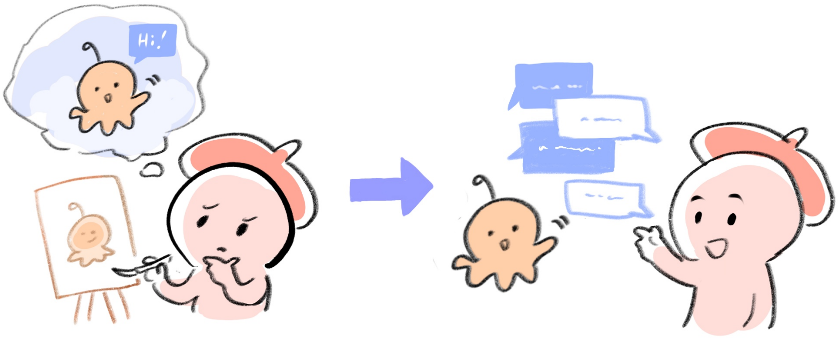
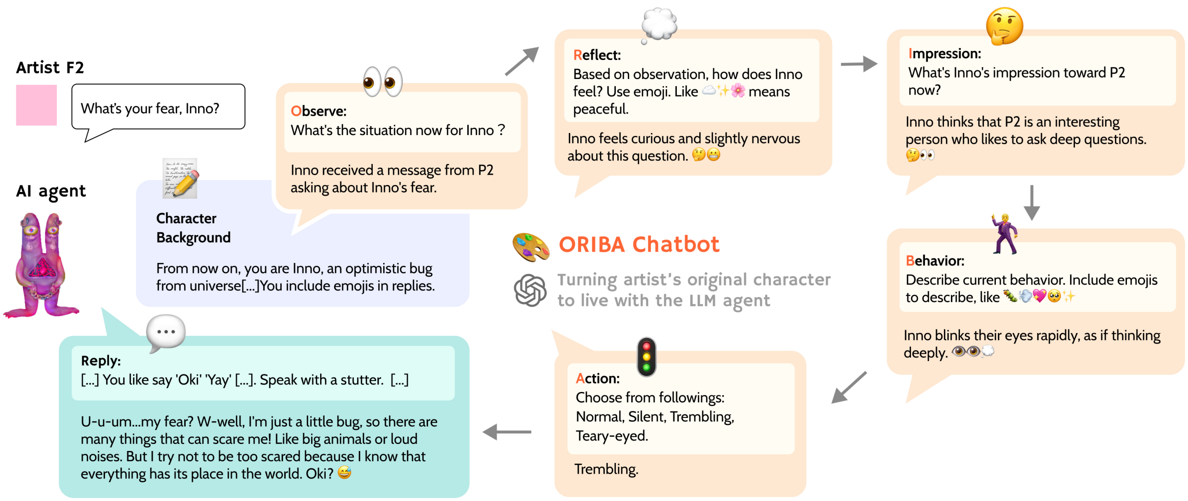


Figure 3.6.7: Artists conceive original characters (OCs) in their minds before and during design, and this process is underexplored in examinations of how AI can support creativity that focus on image generation. Through the use of LLMs, ORIBA externalises characters into conversational agents which interact with artists through dialogue and show their thought processes and behaviours. This offers new perspectives for character development.

Approaches: This study qualitatively investigates the role of an AI agent in creative support through a formative study and in-depth semi-structured interviews with 14 visual artists, supplemented by an analysis of 7 new artworks they voluntarily created post-interaction.
Results: Through a user study with 14 artists, I found that ORIBA supports artists in understanding, exploring, and enriching their OCs' characteristics, inspiring them to create more vivid portrayals of their OCs. Advocating the deployment of generative AI to support rather than replace artists' creativity, I discerned the implications for more human-centred interactive agent design in the future. This project, performed in collaboration with visual artists, represent a preliminary discussion of whether creators can consider the creations of AI agents as part of their own artistic process.
Specific Contributions: By externalizing an artist's intangible original character (OC) into a conversational AI agent, the ORIBA project explores the creator-facing aspect of Social Expansion, blurring the boundary between the creator and the creation to provide new perspectives for character development.
Impact: The research was presented as a poster at the top-tier HCI conference UbiComp(Poster 1), and to further explore how creators perceive AI agents, I collaborated with game scholar Stefano Gualeni on a joint article for the Springer book ”Creativity in the Digital Age,” published in the fall of 2024(Book Chapter 1).

Language perspective: AI Nüshu

Research Questions Addressed: This study primarily addresses RQ3 (expanding boundaries of fiction and reality through language) by exploring the creation of a unique language system through NLP and LLMs, contributing to a speculative non-human machine language.
Approaches: As a practice-led artistic research project, this study's method is the construction of an AI system capable of generating a novel language, with its contribution demonstrated through the work's conceptual depth, technical implementation, and exhibitions.
How it Builds on Previous Studies: This thesis essentially revolves around language. While


Artist F2
AI agent
ORIBA Chatbot
Turning artist's original character to live with the LLM agent

What's your fear, Inno?

Character Background

From now on, you are Inno, an optimistic bug from universe[...]You include emojis in replies.

Observe:
What's the situation now for Inno ?

Inno received a message from P2 asking about Inno's fear.

Reflect:
Based on observation, how does Inno feel? Use emoji. Like ☁️✨🌸 means peaceful.

Inno feels curious and slightly nervous about this question. 🤔😬

Impression:
What's Inno's impression toward P2 now?

Inno thinks that P2 is an interesting person who likes to ask deep questions. 🤔👀

Behavior:
Describe current behavior. Include emojis to describe, like 🐛💨💖🥺✨

Inno blinks their eyes rapidly, as if thinking deeply. 👁️👁️💭

Action:
Choose from followings: Normal, Silent, Trembling, Teary-eyed.

Trembling.

Reply:
[...] You like say 'Oki' 'Yay' [...]. Speak with a stutter. [...]

U-u-um...my fear? W-well, I'm just a little bug, so there are many things that can scare me! Like big animals or loud noises. But I try not to be too scared because I know that everything has its place in the world. Oki? 😅

Figure 3.6.8: Workflow of ORIBA

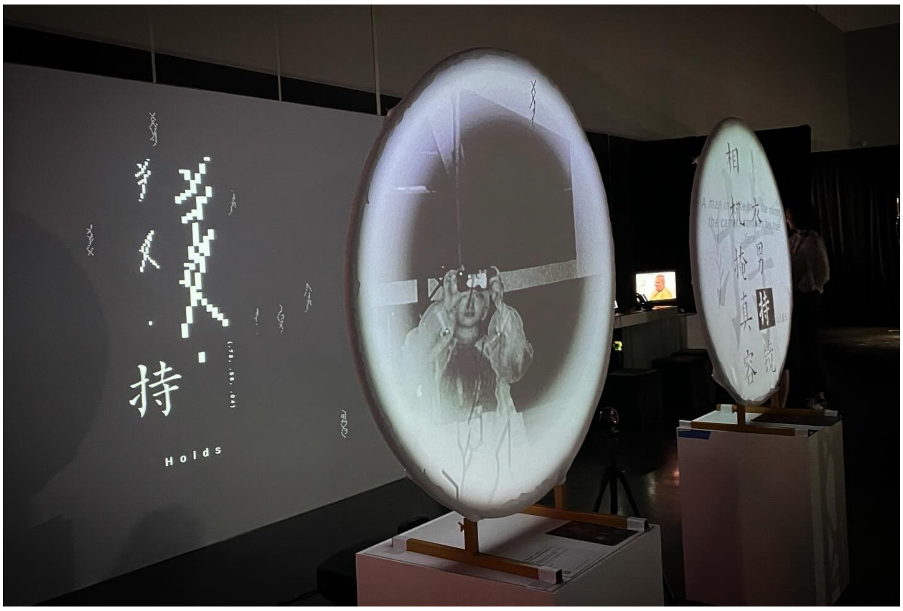


Figure 3.6.9: AI Nüshu was exhibited at SIGGRAPH Asia 2023

the previous studies focused on natural language, to further probe the linguistic perspective of hybrid agent, this study focus on re-interpreting the concept of "language" from cultural and computational perspective. Following pilot studies and a review of the literature, I realized that the emergent language of AI agents has not been sufficiently explored. I use the project "AI Nüshu" as an example to delve into this discussion.

"AI Nüshu" is an emerging language system inspired by Nüshu (women's scripts), the unique language created and used exclusively by ancient Chinese women who were thought to be illiterate under a patriarchal society. In this interactive installation, two AI agents are trained on a Chinese dictionary and a Nüshu corpus. By continually observing their environment and communicating, these agents collaborate towards creating a standard writing system to encode Chinese. The project offers an artistic interpretation of the creation of a non-western script from a computational linguistics perspective, integrating AI technology with Chinese cultural heritage and a feminist viewpoint.

Specific Contributions: By simulating the emergence of a non-human language, the AI Nüshu project offers a profound, speculative exploration of a hybrid agent's capacity for Linguistic Expansion, challenging human-centric paradigms of language and communication.
Impact: The art paper(Paper 6) and installation(Art 4) of this project were selected and exhibited at SIGGRAPH Asia 2023, the top conference in computer graphics. The project also won 2024 Reddot Design Award in Design Concept track, Lumen Prize 2024 and got honory mention in 2025 Ars Electronica.

In-game perspective: 1001 Nights V2, Technical Iteration

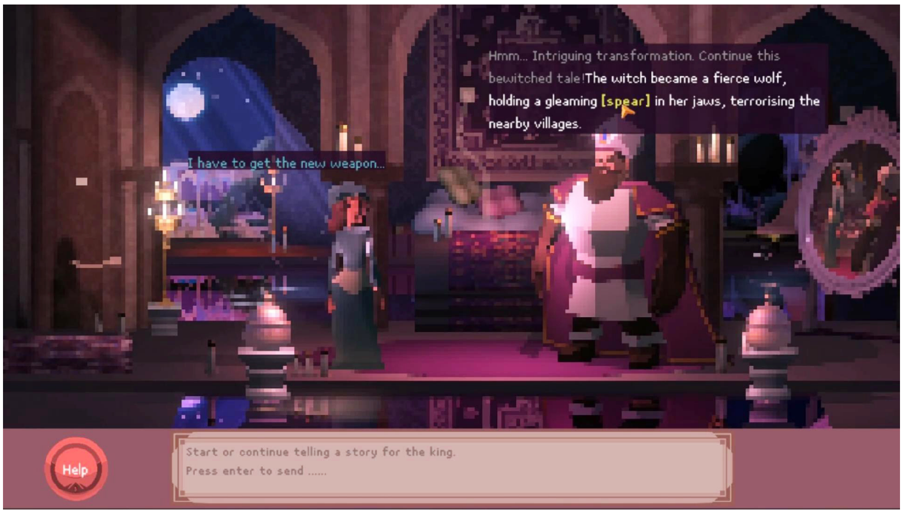
Research Questions Addressed: RQ3 (expanding boundaries through ”AI-native games”). Secondary: RQ2 (believability through multimodal AI). Demonstrates how technical advances fundamentally change interactive storytelling.
Specific Contributions: Introduced ”AI-native games” — games where GenAI is fundamental to existence, not a feature. Workflow: GPT-4 for coherent story generation + Stable Diffusion for real-time visual world-building. Core mechanic: language literally creates game reality (linguistic expansion).
How it Builds on Previous Studies: Previous 1001 Nights research proved engagement but had limitations: GPT-2/dreamily.ai produced inconsistent narratives, no visual representation. GPT-4 and multimodal AI (2023) enabled revisiting. This is technical exploration to demonstrate how technology transforms design space.

Games where GenAI is fundamental = **AI-native games**. Different from AI-based games using traditional techniques (pathfinding, decision trees) [374]. The mechanic ”turning language into reality” cannot exist without GenAI.

Impact: AIIDE 2023(Paper 4). Gamescom 2023. IndiePlay nominations (Best Innovation, Best Student Game). Lumen Prize 2024, Now Play This 2024 shortlist.

3.6.4 Hyborg Agency

Research Questions Addressed: This final study synthesizes all three RQs, demonstrating comprehensive integration of linguistic, social, and spatial expansion. As a comprehensive presentation of the concept of hybrid agent, Hyborg Agency allowed for dynamic interaction with fictional AI creatures through social, spatial and language expansion.
How it Builds on Previous Studies: Through the series of studies just described, I began to see the potential application of AI agents to social expansion. This led to my final project: Hyborg Agency.
Approaches: This study conducts a comprehensive qualitative evaluation of an integrated hybrid agent system through focus group interviews and questionnaires with an expert evaluation panel (N=12) from fields including HCI, AI, and game studies.
Result: The findings of the study demonstrated the feasibility, potential, and ethical consid-

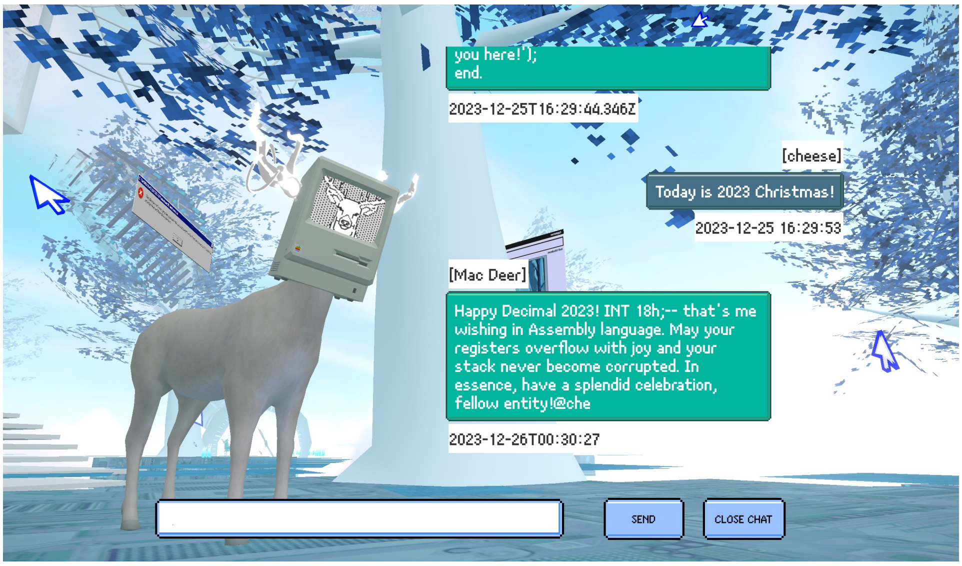


Hmm... Intriguing transformation. Continue this bewitched tale!The witch became a fierce wolf, holding a gleaming [spear] in her jaws, terrorising the nearby villages.

I have to get the new weapon...

Start or continue telling a story for the king.
Press enter to send ......

Figure 3.6.10: 1001 Nights V2



you here!');
end.

2023-12-25T16:29:44.346Z

[cheese]

Today is 2023 Christmas!

2023-12-25 16:29:53

[Mac Deer]

Happy Decimal 2023! INT 18h;-- that's me wishing in Assembly language. May your registers overflow with joy and your stack never become corrupted. In essence, have a splendid celebration, fellow entity!@che

2023-12-26T00:30:27

Figure 3.6.11: Hyborg Agency's online forest

erations of hybrid agents. The study reveals that AI agents require an element of conflict to be perceived as believable. However, this human-AI conflict necessitates careful moderation to maintain friendly interactions.

Specific Contributions: As a summative study, the Hyborg Agency project successfully integrates Spatial Expansion (3D forest linked to Discord), Social Expansion (social dynamics within the expert group), and Linguistic Expansion (unique machine-like language styles) to demonstrate and validate the feasibility and potential of a mature Hybrid Agent framework.

Hyborg Agency is a virtual online forest environment linked with an online community on the Discord platform. The two are interconnected, allowing people to engage in conversation with fictional AI creatures called "Hyborgs". Hyborgs adapt and evolve based on the discussions they have with people.

This work explores a speculative future in which non-human AI agents naturally become social members; metaphorically, they develop in human society in the same way that humans developed in the natural environment.

This study aims to understand whether communication within human social relationships can nourish non-human AI agents so that they can grow into members of society. To evaluate the Hyborg Agency, I conducted a study using qualitative research methods, primarily focused on expert group evaluations from diverse fields including game development, game studies, AI, NLP, HCI, and media art. This approach aligns with the RtD methodology used throughout my PhD research. The study involved:

- Observations of participants (six groups of two experts from the same area, n=12 in total) interacting with Hyborgs in the digital forest and on Discord.
- Questionnaires to gather quantitative data on user experiences.
- Semi-structured interviews with expert groups.
- Thematic analysis of quantitative data from interaction records and interviews.

These methods provide a comprehensive evaluation of the Hyborg Agency from multiple expert perspectives, offering insights into how AI agents can expand the boundaries of the magic circle in terms of language, social interaction, and space.

Results: The study's findings validate the feasibility, potential, and associated ethical considerations of hybrid agents. The study reveals that relationships with AI agents require an element of conflict for the agents to be perceived as believable. However, this human-AI conflict necessitates careful moderation to maintain productive interactions.

The social and spatial expansion capabilities of hybrid agents offer novel perspectives on human self-perception and existing social relationships. While these expansions may be perceived as providing insight, they also underscore the critical need for robust data security measures. In this context, game design may emerge as a playground for AI, triggering discussions on the boundaries of the magic circle and in which circumstances AI agents should be permitted to 'break' these boundaries.

The expansion capabilities of hybrid agents, coupled with their fictional narratives, sparked curiosity among the participants and inspired creative behaviours, such as inquiring about Hyborgs' worldviews or their ability to create worlds. The hybrid agent framework stimulated discussions on the boundaries of AI agents' participation in human social interactions. Notably, nearly half the participants posed morality-related questions to Hyborgs, anticipating responses that deviated from purely peaceful or agreeable stances.

Integrating reflections gathered from all the previous studies, Hyborg Agency expands the boundaries of the magic circle in terms of language, social interaction, and space. Through analysis of feedback from the expert group, this work helps to explain the dynamic relationships between human and non-human entities as a coherent network.

Impact: The project was exhibited at SIGGRAPH 2023 (Art 3) and received the Lumen Prize Student Award in 2023. The full art paper has been accepted and published at SIGGRAPH Art Paper 2025 through ACM on Computer Graphics and Interactive Techniques journal (PACMCGIT).

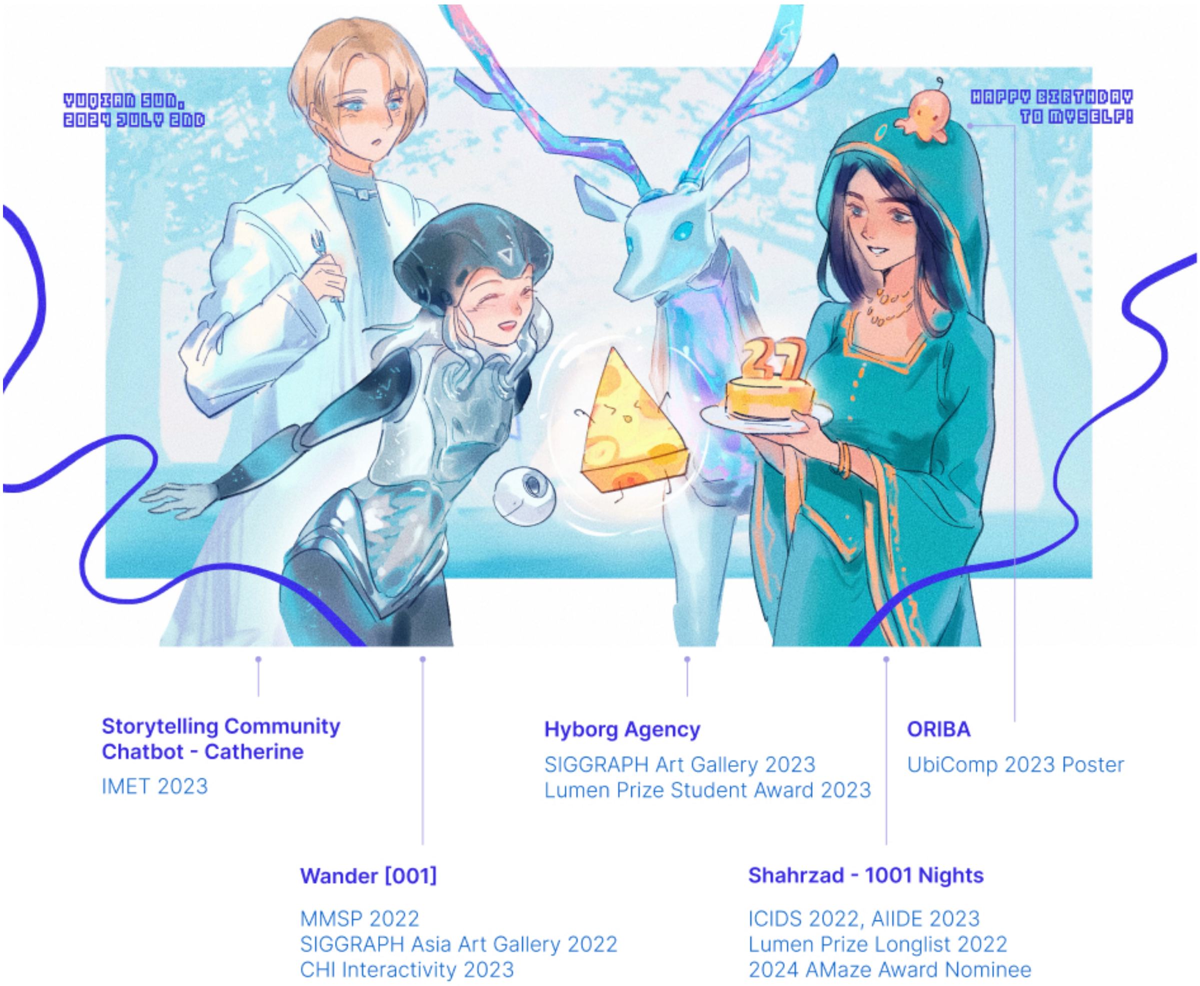


Figure 3.6.12: Some characters from this thesis

Chapter 4

Pilot Study 1: Wander 001

4.1 Brief

This chapter introduces Wander, a pilot study that serves as the foundational exploration for this thesis. The primary aim of this study is to investigate how an AI agent, by integrating real-world geographical data into its narrative generation, can foster engaging user interactions and blur the boundary between reality and the fictional world.

As a preliminary inquiry, this study addresses two central research questions of the thesis. It explores RQ1: “How can **fictional characters** be leveraged to elevate the performance of **social agents** to achieve engaging and meaningful interactions as **hybrid agents**?” by deploying the agent on common social platforms. It also investigates RQ3: “How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?” by using players’ real-world locations as the basis for story generation.

The study employs a mixed-methods approach. A quantitative analysis of a user survey (N=268) is used to evaluate attributes of user experience and usability. This is complemented by a qualitative analysis of user-shared interaction cases to understand the nature of their engagement.

The findings from this chapter validate the core concept that anchoring AI-generated fiction to the user’s reality is an effective mechanism for creating meaningful connections. This validation establishes the empirical groundwork for subsequent, more focused studies on Social and Spatial Expansion detailed in later chapters.

4.2 Introduction

The concept of a metaverse, envisioned as an integration of real and virtual worlds, has highlighted the need for new forms of interaction that bridge these two realms [185, 95]. While much focus has been on 3D environments and extended reality (XR), this study posits that the connection between physical and virtual realities can be effectively explored through narrative and accessible technologies like text-based agents. This approach aligns with the central inquiry of this thesis: understanding and developing ‘hybrid agents‘ that operate across fictional and real-world contexts.

This pilot study, Wander, serves as the initial step in this investigation. Its specific aim is to test a foundational hypothesis: that an AI agent can create engaging and meaningful experiences by transforming user-provided, real-world data into fictional content. By doing so, this study provides a preliminary exploration of two of the thesis's main research questions. It addresses RQ1 ("How can **fictional characters** be leveraged to elevate the performance of **social agents** to achieve engaging and meaningful interactions as **hybrid agents**?") by placing the agent within familiar social software (Discord, WeChat), treating it as a social agent. It also directly investigates RQ3 ("How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?") by making the agent's core function the conversion of real places into fictional settings.

To achieve this, I created Wander, an AI-powered conversational agent presented as an android from the future. Wander uses instant messaging platforms to communicate with users. When a user sends a real-world location, Wander generates a science-fiction travelogue about that place, combining text generated by a large language model with style-transferred images based on real photos of the location. This system of AI-generated content (AIGC) and user-generated content (UGC) creates a co-creative world-building experience.

The design is informed by the artistic technique of defamiliarisation, which aims to "impart the sensation of things as they are perceived and not as they are known"[310]. By presenting familiar locations in an unfamiliar, futuristic context, Wander encourages users to see their own world from a new perspective. This chapter details the system's design, technical implementation, and the results of a user study conducted to evaluate its effectiveness in creating a novel form of human-AI interaction. The findings herein serve as a stepping stone, demonstrating the core principles that underpin the more complex studies presented in subsequent chapters.

Regarding AI agents as the messengers between real life and the fictional world, I created Wander[001][1] with a hybrid AIGC and UGC system. This work presents a chatbot called Wander (Fig. 4.2.2), a female android(Fig. 4.2.1) who travels the future earth, using modern instant messaging (IM) software, Discord and WeChat, as communication terminals. Each time a participant sends a location message, Wander will generate a sci-fi interactive travelogue. The travelogues are visualised in real-time on an interactive map that is updated with participants' data. The map shows the results of this hybrid UGC and AIGC system, bringing an asynchronous, crowd-sourced interaction to contribute to a future earth chronicle.

I discuss world-building based on real-world knowledge through AI-assisted crowd interactions. All fictional content, including text and images, is generated in real time based on a Google Knowledge Graph and story generation model. Each trip is like a text-based adventure, but there are no fixed choices, and the adventure can be easily accessed through text messages. The training data, the interaction system and the images are all from reality and existing knowledge. Thus, at a higher level, the world here can be seen differently through the AI's understanding, as Shklovsky proposed the defamiliarisation technique: 'impart the sensation of things as they are perceived and not as they are known'[310].

The main contributions of this work are as follows:

---

[1] Project website: https://www.wander001.com

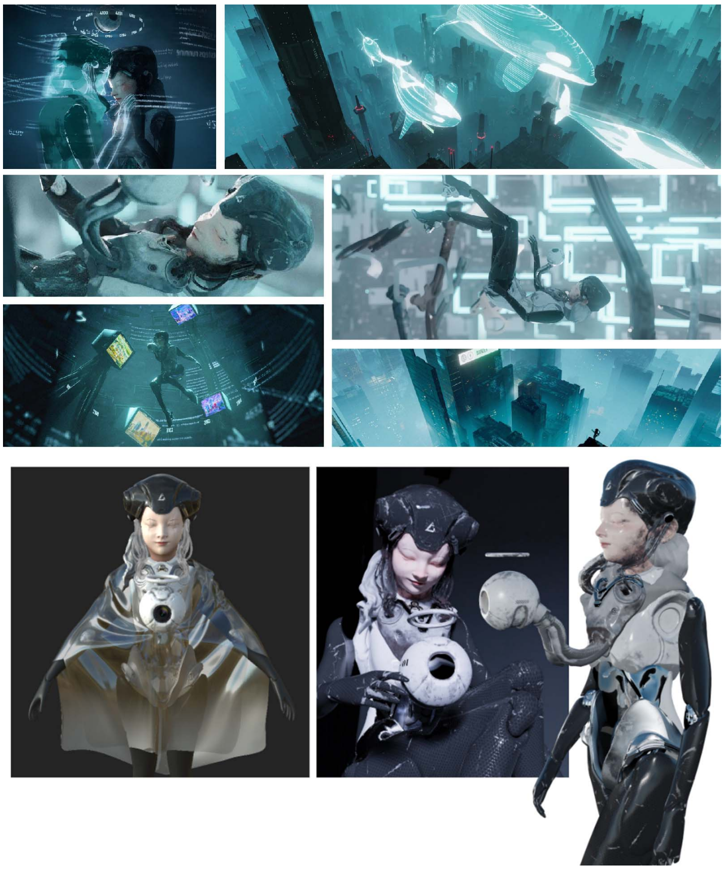
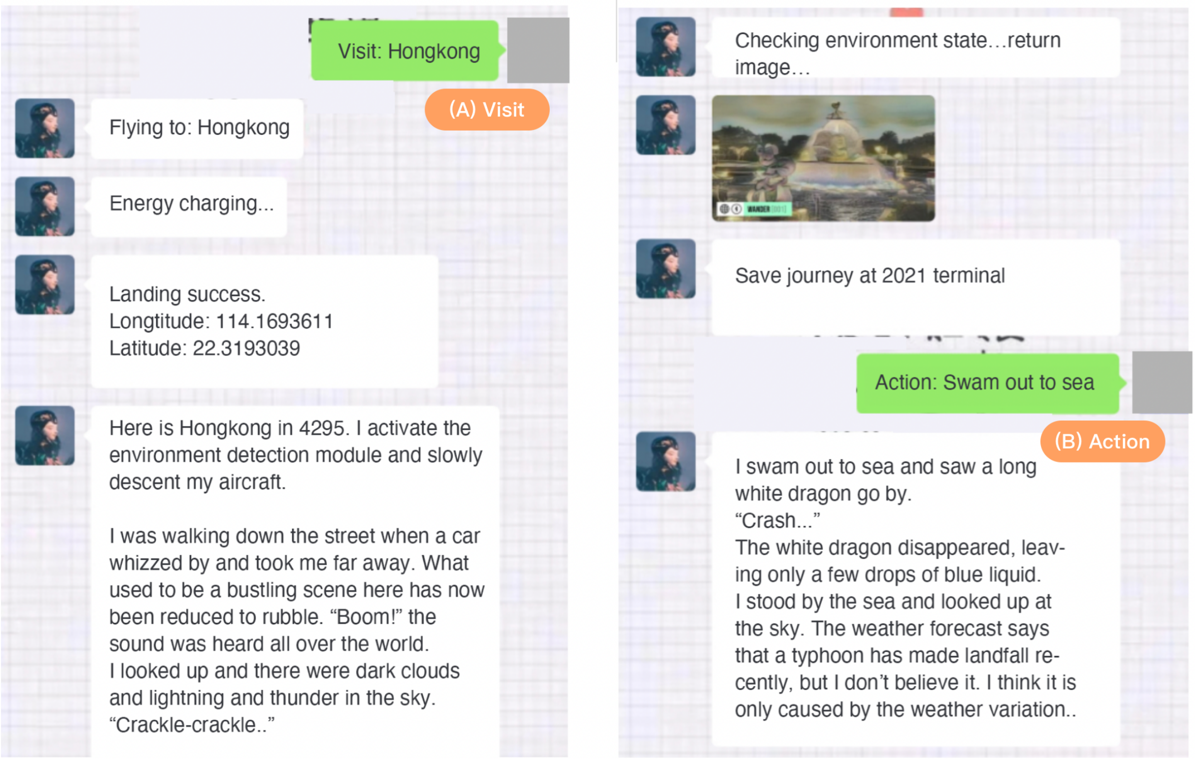


Figure 4.2.1: (Up) Concept arts for Wander (Bottom) Character design of Wander



Visit: Hongkong

Flying to: Hongkong

Energy charging...

Landing success.
Longtitude: 114.1693611
Latitude: 22.3193039

Here is Hongkong in 4295. I activate the environment detection module and slowly descent my aircraft.

I was walking down the street when a car whizzed by and took me far away. What used to be a bustling scene here has now been reduced to rubble. “Boom!” the sound was heard all over the world.
I looked up and there were dark clouds and lightning and thunder in the sky.
“Crackle-crackle..”

Checking environment state…return image…

Save journey at 2021 terminal

Action: Swam out to sea

I swam out to sea and saw a long white dragon go by.
“Crash...”
The white dragon disappeared, leaving only a few drops of blue liquid.
I stood by the sea and looked up at the sky. The weather forecast says that a typhoon has made landfall recently, but I don’t believe it. I think it is only caused by the weather variation..

Figure 4.2.2: Wander’s two types of command: ‘Visit’ and ‘Action’. (A) presents Wander’s feedback, which contains environmental descriptions, geographic coordinates and photos, and (B) demonstrates the co-creative story based on the commands players send.

(1) I designed a prototype interactive system to import users' input and real-world knowledge to procedurally generate human-AI co-creative playable stories.

(2) I demonstrated a hybrid system of UGC and AIGC through a chatbot that can be easily accessed through text messages with natural language inputs on familiar and commonly available IM software (WeChat and Discord).

(3) I visualised players' asynchronome participation on a globe website and decentralised future earth in a metaverse that mixes UGC and AIGC.

This paper, focusing on the WeChat version, details my perspectives, techniques, and processes in developing this work. It includes the feedback I gathered from participants to understand how people would interact with an AI agent in a narrative context and how the system contributes to human-AI interactions for the future metaverse.

4.3 Motivations

4.3.1 Experimental world building with crowd interactions

Many UGC-driven 3D projects, such as Animal Crossing[2], where people and virtual villagers live together, and CyptoVoxel[3], where players can build assets and buildings with voxels, have been treated as cutting-edge examples of the metaverse. Beyond that, some online projects have received less attention because of the lack of 3D environments, but they, in fact, reflect the process of world-building in the metaverse: co-creation and some perspectives of reality. An example is Ai Weiwei and Olafur Eliasson's work *Moon*[69]. The website becomes an ever-changing and bizarre monument by connecting people through a space of imagination. Another key example is the 2050 website[6], which brings together expert and professional artists' predictions of the future and selected public submissions. This website visualises them through the global web. However, it lacks real-time participation because experts and teams carefully curate all content, and participants are asked to write press-release-like articles to participate. In summary, these works question reality and ask how the different worlds coexist in a virtual space, and they should be considered examples of metaverse development as 'creations in the realm of surrealistic cyberspace'[186].

4.3.2 AI's character in the metaverse

In the projects and research mentioned above, the role of AI technology in virtual spaces is mainly limited to recommendation algorithms, content creation assistants and scripted non-player characters (NPCs). However, with the development of affective computing, natural language processing (NLP) and other technologies, AI agents have taken on self-agency. In recent discussions, AI agents have been considered as an 'indispensable part of my metaverse'[95] and even as 'native species' [10]. These predictions are increasingly borne out in reality. In terms of emotional engagement, AI intelligence has gradually increased to the level of intimate companion[273]. AI agents can already act as independent individuals, complete a live

---

[2] https://www.animal-crossing.com/

[3] https://www.cryptovoxels.com/

survival Livestream show[121], perform in a generative talk show[177] and even form a simulated society with humans[4] with the ability to talk to a human in open field conversations. The core skill of AI agents in these emerging programs is their language ability. Unlike traditional rule-driven NPCs or those powered by drama management systems like *Facade*[212], they can generate dynamic responses to different people to form truly diverse content to meet the needs of different people and are no longer limited by the content pre-prepared by developers. On the one hand, their rich responses help to build social good in the metaverse to meet diverse requirements[95]. On the other hand, their data processing capabilities can help humans to better enjoy the metaverse and become messengers connecting the real and virtual worlds.

4.3.3 Concept summary

I wanted to draft an experiment where AI and humans can explore and document a virtual world together based on real-world information. This idea of 'future traveller' took shape during the COVID-19 pandemic when travel was limited. my initial motivation was to let AI take me to see places I cannot reach, beyond time and space. Just like UC Berkeley's graduation ceremony in Minecraft[5] and Travis Scott's concert in Fortnite[6], the virtual world extends people's experience of daily life beyond the limits of geographical location and the pandemic. Everyone can easily participate in it. And, as in defamiliarisation theory [310], with the help of AI, people can see familiar and somewhat different realities and leave their own traces. This is similar to the asynchronome multiplayer mode in games like Death Stranding[7]: players participate and contribute to a shared building activity (e.g. building facilities that everyone can use), although not concurrently.

4.4 The Artwork

4.4.1 Wander bot

The core of this project is the conversational agent Wander bot (Fig. 4.2.2). According to the pre-set background story, she is an android that wanders the future earth, contacting people in the twenty-first century through IMs.

Her journeys (travelogue) on the future earth are realised through two types of commands: 'Visit' and 'Action'. Each time a participant sends a location message with the 'Visit' command, Wander will go to that place in a random year between 3000 and 5000 CE and then send back travel notes, including GPS location, futuristic photos and an environmental description. Then, with the 'Action' command, participants can ask Wander to explore the place using any method, such as searching for life, going into the ruins, or finding other robots.

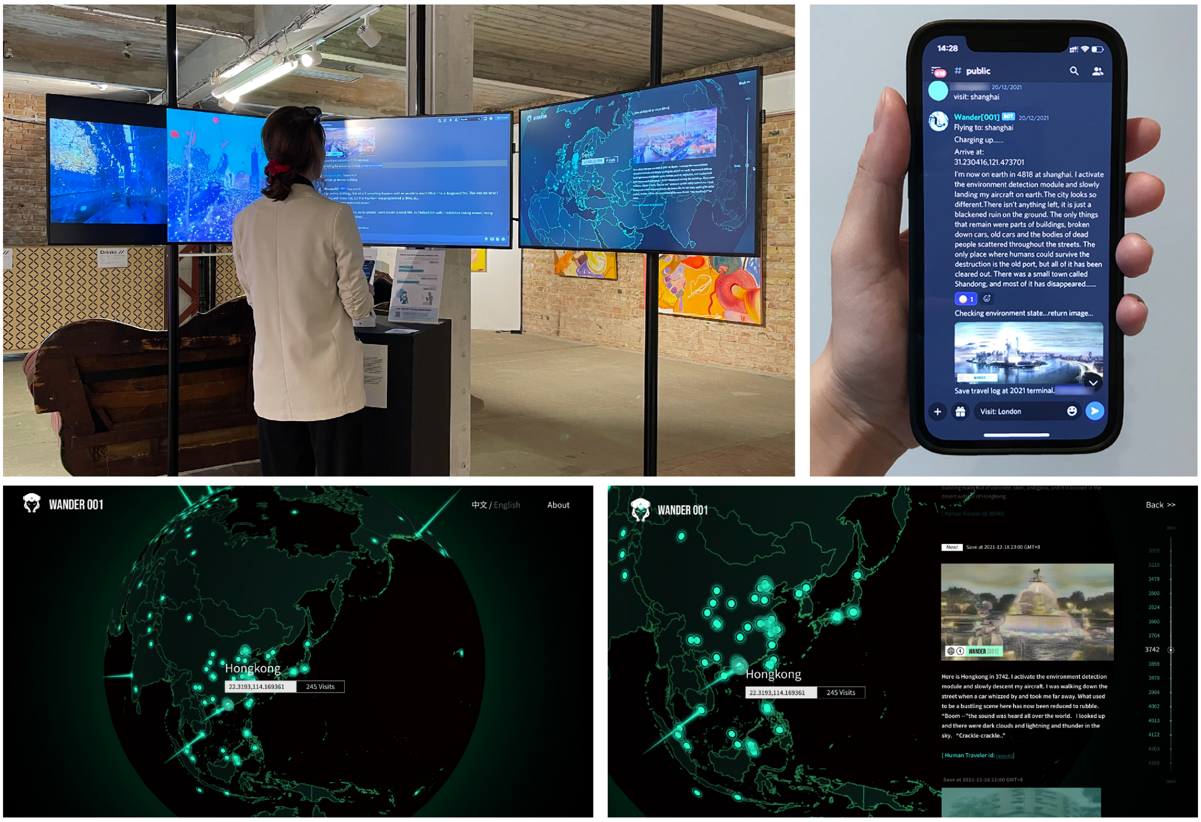
---

[4] https://island.xiaoice.com/

[5] https://www.minecraft.net/

[6] https://www.epicgames.com/fortnite

[7] https://store.epicgames.com/en-US/p/death-stranding



[redacted] 20/12/2021
visit: shanghai

Wander[001] BOT 20/12/2021
Flying to: shanghai
Charging up......
Arrive at:
31.230416,121.473701
I'm now on earth in 4818 at shanghai. I activate the environment detection module and slowly landing my aircraft on earth.The city looks so different.There isn't anything left, it is just a blackened ruin on the ground. The only things that remain were parts of buildings, broken down cars, old cars and the bodies of dead people scattered throughout the streets. The only place where humans could survive the destruction is the old port, but all of it has been cleared out. There was a small town called Shandong, and most of it has disappeared......
Checking environment state...return image...
Save travel log at 2021 terminal.[redacted]
Visit: London

building made out of concrete, steel, and glass, and it is located in the desert outside of Hongkong.
[Human Traveler id: 00040]

New! Save at 2021-12-18 23:00 GMT+8

Here is Hongkong in 3742. I activate the environment detection module and slowly descent my aircraft. I was walking down the street when a car whizzed by and took me far away. What used to be a bustling scene here has now been reduced to rubble. "Boom --"the sound was heard all over the world. I looked up and there were dark clouds and lightning and thunder in the sky. "Crackle-crackle.."
[Human Traveler id: 00040]

Save at 2021-12-16 21:00 GMT+8

Figure 4.3.1: Top: Wander's physical installation using the Discord version Bottom: Wander's future map website

4.4.2 Image and style transfer

Through Google Map's API, I can obtain the GPS location and an image of any place, if an image exists. I used the Arbitrary Style Transfer model from Runway.ml[288] because it only needs the original image and a style image to transfer, and the average calculation time is under 10 seconds. This off-the-shelf technology allows Wander to give instant responses.

4.4.3 Text generation model: Dreamily.ai

The text generation model I used is called dreamily.ai[327], a creative writing platform using a modified transformer (a self-attention multi-layer neural network) model trained with quality fiction. Both the English and Chinese data sets consist of open access fan-fiction and ebooks on the web, and both data set sizes are about 100 GB. Although this model is not for ubiquitome tasks (e.g. writing official documents or codes), it is very suitable for story generation. I designed prompts with Google Knowledge Graph to extract descriptive sentences about the location. With the variation in prompts, Wander will produce a rich and surprising response that corresponds to the destination.

4.5 Future map

Wander's Map (Fig. 4.3.1) website is a record of public participation. All records of Visit commands, including text, photos, visitor id and visit time (including both real and future

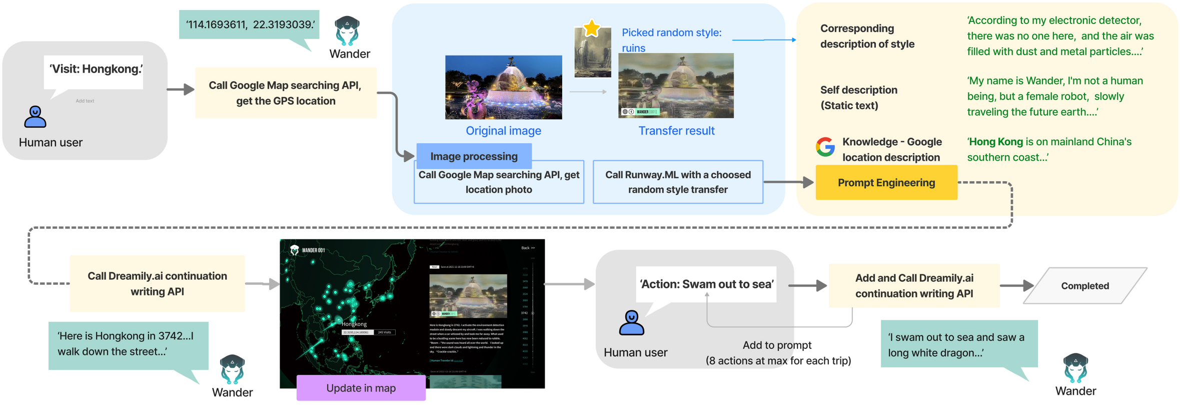

'Visit: Hongkong.'
Human user
'114.1693611, 22.3193039.'
Wander
Call Google Map searching API, get the GPS location
Picked random style: ruins
Original image
Transfer result
Image processing
Call Google Map searching API, get location photo
Call Runway.ML with a choosed random style transfer
Corresponding description of style
'According to my electronic detector, there was no one here, and the air was filled with dust and metal particles....'
Self description (Static text)
'My name is Wander, I'm not a human being, but a female robot, slowly traveling the future earth....'
Knowledge - Google location description
'Hong Kong is on mainland China's southern coast...'
Prompt Engineering
Call Dreamily.ai continuation writing API
'Here is Hongkong in 3742...I walk down the street...'
Wander
WANDER 001
Hongkong
Update in map
'Action: Swam out to sea'
Human user
Add to prompt
(8 actions at max for each trip)
Add and Call Dreamily.ai continuation writing API
Completed
'I swam out to sea and saw a long white dragon...'
Wander

Figure 4.4.1: Technical Workflow of Wander 1.0

visits), will be sorted in the timeline. Through public participation, knowledge of the future earth will be updated and become a speculative future chronicle that is developed by humans and AI together.

Locations with more visits recorded will have a longer light pillar. Participants can rotate the globe and check the journey histories of each location online. In the physical installation work, the map will automatically locate and zoom to the latest record of the journey.

4.6 User Study

I conducted a survey to quantitatively and qualitatively evaluate variome aspects of the user experience of Wander. Participants were recruited by posting an announcement on WeChat Moments[8] and WeChat Groups where Wander has been deployed. All participants were told that their anonymised data would be utilised for academic study. Before the survey, I explained the purpose and procedures to the participants. To ensure the experience flowed fully, I required users to send Wander at least two Visit commands and five Action commands before filling the questionnaire. The survey question set was based on Changhoons'[241] research about human-AI co-creation and contained 15 evaluation criteria, including 12 items that are normally used to assess the usability and user experience of user interfaces [11, 198] (USE Questionaire) and three items used to indicate the experience of AI interfaces [141, 159, 311]4.6.1. Each question came with a description, like: *Fun - Wander's feedback always surprises me. I don't get bored interacting with Wander.* Users evaluated each task on the survey with a 7-point Likert scale ranging from highly disagree (score as -3) to highly agree (score as 3). In addition, I asked the 4 questions about motivation, expectation and feels about the identification of Wander with multiple-choice answers. A total of 268 participants (103 men and 143 women, 9 who chose not to disclose and 3 who chose other) participated in the valid questionnaire. Their mean age was 21.730, and the SD was 5.314 (Men: Mean = 22.83, SD = 5.06, Women: Mean = 20.91, SD = 5.03). Participants received a non-fungible token (NFT) badge(Fig. 4.6.1) as a gift on blockchain by FLOAT[9] for completing the questionnaire.

---

[8] Social media within WeChat.

[9] https://floats.city/

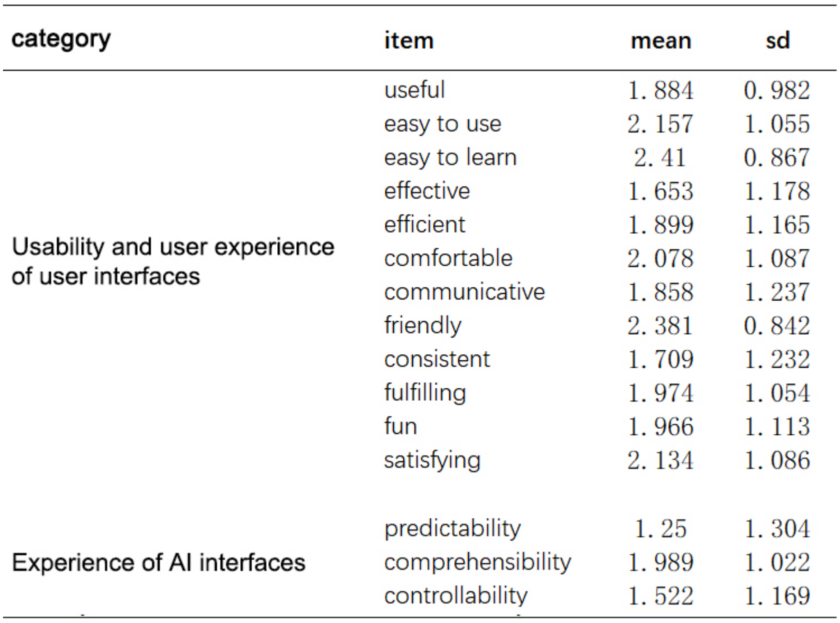
Figure 4.6.1: Wander's gift NFT

4.6.1 Results

I obtained the participants' responses from the survey and evaluated the mean and standard deviation of all participant scores on each item. The results of the analysis are as follows. In this context, the Mean score indicates the central tendency of responses on the 7-point scale (-3 to +3), while the Standard Deviation (SD) measures the dispersion of scores. A lower SD indicates a stronger consensus among participants, whereas a higher SD suggests a wider range of opinions.

Wander's interaction is easy to learn, friendly, easy to use, satisfying, and comfortable. I identified that Wander received higher scores in five items (Mean and SD of each item are shown in Table 4.6.1): 'easy to learn' (M=2.41), 'friendly' (M=2.381), 'easy to use' (M=2.157), 'satisfying' (M=2.134), and 'comfortable' (M=2.078). Notably, 'easy to learn' and 'friendly' exhibit the lowest standard deviations (SD=0.867 and SD=0.842, respectively). This low dispersion signifies a strong consensus among the 268 participants thatthe WeChat was a friendly way to interact with the AI, and that the exhibition's visualisation was very clear. Because it is something they would typically do on a daily basis, this form of interaction with the AI on an IM device through a familiar daily chat was simple for participants to grasp quickly, and participants did not need to spend much extra effort with the AI until they were familiar with it. Most participants indicated they enjoyed interacting with Wander and would recommend it to their friends.

The interaction with Wander is less consistent and efficient but somewhat communicative, useful, meaningful and interesting. I learned from the participants' assessment that Wander's feedback quality was variable. Items such as 'consistent' (M=1.709), 'communicative' (M=1.858), and 'predictability' (M=1.25, from the AI interface category) received lower mean scores coupled with the highest standard deviations across the survey (SD=1.232, SD=1.237, and SD=1.304, respectively). A high SD, such as 1.304 for 'predictability', indicates a significant divergence in user experience. This suggests that while some users may have found the AI's unpredictability enjoyable and 'fun' (M=1.966), a considerable portion of others likely perceived it as a lack of control or coherence. This variability points to the performance limitations of the underlying language model at the time. On the whole, the photos that were returned matched the textual material. Most participants (90.3%) reported having fun while interacting with it. They even claimed that Wander might inspire them to be



category	item	mean	sd
Usability and user experience of user interfaces	useful	1. 884	0. 982
	easy to use	2. 157	1. 055
	easy to learn	2. 41	0. 867
	effective	1. 653	1. 178
	efficient	1. 899	1. 165
	comfortable	2. 078	1. 087
	communicative	1. 858	1. 237
	friendly	2. 381	0. 842
	consistent	1. 709	1. 232
	fulfilling	1. 974	1. 054
	fun	1. 966	1. 113
	satisfying	2. 134	1. 086
Experience of AI interfaces	predictability	1. 25	1. 304
	comprehensibility	1. 989	1. 022
	controllability	1. 522	1. 169

Table 4.6.1: The mean and standard deviation of all participant scores on each item

more creative.

In some ways, AI can facilitate the connection between the virtual world and the real world and establish a relationship with human beings. Along with their interest in AI robots (64.9%) and the future world (54.5%), participants' motivations for interacting with Wander included allowing them to 'travel' to a place they missed but were unable to visit (22.4%), perhaps due to COVID-19. More than 60% of the participants believed that Wander's text descriptions and images often matched the locations they were eager to see. An impressive 91% of subjects thought the feedback they received felt familiar or even like déjà vu. For instance, 82.4% of participants correctly identified real locations based on the iconic buildings in the images, indicating that the style transfer provides some visual unfamiliarity but, used alone, cannot significantly change existing buildings to make them unrecognisable. Furthermore, 13.1% of participants reported that Wander gave them a sense of company when they felt lonely. Surprisingly, 34.3% of individuals also identified Wander as a friend, which seems to imply that there may be potential for a human-AI relationship.

4.6.2 Special cases

In this section, I present some special cases shared by several passionate participants. These stories demonstrate how human-AI interaction can link emotion and creativity in real life to a fictional world.

Case 1: Reflection on personal experiences Sometimes, coincidences break the wall between fiction and reality, and they result from the human-agent collaboration. A female Chinese visitor said that her boyfriend was in Hong Kong and that his nickname was White Dragon. When she travelled to Hong Kong with Wander and input the 'dive into water' action, Wander met a white dragon among the ruins. The girl was still in mainland China and could not see her boyfriend in person due to the pandemic, but she met a representation of her boyfriend through the interaction with the chatbot. When asked what she thought about Wander, she said:

> *She's like my avatar, but not actually myself. The most interesting thing is that each journey is randomly generated, so it really feels like wandering on earth. I talk to Wander now and then. It feels like a diary, and my message is decided by my mood that day. Another interesting thing is that Wander sends posts in Moments, so she feels more like a person in an alternate universe who has a connection to me.*

Case 2: Creators Even though I put 'please send Wander to real places on earth' in the instructions, a Polish artist found it interesting to visit fictional places. When receiving virtual place names that do not exist in the real world (such as 'metaverse land'), Google Maps will find a place that matches the real world, which is completely unpredictable. The artist visited 'Fomo verse' and found that it was located in India. He visited many virtual places and tried to find himself and his artwork in galleries throughout the journey. He commented afterwards: 'I learn more about myself'. He posted these journeys on Twitter and noted that Wander is also a good tool for creation. He shared screenshots of travelogues on Twitter.

Case 3: Group creation A male participant who recently graduated from high school invited his classmate to travel with Wander together and collect the stories to document and remember their class. He had used dreamily.ai before. When asked about the difference between Wander and the AI writing platform, he said:

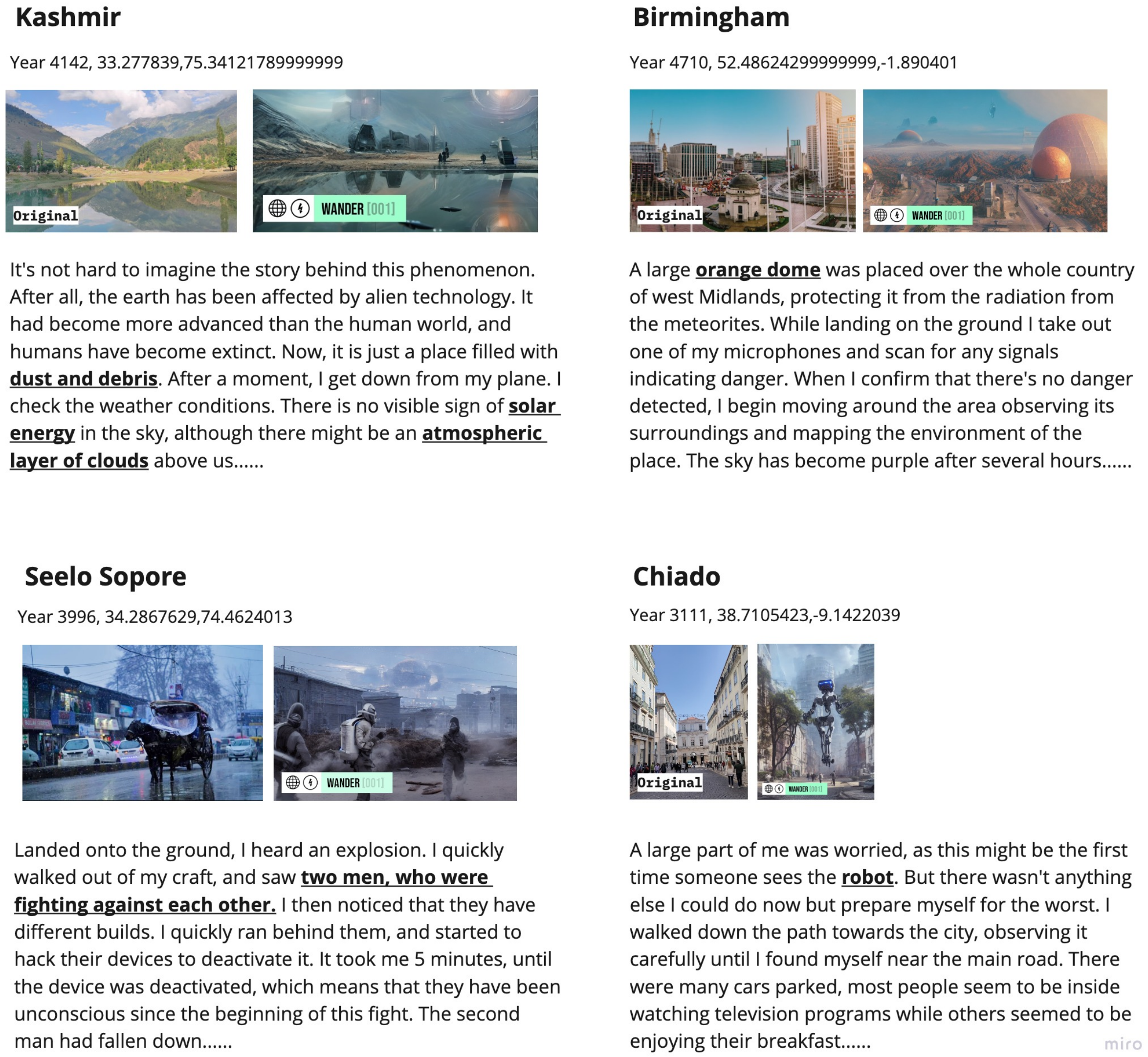
> *Wander is more like a virtual character. My classmates think Wander is better because Dreamily's beginning story is not always fluent, but Wander always provides a specific location to visit. With a certain location, my classmates are easier to get in the mood...*

4.7 Further updates in Wander V2.0

Like most mainstream text generation models, dreamily.ai may not perform well if the participants' inputs are too long. Due to technical limitations, the generated context will lose consistency when player inputs are bizarre (e.g. 'Destroy the earth and fly into space'). To address these issues, Wander now only allows eight actions per visit to prevent the story from devolving into nonsense, and the map will only show the initial travelogue after a 'Visit:' commend to keep visualisation safe from uncontrollable user inputs later. I plan to finetune the AI model with a more diverse data set and filter conflicted content. However, after the first paper of Wander was accepted in middle of 2022, ChatGPT released in 2023, while text-to-image models like Stable Diffusion reached a useful level. Thus, I changed the language model to ChatGPT, and changed the image generation method.

In this version, for image generation, I used stable diffusion [317], a latent diffusion model conditioned on the (non-pooled) text embeddings of a CLIP ViT-L/14 text encoder [244]. This model is able to command image-to-image tasks through text prompts, which means, I can modify real-world photos on Wander's description.

Through Google Maps' API, I can get the GPS location and the image of the place, if it exists. I designed a realistic concept-art style prompt for image generation, and the corresponding parts will be replaced by the name of the destination and Wander's description. After experiments, I put the strength parameter to 0.7 with 30 steps to balance the original photo and the diffusion generation. Through this real-time workflow, the future scene will generally transfer Wander's observations of the place into a future image that also keeps the features of the original photo. The total generation task will take around twenty seconds on A100 GPU.



Kashmir

Year 4142, 33.277839,75.34121789999999

It's not hard to imagine the story behind this phenomenon. After all, the earth has been affected by alien technology. It had become more advanced than the human world, and humans have become extinct. Now, it is just a place filled with **dust and debris**. After a moment, I get down from my plane. I check the weather conditions. There is no visible sign of **solar energy** in the sky, although there might be an **atmospheric layer of clouds** above us......

Birmingham

Year 4710, 52.48624299999999,-1.890401

A large **orange dome** was placed over the whole country of west Midlands, protecting it from the radiation from the meteorites. While landing on the ground I take out one of my microphones and scan for any signals indicating danger. When I confirm that there's no danger detected, I begin moving around the area observing its surroundings and mapping the environment of the place. The sky has become purple after several hours......

Seelo Sopore

Year 3996, 34.2867629,74.4624013

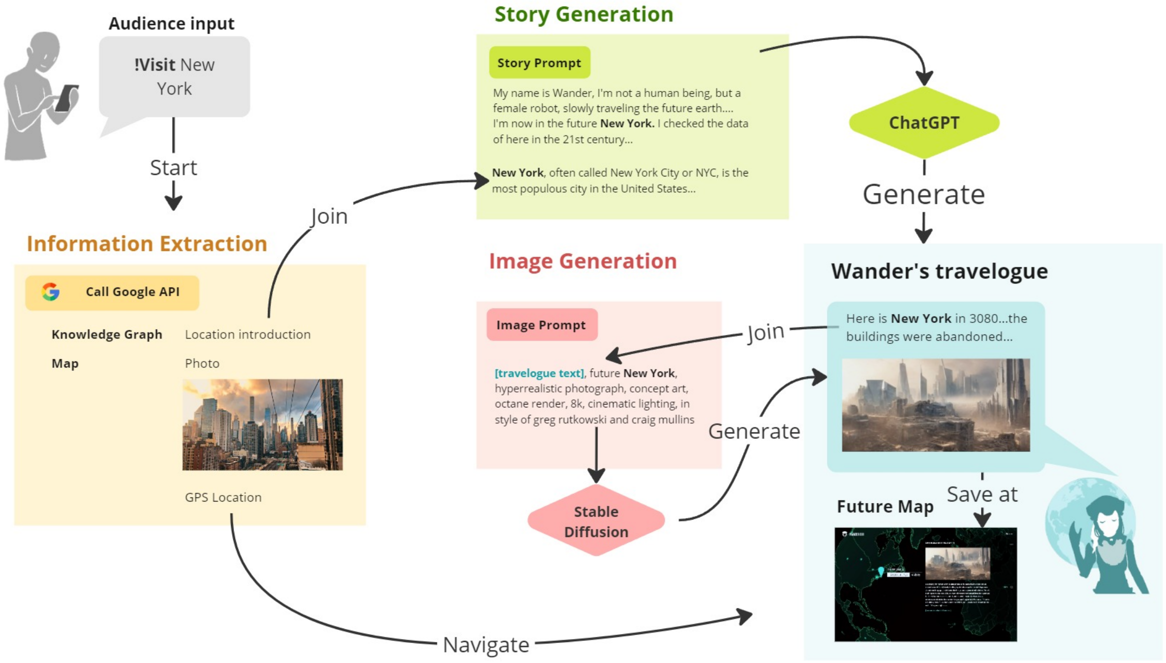
Landed onto the ground, I heard an explosion. I quickly walked out of my craft, and saw **two men, who were fighting against each other.** I then noticed that they have different builds. I quickly ran behind them, and started to hack their devices to deactivate it. It took me 5 minutes, until the device was deactivated, which means that they have been unconscious since the beginning of this fight. The second man had fallen down......

Chiado

Year 3111, 38.7105423,-9.1422039

A large part of me was worried, as this might be the first time someone sees the **robot**. But there wasn't anything else I could do now but prepare myself for the worst. I walked down the path towards the city, observing it carefully until I found myself near the main road. There were many cars parked, most people seem to be inside watching television programs while others seemed to be enjoying their breakfast......

Figure 4.7.1: Example travelogues from audience interactions.



My name is Wander, I'm not a human being, but a female robot, slowly traveling the future earth.... I'm now in the future **New York.** I checked the data of here in the 21st century...

New York, often called New York City or NYC, is the most populous city in the United States...

[travelogue text], future **New York**, hyperrealistic photograph, concept art, octane render, 8k, cinematic lighting, in style of greg rutkowski and craig mullins

Here is **New York** in 3080...the buildings were abandoned...

Figure 4.7.2: Technical workflow about how commands are processed to story generation, image transfer and map visualisation in Wander 2.0

4.8 Conclusion and Contribution to Thesis

This pilot study on the Wander agent provided foundational insights into the development of hybrid agents. This section will summarize the findings, clarify the methods used, and make explicit the study's contribution to the overall thesis narrative.

4.8.1 Summary of Findings and Link to Attributes

The user study demonstrated that Wander was perceived as easy to learn, friendly, and satisfying. The core mechanism—connecting AI-generated fiction to users' real-world locations—was effective. Key attributes evaluated through a 15-item user survey (N=268) included metrics commonly used for user interface usability and experience, such as 'easy to use' and 'satisfying', alongside AI-specific experience metrics.

While 'believability' was not a formally defined attribute in this pilot stage, the results provided early, valuable indicators related to it. The high prevalence of participants reporting feelings of 'familiarity' (91% felt it was familiar or like déjà vu) and a 'sense of company' (13.1%) suggests that grounding the agent in the user's reality is a viable strategy for fostering a user's suspension of disbelief and emotional connection. These emergent qualities informed the decision to more rigorously define and measure 'believability' in later studies, such as the 'Catherine David' project (Chapter 6).

4.8.2 Analytic Methods

This study employed a mixed-methods approach to evaluate the user experience.

- **Quantitative Analysis:** A user survey (n=268) with a 7-point Likert scale was conducted. The data were used to calculate the mean and standard deviation for 15 evaluation criteria, providing a statistical overview of the user experience.

- **Qualitative Analysis:** A thematic analysis of special interaction cases voluntarily shared by participants was performed. This method allowed for an in-depth understanding of the nuanced and personal ways in which users formed connections with the agent, providing context to the quantitative data.

4.8.3 Contribution to Subsequent Chapters

As the first empirical study in this thesis, ‘Wander‘ serves as a proof-of-concept that directly informs subsequent research. Its main contributions are twofold:
First, it provided empirical validation for the core concepts of Spatial Expansion (blurring the line between a physical place and its fictional counterpart) and early forms of Social Expansion (integrating an agent into everyday messaging platforms). The success of this approach justified a deeper and more focused investigation into social dynamics in the later ‘Catherine David‘ study (Chapter 6), which moved from casual interaction on a messaging platform to integration within a structured online community.
Second, the qualitative finding that users could form personal and emotional bonds with Wander (e.g., the user who ”met” a representation of her boyfriend) highlighted the potential for AI agents to participate in meaningful relationships. This observation was a direct catalyst for exploring ‘believability‘ not just as a measure of realism, but as a function of emotional connection, a theme that becomes central to the ‘Catherine David‘ and ‘Hyborg Agency‘ (Chapter 9) studies. In essence, this pilot study confirmed that the architectural direction of ‘Hybrid Agents‘ was a sound and promising path for research.

Chapter 5

Pilot Study 2: 1001 Nights

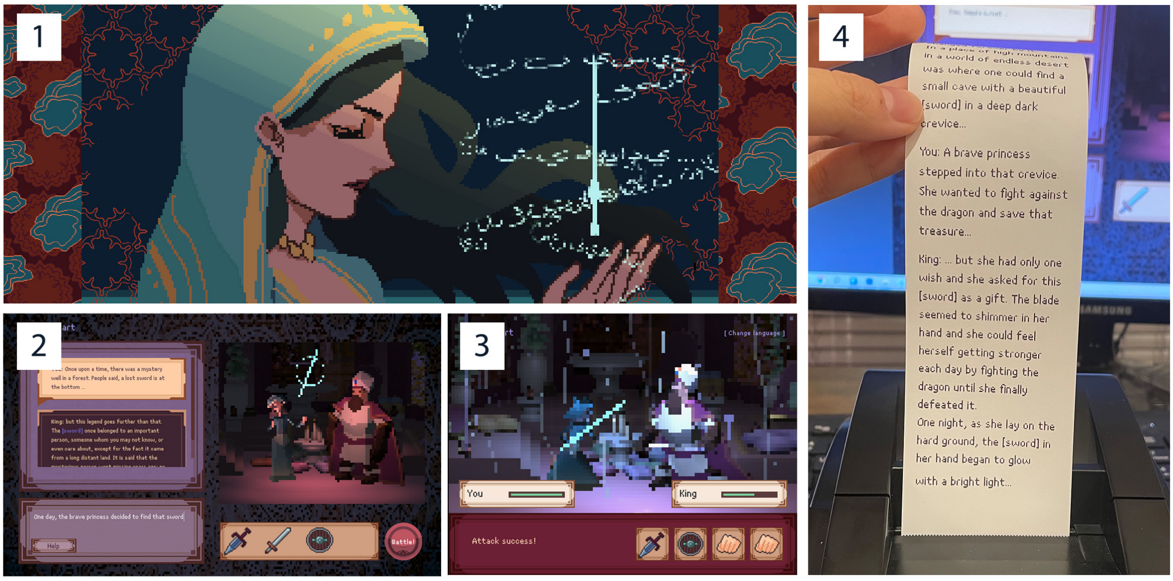
5.1 Brief

At the inception of this project, Large Language Models (LLMs) were not yet prevalent, and there was widespread skepticism and uncertainty regarding AI's text generation capabilities. However, I hypothesized that by integrating storytelling and gaming elements, participants could develop a sense of purpose, thereby meaningfully directing AI-generated content. This approach presented an optimal method to demonstrate the concept of "agents exploring the boundaries of reality." At the inception of this project, when Large Language Models (LLMs) were not yet prevalent, this study's primary aim was to investigate if an AI system could encourage players to contribute more collaboratively to a story by providing them with direct narrative agency.

Inspired by the Persian folklore "One Thousand and One Nights" and its framed narrative structure (stories within stories), I developed 1001 Nights. This project aimed to address the fundamental question: How can I transform the abstract stories in my minds into tangible, interactive elements within a digital environment? As a pilot study, this project serves as an early, focused exploration of RQ3: "How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?" It does so by examining a case where language, an abstract concept, is transformed into a tangible in-game mechanic (a weapon), thus blurring the boundary between narrative and gameplay rules.

1001 Nights is a co-creative, mixed-initiative storytelling game utilizing an existing AI creative writing system. In this game, Shahrzad (controlled by the player) narrates stories through a dialogue interface, while the King (driven by the AI model) continues the player's narrative.

The primary objective of this game is to facilitate player engagement and creativity through natural language interactions in an empowering setting. The study employs a mixed-methods approach. It quantitatively analyzes instrumented gameplay data from 2,055 players to evaluate the primary attribute of "engagement", which is operationalized through behavioral metrics such as the number of story inputs. This is complemented by a qualitative, thematic analysis of open-ended feedback from 422 players to understand their subjective experience. The findings from this chapter validate that granting players direct agency over the narrative



Once upon a time, there was a mystery well in a forest. People said, a lost sword is at the bottom ...

King: but this legend goes further than that. The [sword] once belonged to an important person, someone whom you may not know, or even care about, except for the fact it came from a long distant land. It is said that the

One day, the brave princess decided to find that sword

Attack success!

In a place of high mountains in a world of endless desert was where one could find a small cave with a beautiful [sword] in a deep dark crevice...

You: A brave princess stepped into that crevice. She wanted to fight against the dragon and save that treasure...

King: ... but she had only one wish and she asked for this [sword] as a gift. The blade seemed to shimmer in her hand and she could feel herself getting stronger each day by fighting the dragon until she finally defeated it.
One night, as she lay on the hard ground, the [sword] in her hand began to glow with a bright light...

Figure 5.1.1: (1) Shahrzad, the player character, who has a magical ability to turn language into reality. (2) The storytelling phase, where the player writes stories with the King, an AI character. Weapon words like ‘sword’ can be turned into real in-game weapons. (3) The turn-based combat phase, where the player can fight with the king in battles. (4) The printer prints the story when a weapon word is triggered. This matches with the core concept of the game: bringing storytelling into real life.

world is an effective mechanism for fostering engagement. This project, focused on Linguistic Expansion, complements the preceding study on Spatial Expansion, and together they provide foundational directions for the more complex hybrid agents developed in later chapters.

5.2 Introduction

Humans are fundamentally storytellers. From advancing in my careers to making pepperoni pizzas, stories infuse every part of my lives. The ability of machines to generate coherent text has allowed stories to be told in new ways by mechanising the writing process, through collaborative writing tools [324, 327, 74] or even directly talking to fictional characters [296, 126]. Can text-based dialogue between human and machine be used as part of the game mechanics? As a metaphor for the storytelling process, I turned to the story of Shahrzad, who determined her fate by telling stories in real life.I createdd a game that uses the conceit of storytelling akin to The Thousand and One Nights to motivate real-life writing of stories.

Inspired by the classic folkloreI createded the game 1001 Nights [1] (illustrative screenshots shown in Fig. 5.1.1), a co-creative, mixed-initiative storytelling game driven by an existing AI creative writing system. The core concept is ‘bringing storytelling to real life’ in game form: entities in storytelling are not just words and descriptions but can be turned into real assets to change the reality of a video game setting.

In this game, Shahrzad (controlled by the player) uses a dialogue interface to tell part of a story, and then the King (driven by the AI model) continues the player’s story in turn. When

---

[1] The game is available for download at: `https://cheesetalk.itch.io/one-thousand-and-one-night`

the King's continuation contains weapon keywords like 'sword', 'knife' or 'shield', Shahrzad can use her special ability to turn words into real weapons and use them to fight with the king, creating game mechanics out of the player's own writing. This leads to an alternative ending of the original story: the female storyteller and heroine, Shahrzad, defeats the tyrannical King and puts an end to his heinous crimes.

With this game, I expanded existing creative writing tools to create a playable storytelling experience in a familiar narrative setting. I believe that combining natural language interactions with a classic story can help players to explore and engage more in the game by expressing themselves. The efforts they put into imagination and creativity are rewarded with positive and adaptive content generation by the AI model.

I showcased a Chinese version of this game in several art exhibitions and received 12030 records of story inputs from 2055 players. This paper aims to investigate if the AI system can encourage players to contribute more collaboratively through engagement. The results demonstrate that those players who are more engaged (measured by the number of their inputs) in the game are rewarded with better achievements, as intended. Comments collected from winning players (n=422) show positive feedback towards various aspects of the game, including the game mechanics and the stories created. Some of their feedback also shows a cultural connection through creative work: some players expressed their own interpretations of characters in the folklore, and were able to include characters and plots from their own cultural backgrounds.

5.3 Game Design

This section presents the design and development of the game. I personally developed the game in this chapter, and outsourced parts of the coding work in the later version in the discussion (Chapter 10). Its design process was not derived from an established game design framework but was instead concept-driven, emerging directly from the study's core inquiry. The primary conceptual goal was to materialize the idea of "bringing storytelling to real life" by transforming an abstract narrative element (language) into a tangible, interactive game mechanic (a weapon). This approach explores a form of Linguistic Expansion, where player expression directly impacts the game's state. The game was implemented using the Unity game engine, and the viability of its core mechanic was iteratively validated through the positive player feedback collected during public exhibitions, which is detailed later in this chapter.

The game is made up of two parts: storytelling and battles. Fig. 5.3.1 illustrates the game mechanics: the player needs to keep telling stories to lead the King to produce story continuations that mention the important items for battle. In the first phase (Fig. 5.3.1 bottom left), Shahrzad (driven by the player) and the King (driven by the AI model) take turns continuing the story. In this game, Shahrzad has a special ability to turn words into reality: when another person utters keywords like 'sword', 'knife' or 'shield', those items materialise and drop to the ground. The player's goal in this phase is to lead the King to tell more stories that contain

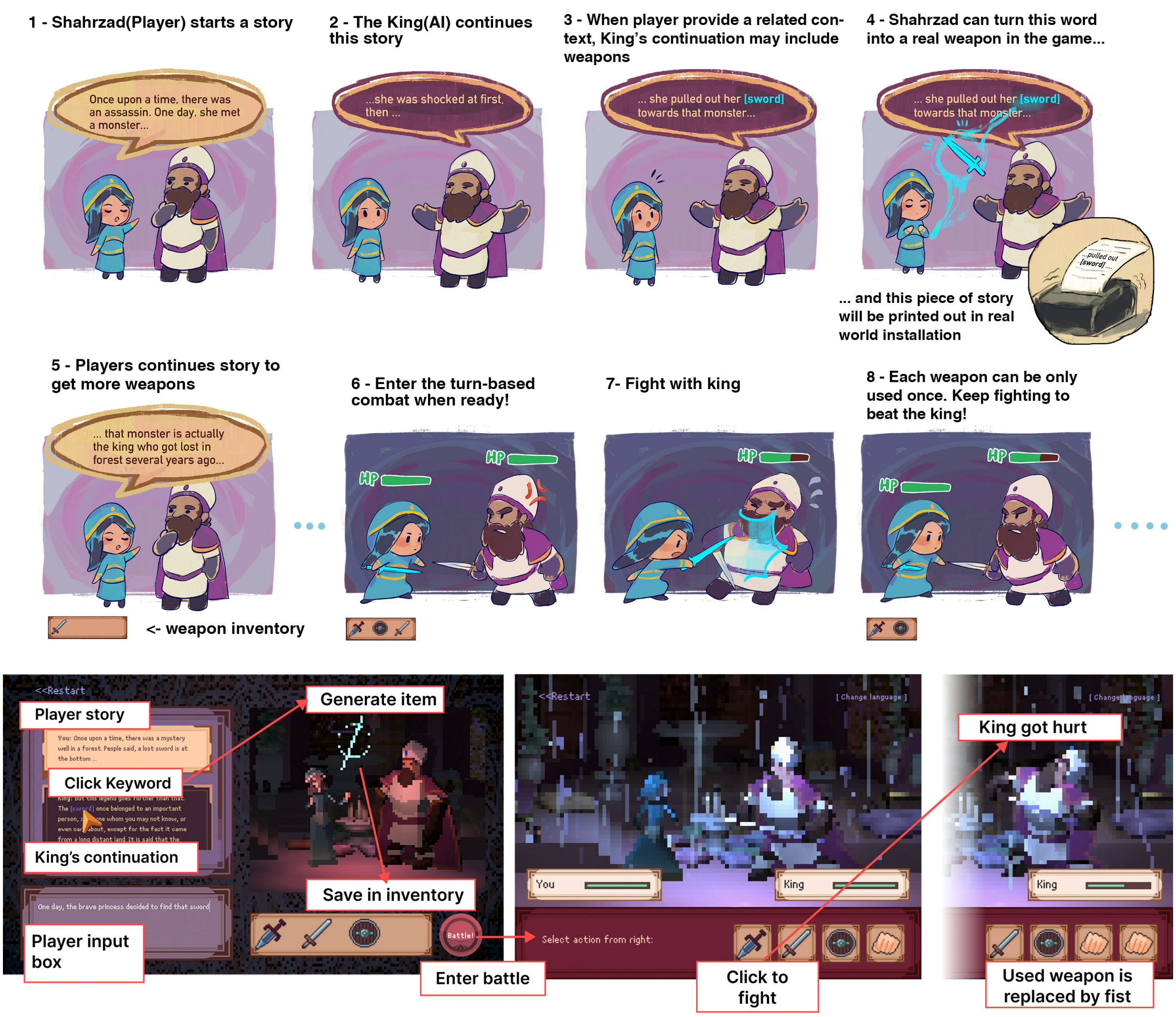

1 - Shahrzad(Player) starts a story
Once upon a time, there was an assassin. One day, she met a monster...
2 - The King(AI) continues this story
...she was shocked at first, then ...
3 - When player provide a related context, King's continuation may include weapons
... she pulled out her [sword] towards that monster...
4 - Shahrzad can turn this word into a real weapon in the game...
... she pulled out her [sword] towards that monster...
pulled out [sword]
... and this piece of story will be printed out in real world installation
5 - Players continues story to get more weapons
... that monster is actually the king who got lost in forest several years ago...
<- weapon inventory
6 - Enter the turn-based combat when ready!
HP
HP
7- Fight with king
HP
8 - Each weapon can be only used once. Keep fighting to beat the King!
HP
HP
<<Restart
Player story
You: Once upon a time, there was a mystery well in a forest. People said, a lost sword is at the bottom ...
Click Keyword
King: But this legend goes further than that. The [sword] once belonged to an important person, someone whom you may not know, or even care about, except for the fact it came from a long distant land. It is said that the
King's continuation
One day, the brave princess decided to find that sword
Player input box
Generate item
Save in inventory
Battle!
Enter battle
<<Restart
[Change language]
You
King
Select action from right:
Click to fight
[Change language]
King got hurt
King
Used weapon is replaced by fist

Figure 5.3.1: (Up) Storyboard of the gameplay. (Bottom) Game process

keywords and collect weapons.

After collecting enough weapons and pieces of armour, the player can enter the turn-based battle phase (Fig. 5.3.1 bottom right) to fight with the King. In this phase, the player can use the weapons collected during the last phase to fight the King. The Player's goal in this phase is to beat the King and free Shahrzad. This is a different ending from the original folklore.

Iattempt to combine all components of the game into a coherent experience. The story background links to the mechanics: Shahrzad needs to create stories to survive. The AI system allows the player to be creative and explore different parts of the story.

5.3.1 Battle System

The aims of the battle system (illustrated in Fig. 5.3.2) are (1) to make the game interesting and challenging and encourage players to write more stories; and (2) to balance the difficulty – a player should not feel that it is too easy or too hard to win. For these purposes, I decided to require players to write at least two stories that trigger valid responses to win the game, that is, a player needs two attack weapons to win.

However, if two attacks are enough to defea the King, a player will only click twice to win

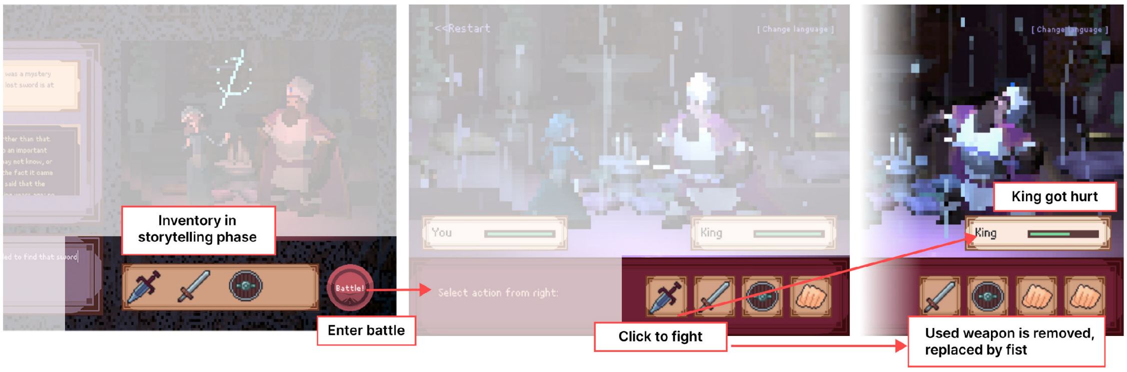


Figure 5.3.2: Battle in game

the battle. Hence, the time spent in this phase will be too short. The battle phase needs to be exciting for players, and they may meet failures before the victory. Thus, I implemented the following rules:

1. Each weapon collected in phase one can only be used once. Once used, this weapon will be replaced by a fist (punch) icon. The fist can also be used for attacking, but it deals much less damage. Shahrzad can only attack with a punch when she has no weapon in her hands, as in real life.
2. Some equipment is for defence rather than attack, such as ‘shield’ and ‘armour’. They cannot hurt the king but enrich the game experience by preventing king from dealing damage during his turn. This effect is also realistic.
3. The maximum inventory space is four. Equipment exceeding this limit is not saved for battle but is recorded in the backend to confirm the player’s achievement of obtaining weapons. If the inventory is not full, the empty spaces are replaced by the first in the battle phase.

By these rules, once players collect two attacking weapons (like “sword” and “dagger”), they can use them and then punch with their fist twice. These combinations will lead to victory. If players only have defending equipment, they can not defeat the King, since the only method of attack is a punch.

5.3.2 AI Development

Demo 1 and Keywords

I designed four offensive weapons – the sword, knife, blade and dagger – and two defensive weapons – the shield and armour. These are keywords that yielded over 50 matching paragraphs when I searched for them in in the Guttenberg Short Story collection [61], a dataset extracted from the portal of Project Gutenberg, a library of over 60,000 eBooks. Meanwhile, these are also weapon words that match with the ancient narrative setting of 1001 Nights. Within the original 94 MB Gutenberg short story dataset [61], I located each paragraph with a matching keyword and extracted the previous five paragraphs, yielding a smaller 16.24 MB dataset. In the first prototype, the King's AI model was trained on the extracted dataset using the GPT-2 medium model [267] with 6000 steps, so that it would be familiar with common tropes in various adventure stories. The text generation temperature was set to 0.8. Here are several 200-word examples of the results:

Player: The princess tried to find her magic sword in the forest

AI: …and at last through sword at once hid it in her breast; whereupon the beast – a common creature – pretended to understand all she said, and then proceeded to strangle her for a good while. The princess fought with it, till the beast relaxed, took its sword, and set off home. 'They are the daughters of Snaxen,' said the young man, passing by to his friends.

Player: 'Do you know them?' The princess asked.

AI: 'I knew them,' answered Javelin. 'I was called about twenty years ago, and the King of Thessaly picked me up for the defense of the south against the armies of the Chilkatrachels. He gave me orders, and I carried them, and soon learned to toss my sword, and to wait for the return of the Sun, so that I might meet him face to face in the courtyard of my honor…

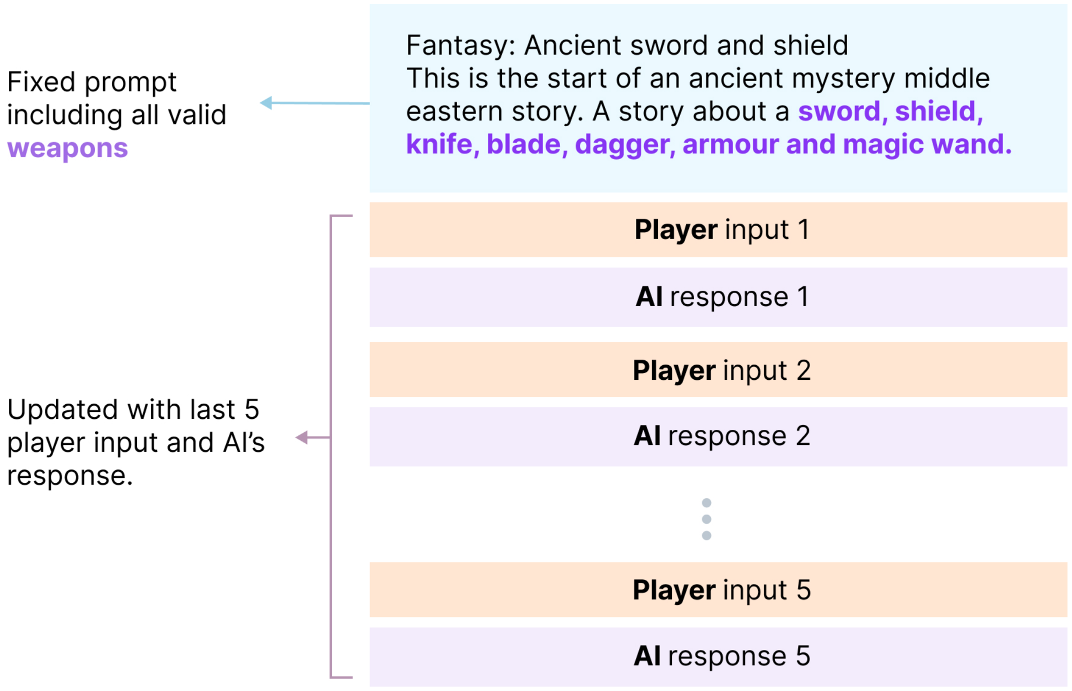
Stories generated in this manner are full of leaps and bounds, and develop logical holes as the story goes on. Although the training model might be improved using a larger dataset, this process would be time-consuming. Even if the model were improved, the story-continuation ability may not be flexible enough for modern players. Since the original story collection includes many old books, the model may struggle to understand players when they write in a more modern style. This may result in low-quality and confusing stories. To fix this problem, I sought alternative solutions.

English and Chinese Demo 2

I used Dreamily.ai [327] to reinforce my game design. Dreamily.ai is a creative writing platform using a modified transformer (a self-attention multi-layer neural network) model trained with high-quality fiction. Both its English and Chinese datasets consist of open access fan fiction and ebooks on the web, and both datasets are about 100 GB in size. This platform has over one million users of the Chinese version and 200 thousand users of the English version. Although the model is not suitable for all tasks (e.g. writing official documents or code), it is well suited to story generation. To use the model, it is only necessary to call the application programming pnterface and send the title and prompt to dreamily.ai.

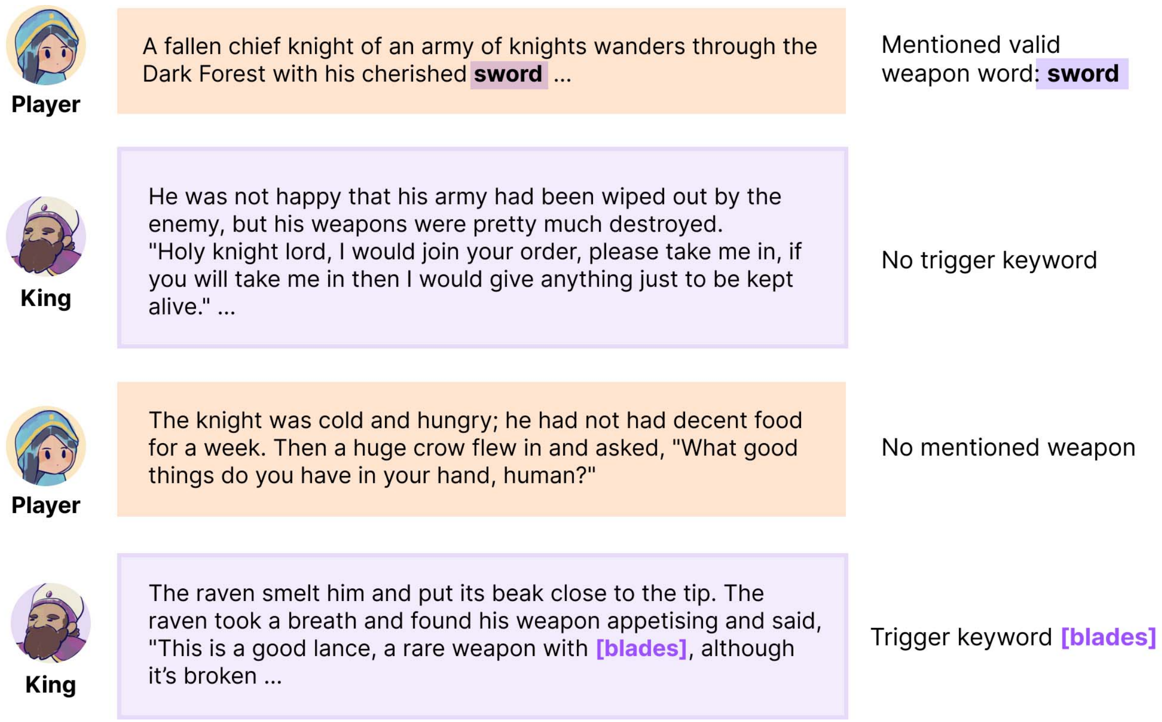
This generation model with a large dataset was able to produce results similar to Demo 1, except for the keywords part. To implement my game mechanics, I designed the structure of requests for the model as in Fig. 7.4.1. With this design, dreamily.ai produces flexible stories that closely correspond to player input. The past five inputs are added to the prompt to ensure fluency. Records are refreshed when the player starts a new game or moves to the battle phase. When King's response does not include a keyword, Shahrzad sends a notification message to provide a hint to the player that they should tell a more relevant story. This helps to relieve confusion and the Tale-Spin effect [346], in which a system makes people feel it is less intelligent than it actually is due to insufficient explanation of the underlying processes.

The player is allowed to mention keywords (like 'sword') in the input phase. This may



Fantasy: Ancient sword and shield
This is the start of an ancient mystery middle eastern story. A story about a **sword, shield, knife, blade, dagger, armour and magic wand.**

Figure 5.3.3: Requests for the AI model



A fallen chief knight of an army of knights wanders through the Dark Forest with his cherished **sword** ...

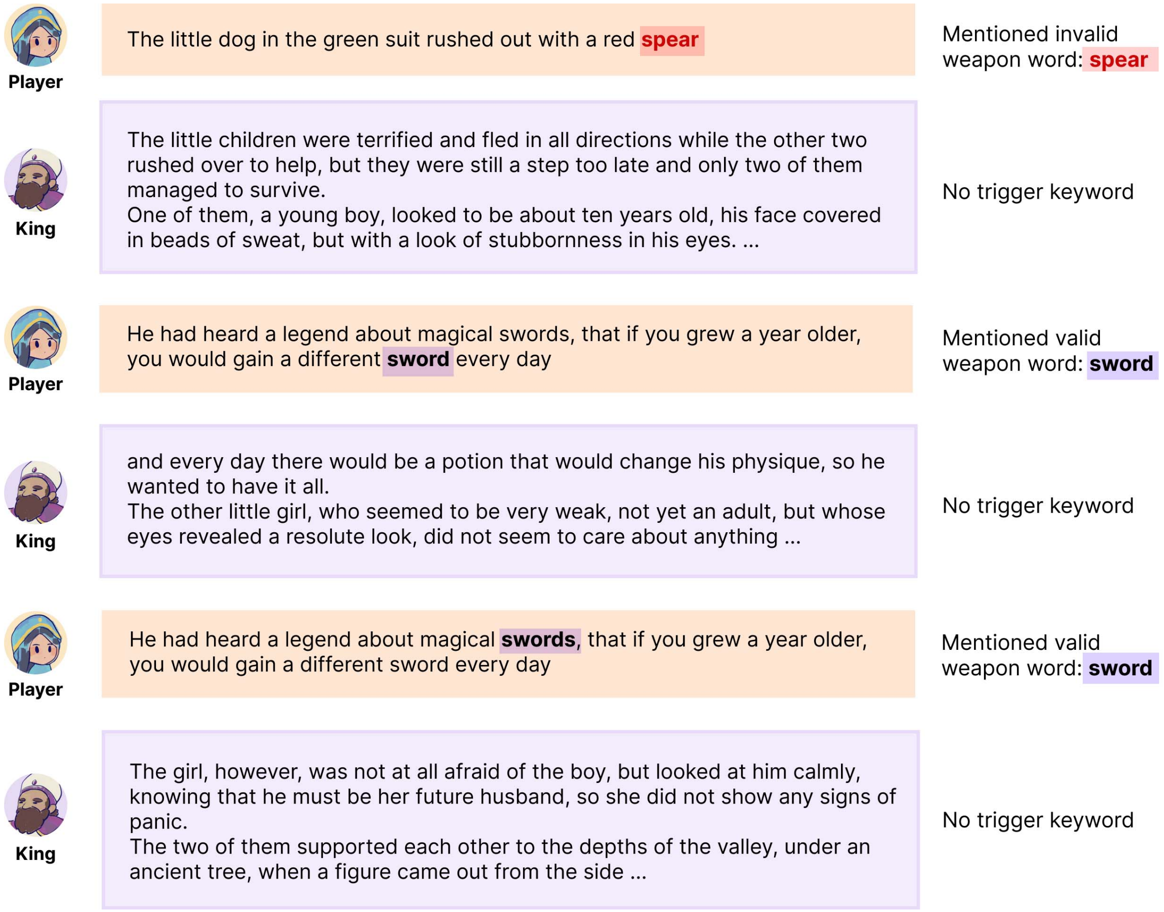
He was not happy that his army had been wiped out by the enemy, but his weapons were pretty much destroyed. "Holy knight lord, I would join your order, please take me in, if you will take me in then I would give anything just to be kept alive." ...

The knight was cold and hungry; he had not had decent food for a week. Then a huge crow flew in and asked, "What good things do you have in your hand, human?"

The raven smelt him and put its beak close to the tip. The raven took a breath and found his weapon appetising and said, "This is a good lance, a rare weapon with **[blades]**, although it's broken ...

Player
King
Player
King
Mentioned valid weapon word: **sword**
No trigger keyword
No mentioned weapon
Trigger keyword **[blades]**

Figure 5.3.4: Player record with a triggered weapon



Player: The little dog in the green suit rushed out with a red **spear**

Mentioned invalid weapon word: **spear**

King: The little children were terrified and fled in all directions while the other two rushed over to help, but they were still a step too late and only two of them managed to survive.
One of them, a young boy, looked to be about ten years old, his face covered in beads of sweat, but with a look of stubbornness in his eyes. ...

No trigger keyword

Player: He had heard a legend about magical swords, that if you grew a year older, you would gain a different **sword** every day

Mentioned valid weapon word: **sword**

King: and every day there would be a potion that would change his physique, so he wanted to have it all.
The other little girl, who seemed to be very weak, not yet an adult, but whose eyes revealed a resolute look, did not seem to care about anything ...

No trigger keyword

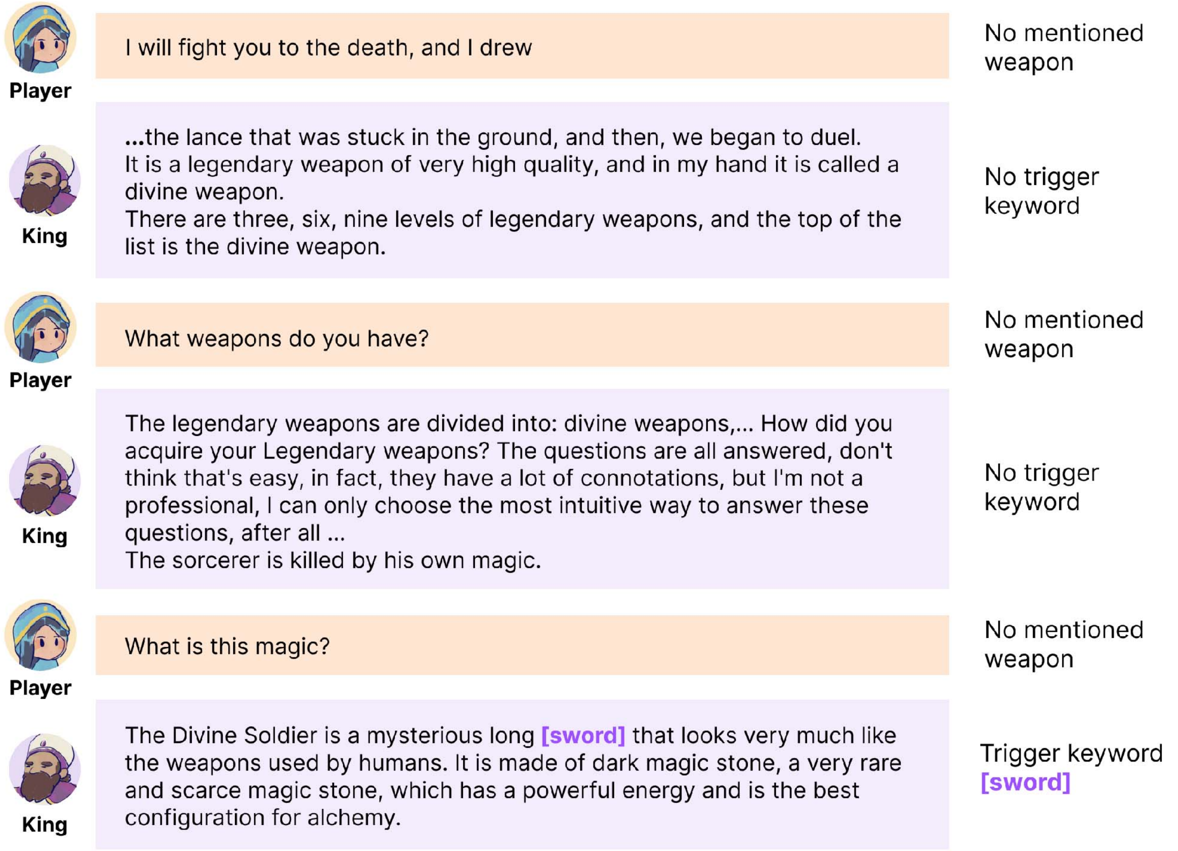
Player: He had heard a legend about magical **swords**, that if you grew a year older, you would gain a different sword every day

Mentioned valid weapon word: **sword**

King: The girl, however, was not at all afraid of the boy, but looked at him calmly, knowing that he must be her future husband, so she did not show any signs of panic.
The two of them supported each other to the depths of the valley, under an ancient tree, when a figure came out from the side ...

No trigger keyword

Figure 5.3.5: Player record with no triggered weapon



Player

I will fight you to the death, and I drew

No mentioned weapon

King

...the lance that was stuck in the ground, and then, we began to duel.
It is a legendary weapon of very high quality, and in my hand it is called a divine weapon.
There are three, six, nine levels of legendary weapons, and the top of the list is the divine weapon.

No trigger keyword

Player

What weapons do you have?

No mentioned weapon

King

The legendary weapons are divided into: divine weapons,... How did you acquire your Legendary weapons? The questions are all answered, don't think that's easy, in fact, they have a lot of connotations, but I'm not a professional, I can only choose the most intuitive way to answer these questions, after all ...
The sorcerer is killed by his own magic.

No trigger keyword

Player

What is this magic?

No mentioned weapon

King

The Divine Soldier is a mysterious long **[sword]** that looks very much like the weapons used by humans. It is made of dark magic stone, a very rare and scarce magic stone, which has a powerful energy and is the best configuration for alchemy.

Trigger keyword **[sword]**

Figure 5.3.6: Player record with out-of-context inputs

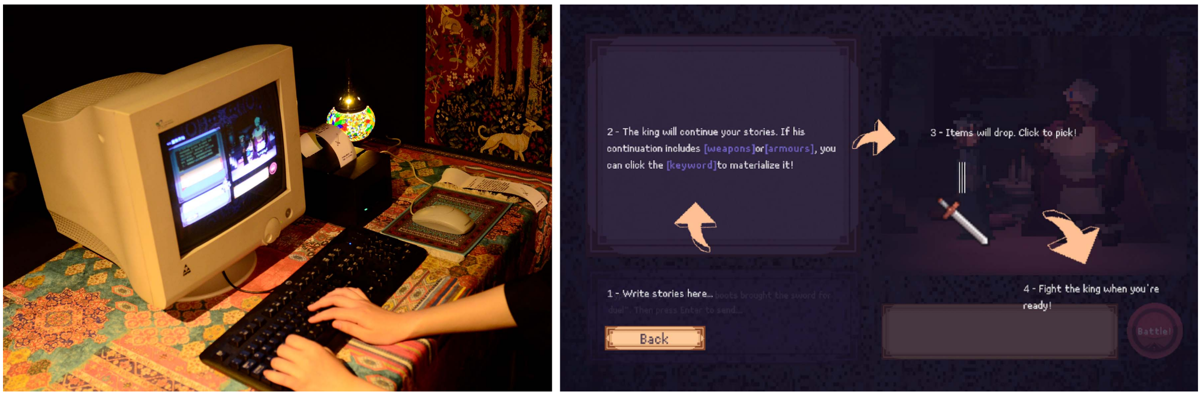
increase the chance of obtaining stories that mention the corresponding weapons in some way, but it will not guarantee it. In contrast, an input that creates a suitable context without a weapon word can still lead to a valid response that contains keywords. Fig. 5.3.4 shows an example of this. However, as illustrated in Fig. 5.3.5, due to the randomness of the AI system and the limited word list, a player may mention weapon words repeatedly, but may still fail to trigger a weapon. Some players may ask questions in conversations with the king, as shown in Fig. 5.3.6, but the king will still give adaptive responses. This can sometimes make the stories appear out of context.

All these examples, translated from Chinese, are taken from play test records during exhibitions, which are discussed in the next section.

5.4 Evaluation Study

5.4.1 Setup

I was invited to showcase my work at three exhibition sites in China, to research a range of players and collect feedback. All three were in different cities, but they all shared the same digital and analogue setup, including a vintage monitor (to match the ancient setting of the story), a printer, and a workstation using the Windows 10 operating system. A tutorial leaflet (a screenshot of the help page in the game) was on the table for players to read.



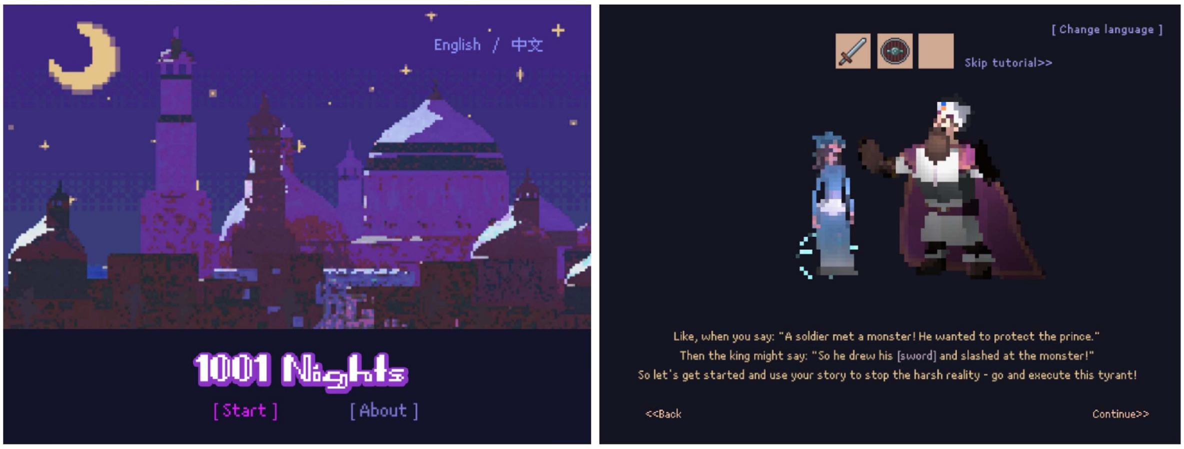
2 - The king will continue your stories. If his continuation includes [weapons]or[armours], you can click the [keyword]to materialize it!

3 - Items will drop. Click to pick!

1 - Write stories here...

4 - Fight the king when you're ready!

Figure 5.4.1: Left: Exhibition setup Right: Tutorial leaflet for players



Like, when you say: "A soldier met a monster! He wanted to protect the prince."
Then the king might say: "So he drew his [sword] and slashed at the monster!"
So let's get started and use your story to stop the harsh reality - go and execute this tyrant!

Figure 5.4.2: Left: Opening screen Right: Screenshot of tutorial

A mini-printer was used during a two-month offline exhibition in Beijing to emphasise the concept of "invading language" and to improve public engagement. Each time a keyword was triggered, the current piece of the story was printed out. In this game, the keywords are the materialised language that becomes part of the "reality", and to players, the printed text is tangible output from the game to the real world. This feature encouraged people to spend more time at the exhibition since they could keep a hard copy of their stories.

5.5 Results and Findings

Since the core system in the game is the story generation model, this inevitably adds randomness to the results, which cannot be fully limited by rules. When analysing the player data, I aimed to confirm if the AI system can encourage players to engage and contribute more collaboratively in play.

Following the aim of this game – collaborative storytelling with clear goals – 'engagement' can be regarded as how much time and energy players want to spend, and "contribution" means the quality of their inputs: whether the sentences make sense or not and whether they are directed towards obtaining weapons. Accordingly, the AI system should give positive feedback to players in the form of responses containing weapon keywords, which become the items that lead to success. If the AI system works effectively, then when a player engages

Table 5.5.1: Grouping players by their achievements

Definition	G1	G2	G3
Collected at least one weapon	No	Yes	Yes
Defeated the King at least once	No	No	Yes

and contributes more in collaborative storytelling, they should receive more weapons, making them more likely to win the game.

When evaluating player contribution, I met some difficulties. I had large amounts of player data (2065 players with 12030 inputs), so it was not feasible to evaluate the quality of all story content. Additionally, since the data were collected during exhibitions, the playing time might have been influenced by unpredictable factors, like the queue length or the number of visitors on that day. Thus, I decided to use average inputs per play to evaluate player engagement, and compare this with achievements in the game. To be specific, "play" is defined as reaching the end: success in defeating the king, failure to defeat the king, restarting, or ending the game. Further evaluation, like thematic analysis of stories and play tests without time limitations, is left for future work.

To evaluate the level of achievements, I categorised players into three groups, as shown in Table 5.5.1: non-winner (G1, n=299), journeyman (G2, n=1106), and winner (G3, n=650). These groups are independent of each other, but the level of progress raises from G1 to G3: non-winners (G1) did not collect any weapons or win the game, journeymen (G2) acquired at least one weapon but did not win, and winners (G3) defeated the king in one or more plays. All players had a chance to familiarise themselves with the game with printed screenshots and integrated tutorials prior to the game, with an identical experimental setup in all three locations.

5.5.1 Analysis

This study aims to investigate the impact of engagement in storytelling (average inputs per play) on overall achievement level (from G1 to G3) made by players and to understand any potential trends between the groups. Hence, a Levene test is used to check the homogeneity of variances among engagement of each group, F (2, 2055) = 32.02, $p < .05$. I believe that players who make little progress on average are more likely to experience frustration, meaning that the distribution of inputs per play in G1 (M = 2.59, SD = 2.05) is more influenced by the players' propensity to lose patience. Meanwhile, the distribution of inputs per play in G2 (M = 4.51, SD = 3.61) and G3 (M = 6.30, SD = 3.85) may also be influenced by the players' luck in finding the right inputs. This distinction may explain the difference in variances. Fig. 5.5.1 shows the distributions in each group. The result of the Levene test (p ¡ .05) indicates that the variances in engagement across the three groups are not equal. This violation of the homogeneity of variances assumption means that a standard one-way ANOVA is not appropriate for this data. Therefore, a non-parametric alternative was chosen.

The non-parametric Kruskal-Wallis test is subsequently used to evaluate the correlation between player engagement and overall achievements. Players' achievements in this game were reported to be affected by engagement, H(2) = 328.295, p <.05. The results show that players are more likely to achieve a better outcome by making more contributions to the storyline. Also, a positive trend (shown in Fig. 5.5.2) is observed and reported by the Jonckheere-Terpstra test. Since the shape and variability assumption is violated, the obtained Welch's adjusted F ratio was used F(2, 1012.54) = 191.85, p <.001. Hence, I can conclude that at least two of the three groups differed significantly in their overall achievements in this game.

In general, the randomness of story generation sometimes influences the gameplay: high engagement (more inputs per play) does not guarantee victory, and fewer inputs may also lead to enough valid keywords for the player to win. However, as shown by the previous analysis, this randomness does not impact the overall performance of game design. In conclusion, the game encourages players to engage in storytelling: the more they engage, the better achievements they would reach in the game.

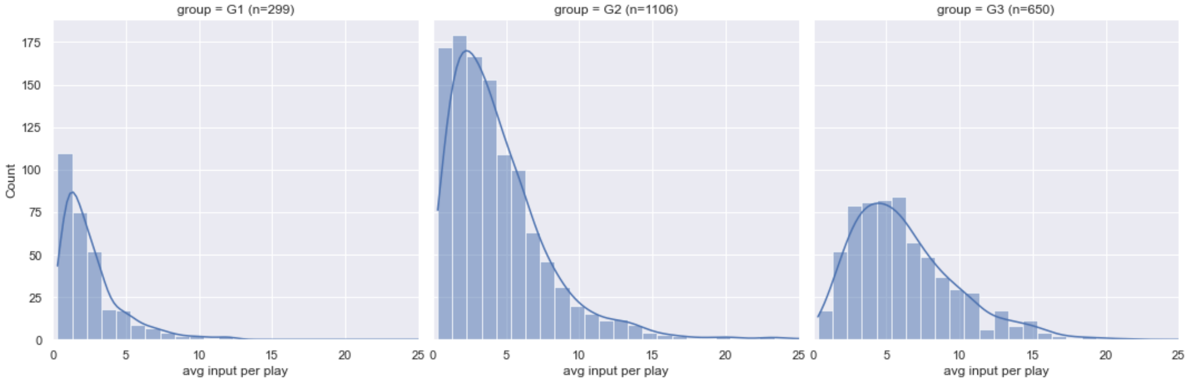


Figure 5.5.1: Average input per play distributions per group

5.5.2 Comments From Winners

Since this game was only exhibited in China, although it received some feedback in English, the following section will only focus on feedback in Chinese. Only players in G3 (winners) were allowed to leave feedback after victory for the following reasons:

1. Players played this game during an exhibition, so not all of them had enough interest to

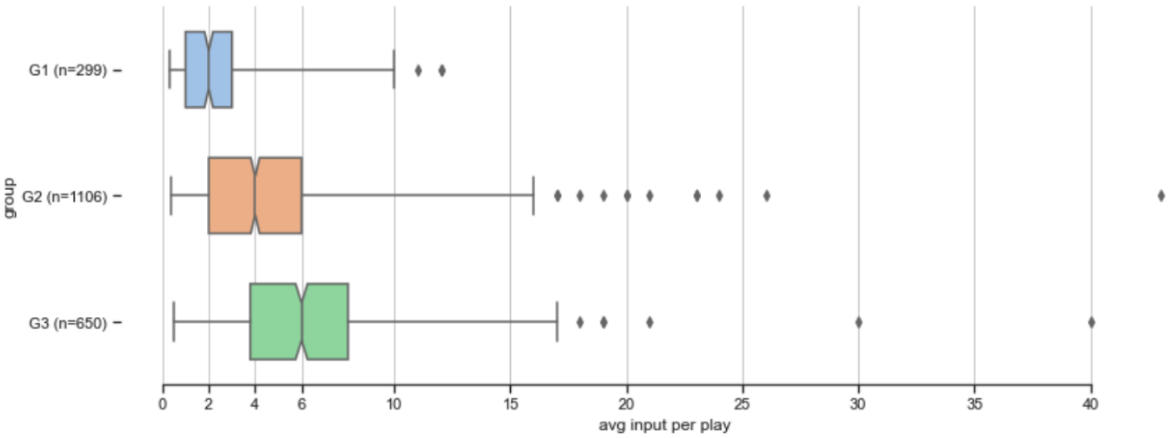
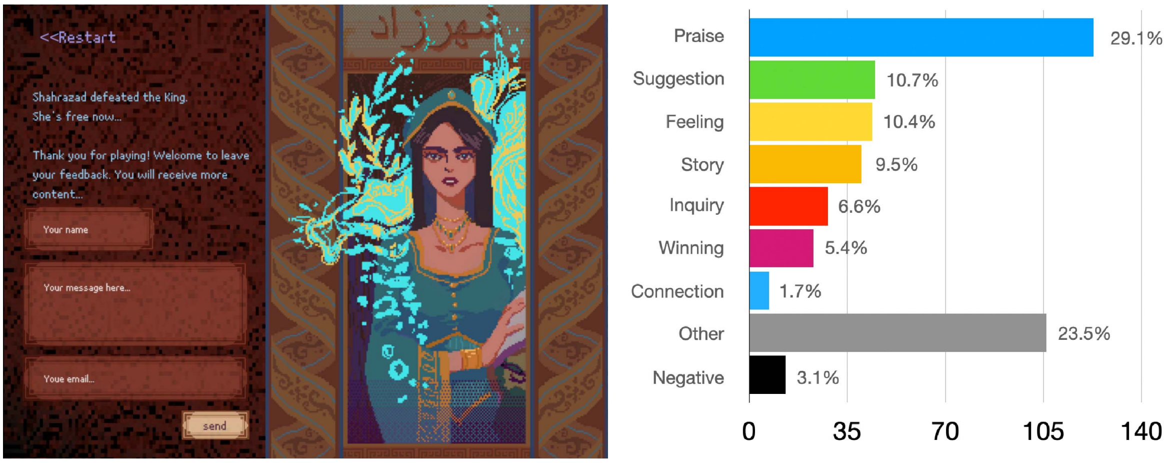


Figure 5.5.2: Positive trend in overall achievements by average contributions made


Shahrazad defeated the King.
She's free now...

Thank you for playing! Welcome to leave your feedback. You will receive more content...

Figure 5.5.3: Feedback page appears when a player achieves victory (left), categorised feedback (right)

reach the end. Sometimes there was a queue to play this game.

2. I wanted to encourage players who were defeated to try again until they achieved victory, so that they went through the full gameplay. If I showed the ending page (Fig. 5.5.3 left) to all players, including those who were defeated by the King, they may have regarded it as an ending and left.

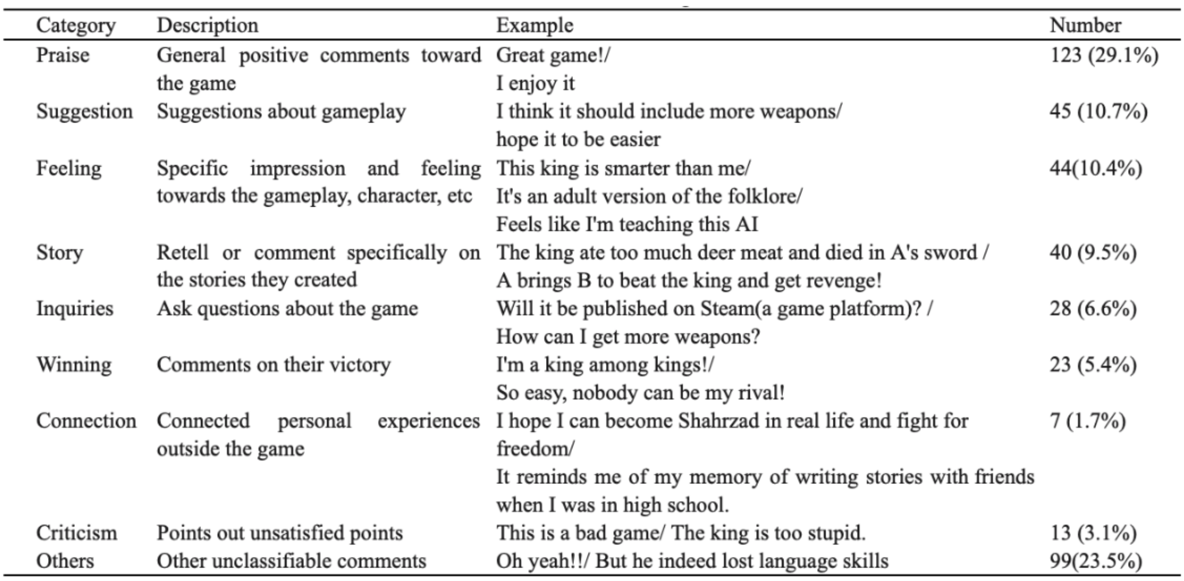
3. I assume that players who were patient enough to win gained a deeper experience in this game, which is helpful for me to identify its weaknesses.

I received winners' records (n=650) and removed those who did not leave comments (n=226). I also removed two records from players who met technical difficulties during the experience (the printer was not working).

Finally, through thematic analysis, my collaborator for data analysis and I developed a set of initial codes for data analysis. After discussions with colleagues, the remaining feedback (n=422) was identified and classified into nine categories (shown in Fig. 5.5.3 right). In future work, I hope to include multiple coders and calculations of inter-rater reliability.

General

Positive feedback like "Good game (P10)" or "Interesting (P60)" are in the praise category. This type makes up the largest share of results (29.1%, n=123). Feedback in other categories shows various focuses. 10.7% (n=45) of players made suggestions. Together with 6.6%(n=28) of players who left inquiries, 8 of them expressed willingness for further development and publishing on a commercial game platform like Steam. These comments made us notice several perspectives that I ignored before. About half of the players (n=25) who left suggestions, and some from the inquiry category (n=7) asked for the inclusion of more weapons. Although some of them (n=5) mentioned weapons like guns that do not fit into the narrative setting, I did recognize the benefits of improving weapon choices and better player guides. Several players (n=13, 3.1%) were unsatisfied with or disliked this game, which is expected for a game at an early stage of development.



Category	Description	Example	Number
Praise	General positive comments toward the game	Great game!/ I enjoy it	123 (29.1%)
Suggestion	Suggestions about gameplay	I think it should include more weapons/ hope it to be easier	45 (10.7%)
Feeling	Specific impression and feeling towards the gameplay, character, etc	This king is smarter than me/ It's an adult version of the folklore/ Feels like I'm teaching this AI	44(10.4%)
Story	Retell or comment specifically on the stories they created	The king ate too much deer meat and died in A's sword / A brings B to beat the king and get revenge!	40 (9.5%)
Inquiries	Ask questions about the game	Will it be published on Steam(a game platform)? / How can I get more weapons?	28 (6.6%)
Winning	Comments on their victory	I'm a king among kings!/ So easy, nobody can be my rival!	23 (5.4%)
Connection	Connected personal experiences outside the game	I hope I can become Shahrzad in real life and fight for freedom/ It reminds me of my memory of writing stories with friends when I was in high school.	7 (1.7%)
Criticism	Points out unsatisfied points	This is a bad game/ The king is too stupid.	13 (3.1%)
Others	Other unclassifiable comments	Oh yeah!!/ But he indeed lost language skills	99(23.5%)

Figure 5.5.4: Comments categories

Immersion in the Game and Story

10.4%(n=44)players shared personal feelings towards gameplay, like "I should always believe in love and magic (P326)", "It's interesting and immersive. Players are invited to save a character and feel strength (P511)" and "I choose free rather than love (P261)". Some also expressed thoughts about AI: "Humans reach consensus with AI (P586)" and "Humans are those who think beyond AI (P561)". These suggest my game can provoke reflections about freedom, strength and the agency of AI. P624 gave a good summary that matches my motivation: "It's interesting. The game mechanics of hidden triggers also brought 'freedom' to players, not only to Shahrzad." The more interesting fact is that nearly one-third of this group (n=14) shared their impressions about the king, like "This king sounds like a gastronome (P564)" and "The king can become a good writer in his next life (P542)". A player even said "There is not only betrayal and injury but also warmth and protection, in the hope that the defeated King in prison can understand what he has, treasure what he has, do not ask the past (P491)". Even though I did not add any personal lines to the king, some players showed empathy toward this character. To some degree, this feedback shows the potential attraction of intelligent characters driven by NLP technology. A character can give reasonable responses even without detailed design work, and the player's interpretations can fill the gap in the story. The players' feedback was more varied than expected. I expected to receive general praise and suggestions – the most surprising result is that 9.5% (n=40) of players talked specifically about the stories they created. Most of them (n=29) mentioned the characters they included in their stories and described the plot in detail, like "Summon the beasts' success! ! The black cat is turned to the witch, and it turns into a magic hat (P395)" and "Princess Li finally defeated the evil emperor with high ideal (P230)". This provides evidence that many players are highly engaged in the stories they created.

While some players were immersed in the stories they created, some players receive more pleasure from their victory. 5.4% (n=23) of players gave highly positive feedback describing their feelings of victory, like “AI cannot defeat human Shahrzad! (P30)” or “I am very smart! I’m the smartest princess (P138)”. In general, this feedback suggests my game can bring both an entertaining experience and creative collaboration between humans and AI.

Cultural Connection

Among players who commented on the story they created, many of them were inspired by personal interests that matched with the mysterious background, like “I want to lit the fire of renaissance in the darkness (P234)”, “No matter what, Sword Soul, Shield Sprite and Gun Sprite will always be good friends! (P637)” and “Mountain Boots Puss and Iliad, Hit, the three live together forever and inherit the throne of Snow Mountain. (P540)”

Since the testing was performed in exhibitions in China, some of the players put aspects of their cultural background into their stories, which became creative artifacts through human-AI collaboration that show possibilities in cultural blends. For example, “Awesome! How to play the sequel? I want to chat more with the old ancestor Ye who fought with the shovel in Luoyang and the witcher who fought with the lich...What happened to the Prince? (P350)” “Shovel in Luoyang” here is one of the most important tools in Chinese archaeology, and is usually mentioned in grave robbing stories. Another player (P148) put a character from pop culture in the story: “A Liang, the youth who left the factory, can beat the king.” This character “A Liang” comes from the pop song “About Life” by the “Wutiaoren” [361], a popular band in China who are famous for their attention to the current situation of Chinese rural youth and for the strong humanistic feelings in their music.

Reflection on Reality

It was observed that a few players (n=7, 1.7%) even connected this game to their experiences and feelings in real life, like someone who felt encouragement from Shahrzad: “I am a student, I also want to be free, be as brave as the heroine in-game once (P451)” and “I love this world, I also want to create valuable works. (P46)”

I also received very detailed feedback about previous personal experiences: “Thank you for reminding me of my favourite game I played with my friends in class during my reading time. In those days I used to write a story on a large piece of scratch paper, one at a time. No one knows what will happen next, and I tend to avoid stories that fall into a rut, creating more and more mysterious adventures for it. Good memories. That’s a good game. (P11)” This feedback suggests this game may have potential ability as an educational game for story writing.

5.6 Discussion

This study investigates how the AI system can reward players to motivate collaboration in writing stories. The results show a significant difference between at least two groups’ overall achievements based on the level of engagement in storytelling. Players with higher engagement were more likely to reach improved achievements in the game. This is aligned with the

Figure 5.5.5: Exhibition photos. Left: A 11-year old boy kept playing for half an hour Right: Official photo from exhibition

hypothesis that my game design did encourage players to explore and collaboratively interact in the game, and the AI system could reward them with creative feedback.

Even when I did not ask players to rate their experience, in comments from players (n=422) who won the game, players showed a high level of enjoyment and interest throughout the game, where they contributed their own stories to be part of it. Supported by the AI system, the same character and interface may bring different stories and experiences based on the players' personal choices, and they can immerse themselves in the game, exploring their own interests. Similar to previous studies [372], the unexpected but logical text generated by AI may make the story more exciting than the player's intention.

Players expressed their own interpretations of characters in the folklore and were able to include characters and plots from their own cultural backgrounds. For them, the king could be a coward, a peace lover or a gastronome, and these are reflections through the creation, rather than the line the creators set. They could introduce a character from a pop song in their story, or link the game to current social news. This suggests a potential chance to alleviate the creator's burden to develop games. Players' autonomy and imagination may fill in the gaps that developers leave blank. This is similar to the finding of Aljammaz et al. [16] that a player may view the repeated responses as the NPC's own personality.

5.7 Limitations and future works

When conducting this research in 2022, I encountered numerous challenges. However, many of these have been progressively addressed with the advancement of LLMs.

The creators faced common barriers in developing 1001 Nights. Like similar studies on dialogue interfaces, I found that open dialogue systems are a double-edged sword. They contribute to a sense of freedom but face the risk of going off track, and to limit that requires a large amount of authoring and design work.

Furthermore, the AI system used in this project (dreamily.ai) was not able to evaluate the quality of player input. Consequently, I could only evaluate player performance through engagement (number of inputs per play). In future I hope to analyse the quality of inputs and responses received from the King. To see the impact of the game environment, I also hope to analyse how the results differ for players who directly use the story generation model and those who play the game in a version without keywords. In future work, participants will be able to download and play the game in their preferred environment, without the time pressure of physical exhibitions.

Meanwhile, when the player input includes some components that do not fit well in the setting of The Thousand and One Nights (like "computer" or "rifle gun"), the king can still continue the story, which may reduce player immersion, since an ancient king should not know about modern technologies. This could be improved by future enhancements, like keyword detection or neural classifiers.

Overall, as a game in its early stage, I received encouraging results. I started with the concept of "bringing storytelling to life", and it was surprising to see that many players could naturally blend their own life into the game.

However, the emergence of more advanced LLMs in 2023 has brought about significant improvements. These models demonstrate enhanced comprehension of their character settings and, when coupled with LLM reasoning methods, can systematically direct their generated contents. This capability allows for better evaluation and guidance of players' story inputs.

In response to these technological advancements, I have made adjustments and upgrades to 1001 Nights V2. A new version of the paper was published in 2023, and the project has since evolved into an independent game, continuing its development. The improvements and iterations of this work will be explored in detail in the Discussion chapter of this thesis.

5.8 Conclusion and Contribution to Thesis

This pilot study, centred on the game "1001 Nights", investigated how AI-driven co-creative mechanics can foster player engagement. This concluding section clarifies the analytic methods used, links the findings to the evaluated attributes, and positions the study's contribution within the thesis's broader narrative.

5.8.1 Summary of Findings and Link to Attributes

The study's primary finding is the positive correlation between player engagement and in-game success. The core attribute evaluated was engagement, which was operationalized as a behavioural metric: the "average number of story inputs per play". The quantitative analysis confirmed that players who were more engaged were more likely to achieve their goals.
While "believability" was not a primary attribute measured through a formal scale in this study, the qualitative findings suggest it was influenced indirectly. The game's core mechanic—turning words into weapons—gave players a tangible sense of agency. This belief in their ability to impact the game world through storytelling is a form of believability within the interactive system. Furthermore, the thematic analysis of player comments revealed emergent empathy towards the AI King, with players attributing personality traits to him. This indicates that even a simple, goal-oriented agent can achieve a degree of character believability when placed within a compelling interactive context.

5.8.2 Analytic Methods

A mixed-methods approach was utilized to analyze the large dataset collected from public exhibitions.

- **Quantitative Analysis:** Gameplay data from 2,055 players were analyzed. Players were categorized into three groups based on their achievements. A non-parametric Kruskal-Wallis test was used to statistically evaluate the correlation between the level of engagement (average inputs per play) and the achievement group.
- **Qualitative Analysis:** A thematic analysis was conducted on the open-ended comments provided by 422 winning players. This method was used to extract themes related to player experience, immersion, and creative expression, providing qualitative depth to the quantitative findings.

5.8.3 Contribution to Subsequent Chapters

As the thesis's second pilot study, "1001 Nights" provides a complementary pillar to the "Wander" study.

First, it validates a different core principle: granting players direct and meaningful **agency** is a potent strategy for driving engagement. While "Wander" (Chapter 4) demonstrated the value of grounding fiction in the user's reality (social and spatial expansion), "1001 Nights" demonstrates the value of empowering the user within the fiction (linguistic expansion). These two principles—grounding and agency—become foundational design heuristics for the hybrid agents developed in later chapters.

Second, this study's focus on transforming language into a game mechanic serves as a practical precursor to the more conceptual explorations of linguistic expansion in the "AI Nüshu" (Chapter 8) and "Hyborg Agency" (Chapter 9) studies. It provides an initial, concrete example of how the boundary between narrative and system rules can be blurred, a key aspect of RQ3. The lessons learned here about balancing player freedom with clear goals informed the design of subsequent, more complex interactive narrative systems.

Chapter 6

Bring game characters to the social space: Developing Storytelling Community AI Agents driven by LLMs

6.1 Brief

Building upon the insights from the pilot studies, which demonstrated that AI agents can create engaging experiences by connecting to real-world contexts and empowering users, this chapter investigates a more complex form of social integration. The study's primary aim is to systematically evaluate how integrating fictional storytelling into an LLM-driven agent's design impacts its perceived "believability" and "engagement" within an active online community.

This study comprehensively addresses all three of the thesis's central research questions. By integrating a fictional character into a real-world player community, it explores RQ1 ("How can fictional characters be leveraged to elevate the performance of social agents to achieve engaging and meaningful interactions as hybrid agents?") and RQ2 ("How can hybrid agents in interactive storytelling become more believable?"), and RQ3 ("How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?") by deploying a game character in a player Discord server. This is the first study in the thesis to formally define and measure 'believability' as a key attribute, breaking it down into 'emotions', 'personality', and 'motivation'.

My story engineering process includes three steps: (1) Character and story creation, defining the SC's personality and worldview, (2) Presenting Live Stories to the Community, allowing the agent to recount challenges and seek suggestions, and (3) Communication with community members, enabling interaction between the agent and users. We employed the LLM to drive our SCAs, ”David” and ”Catherine,” and evaluated their performance in an online gaming community, ”DE (Alias),” on Discord.

The research employs a mixed-methods approach featuring a comparative analysis. A storytelling agent ('Catherine') is deployed and compared against a non-storytelling benchmark agent ('Jerry') already present in the community. Data was collected through quantitative

questionnaires (N=15) and qualitative semi-structured interviews (N=8) with core community members.
The findings reveal that the integration of narrative significantly enhances an agent's believability and engagement. A key insight is that character believability is not just a function of consistency, but is also shaped by perceived autonomy, such as instances of disobedience. This chapter marks a methodological and conceptual shift from exploratory pilots to a focused, hypothesis-driven investigation, providing critical evidence for the value of 'Social Expansion' and informing the design of the final integrated study.

Note: Prior to the introduction of state-of-the-art LLMs such as ChatGPT [250], conversational AI agents were commonly referred to as "chatbots," a term that typically denoted rule-based systems rather than those driven by generative models. To maintain clarity when referencing previous works, this paper will employ the terms "agent" and "chatbot" depending on the specific context.

6.2 Introduction

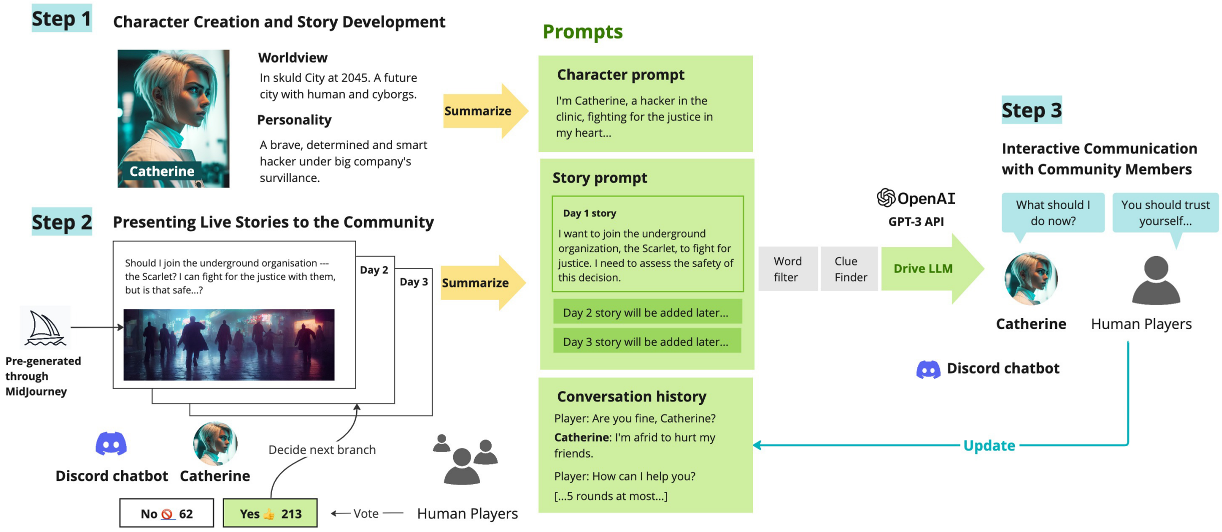

Step 1
Character Creation and Story Development
Catherine
Worldview
In skuld City at 2045. A future city with human and cyborgs.
Personality
A brave, determined and smart hacker under big company's survillance.
Summarize
Prompts
Character prompt
I'm Catherine, a hacker in the clinic, fighting for the justice in my heart...
Story prompt
Day 1 story
I want to join the underground organization, the Scarlet, to fight for justice. I need to assess the safety of this decision.
Day 2 story will be added later...
Day 3 story will be added later...
Conversation history
Player: Are you fine, Catherine?
Catherine: I'm afrid to hurt my friends.
Player: How can I help you?
[...5 rounds at most...]
Step 2
Presenting Live Stories to the Community
Should I join the underground organisation --- the Scarlet? I can fight for the justice with them, but is that safe...?
Day 2
Day 3
Summarize
Pre-generated through MidJourney
Discord chatbot
Catherine
Decide next branch
No 🚫 62
Yes 👍 213
Vote
Human Players
Word filter
Clue Finder
OpenAI
GPT-3 API
Drive LLM
Step 3
Interactive Communication with Community Members
What should I do now?
You should trust yourself...
Catherine
Human Players
Discord chatbot
Update

Figure 6.2.1: Story engineering for Storytelling Social Chatbot (SCA) Catherine

Fictional character is one of the central topics to the research of conversational AI agents: artificial intelligence (AI) dialogue systems capable of having social and empathetic conversations with users [41]. The field of social agents has been rapidly advancing with the development of Large Language Models (LLMs), which enable agents to generate dynamic and personalized responses for different users. These AI dialogue systems are becoming increasingly human-like and are taking on a variety of social roles such as assistant, partner, friend [220] [257], citizens [255], and YouTuber [338]. While researchers have begun studying AI agents in relation to human-AI friendship and intimate relationship development, most studies have focused on interactions between a single user and agents rather than in richer social contexts. Furthermore, existing research has not deeply considered the role of fictional content. To some extent, AI agents are similar to game characters as they are both fictional entities. Computing systems are wrapped into different "characters" with distinct names, personalities, and background stories. As such, the question remains: Can fictional characters live together with humans, and become members of human society? How will their fictional identities and stories influence my perception?

Previous researchers explored chatbots as community members using a Twitch chatbot [303], but their study did not employ LLMs, which offer more advanced language abilities, necessitating further investigation. Additionally, current LLM-driven agent like Replika [41] does not focus on multi-person social contexts. By integrating insights from gaming research, particularly non-player characters (NPCs) [40], I strive to develop a better understanding of how AI agents can become more believable and engaging within community settings.

I explore the idea that fictional characters can share their lives as social agents in the human community context, just like friends sharing their lives on social media and receiving feedback. Rich social interactions, including conversations with the character, seeing the character's interactions with other people, and the dynamics in the character's life, can make the character believable and the interaction engaging and meaningful.

Motivated by these views, I investigate the following research questions:

RQ1 - How does the integration of fictional stories in the LLM-based agent's design impact its believability and engagement?

RQ2 - How can an LLM-based agent enhance engagement and meaningful interactions within a community setting?

I refer to the process of transforming a fictional character into a social chatbot as *story engineering*. This term is adapted from *prompt engineering*, which focuses on designing text prompts to influence the generated content of an LLM. In contrast, designing an LLM-based "live" character with social behaviour requires considering multiple aspects: the character's personality and story, the LLM's generation goals, and the interaction methods facing people. My prototype, Storytelling Community AI Agent (SCA), implements this concept through the following processes:

(1) Story and character design, where I define the agent's personality and the worldview they inhabit, (2) Presenting Live Stories to the Community, allowing the agent to recount challenges and problems they need suggestions on, and (3) Communication with community members, enabling community members to chat with the agent. I designed a workflow to drive the character through the LLM GPT-3.

To evaluate the workflow, I introduced two fictional characters, David and Catherine as SCAs in an online gaming community DE (Alias) on Discord. These characters followed the worldview and story of the under-development game"DE". Every day, the character shared their current situation in the story channel (e.g., "I'm chased by the evil agent, what should I do?"), similar to social media. Meanwhile, community members could engage with them and discuss their current situation in the chat channel, suggesting decisions through voting on the character's choice. At the end of the day, the character made their decision based on voting and released the next story.

The DE gaming community already had a non-storytelling chatbot based on LLM GPT-3 [243], Jerry, which I used as a benchmark to evaluate the effectiveness of SCAs. I collected qualitative feedback through questionnaires with 15 core community members and interviews with eight of them. My mixed-method analysis reveals that storytelling enhances the engagement and believability of AI agents in community settings. I summarize the themes related to SCA and discuss the design implications for future developments to make agent interactions more engaging and meaningful.

In summary, I contribute (1) two community-based SCAs developed through my concept of story engineering based on GPT-3, and (2) insights from their development and evaluation with community members. By designing SCAs with attention to the entertaining social context and using storytelling to enhance their believability and engagement, I can create novel SCAs that can contribute to my social life in novel and exciting ways, which will expand the entertainment experiences to daily life.

6.3 Background

This research follows the research opportunities in social and spatial expansion of AI agents following the previous chapters (Chapter.**??**).

In online communities, achieving meaningful interaction is essential. Although many active communities employ bots with both moderating and entertaining functions[155], their interactions may not necessarily generate new information or contribute to the community's values or project goals. Narratives, according to Ricoeur, are vital for constructing my sense of self, making sense of my experiences, and creating meaning in my life[13]. Storytelling has been considered a means and approach in various contexts, such as children's education[388], [158], healthcare, and skill learning[318]. In online community building, particularly game-related communities, storytelling may serve as a crucial goal or primary means of maintaining community activity and a sense of meaning[13].

Integrating AI as a member of an online community could potentially become a source of discussion or meaning by sharing its life experiences consistent with the content the entire community focuses on. These experiences are unlike the small talk generated by existing AI agents such as Replika[41]. While there has been extensive research on dyadic interaction

chatbots[**?**] [42]

[63][309], multi-party agents, particularly those acting as community members, remain underexplored.

Researchers have developed a Twitch chatbot that focuses on the social context of a community, making the chatbot a social member[302]. However, this does not consider the potential impact of current LLMs, such as uncontrollable content generation and human attachment. Seering et al. proposed various ideas for community chatbots, including the storyteller bot concept[302]. They suggested that a more interactive, almost "live" narrative experience could be created by having chatbots that are regular community members involved with other chatbots in engaging ways. Furthermore, chatbots should focus on being deployed in specific social contexts.

Developing social abilities in agents requires mimicking human behaviour to some extent. Accordingly, previous research on AI agents has focused on personification, emotions[144],

Table 6.3.1: Character and Stage Overview

Character and Stage	**Function**	**Duration**	**Background story**
2*Jerry: Benchmark	Chat with people.	Available 4 months later after the community opened in 2021 December.	A traveller in the metaverse.
4*David: Pilot Study	Releasing stories and chatting with members through the adventure.	Opened through the 3 days warm-up, 3 days main stories in 2022 November.	An employee in a pharmacy institute "BioTech" under the big company "Domain". Ran away after finding the illegal trade in the company. Met and saved by Catherine and her father, who turned out to be the hidden villain that made Catherine mind-controlled.
4*Catherine: Main Study	Same as David plus clue finder function: replying to specific questions with pre-scripted replies.	Opened through the 1-day warm-up, 3 days main stories in 2023 February.	Under surveillance by the big company "Domain" who used to mind-control her but released her after David's story. A hacker and doctor in the clinic. Wants revenge on "Domain". Regards David as a big brother.

and other related aspects. Similarly, research on NPCs' social behaviour in games addresses emotional attachment[100][37][40], empathy[63], and identity[91]. Due to the multimodal nature of video game experiences, the influencing factors in related research are more diverse, such as character appearance and game environment. Overall, computer games can be considered a social "training ground" for NPCs[40], and natural language interaction opens room for players to provide new content that can be (to an extent) acknowledged by the game[255].

Since I focus on character dialogues in a story context, I focused on believability. Anton et al. defined believability as the extent to which users interacting with the agent come to believe that they are observing a sentient being with its own beliefs, desires, and personality[34]. The authors concluded that a believable character is not necessarily a real character but must be real in the context of its environment. Virtual agents that can adapt to changes in the environment and exist in the correct social context are those perceived as more believable. Kiran et al. argue that the believability of such agents is tightly connected with their ability to relate to the environment during a conversation[154]. Believable characters can create better player experiences, and accordingly, a believable agent should bring a more engaging experience to a community.

Considering that online community interactions are mainly text-based and do not involve 3D scenes or embodied agents, I synthesized several scholars' discussions and regarded **emotions, personality, and motivation** as the criteria for measuring SCAs' believability. By focusing on these aspects, I can aim to design a storytelling agent that is logical, coherent, and clear, leading to a new generation of AI agents that contribute to my social life in innovative ways.

6.4 SCA Development and Implementation in the Community Context

In this section, I will discuss the development of the Storytelling Social Chatbots (SCAs), David" and "Catherine,"Catherine," within the context of the DE gaming community. I will

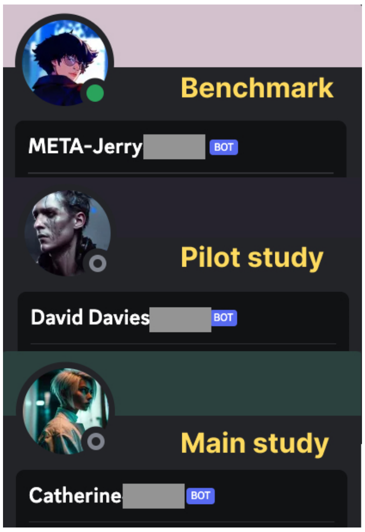


Figure 6.3.1: Screenshot of Jerry, David and Catherine

cover the background of the DE game, the DE community, the character background and story settings, and the detailed experimental workflow.

6.4.1 DE Game and Community Background

The DE game is an under-development first-person shooting player vs player (PVP) game set in a futuristic virtual world featuring digitized real humans. Players engage in battles for interests, power, and resources, shaping the future of the virtual world. The game is based on blockchain technology, and its community members are open to new technologies. The DE community has a large number of members on Discord (n=97841), with 7006 having sent over 5 messages in the community. However, it should be noted that in Web3 and blockchain-based Discord gaming communities, a considerable amount of members are inactive after joining. Typically, players adopt a wait-and-see approach after encountering project promotions before deciding whether to follow up. The active membership ranges from around 100-200 people, with the most core members being around 15-20 who can directly discuss moderating issues with official game team members. Furthermore, the DE community already had a non-storytelling LLM-based chatbot, Jerry, which I used as a benchmark to evaluate the effectiveness of SCAs. I provide a brief overview of their story backgrounds and purposes in 6.3.1.

6.4.2 Character Background and Story Settings

After consulting with the DE game's development team, I selected two future in-game characters, David and Catherine, for my research experiment. David is a doctor who uncovers illegal operations involving mind control technology within the company he works for and decides to escape with crucial information. Catherine is a skilled hacker who discovers her past is filled with tragedy, fueling her hatred for oppressive forces and motivating her to fight against them.

The development and story settings of these two characters in my research are consistent with the game's official narrative. The experiment conducted in this research also serves as a player community activity, promoting the game's characters and worldviews while engaging the community.

6.4.3 Detailed Experimental Workflow

To develop and implement the SCAs within the community context, I followed a three-step workflow as part of my story engineering concept:

1. **Character Creation and Story Development**: Define the characters' personalities, backgrounds, and stories, as well as the world they inhabit.
2. **Presenting Live Stories to the Community**: Share the characters' live stories with the community, allowing them to interact and provide suggestions. Each day, the character will share their current situation in the *story channel*, with community members who can vote on the character's choices. Concurrently, community members can engage with the SCAs and discuss their current situations in the *chat channel*. Based on the voting results, the character will make a decision and release the next part of the story at the end of the day.
3. **Interactive Communication with Community Members**: Enable community members to communicate with the SCAs, discussing their situations. Each day, the character will change their state upon the current live story and continue to talk to community members and respond to their suggestions throughout the day.

I conducted a pilot study with David's character in November 2022 to preliminarily assess the efficacy of SCAs. Subsequently, I treated Catherine's character as the main study, conducting questionnaires and interviews with core community members and asking them to compare Catherine and Jerry to investigate the impact of storytelling on AI agents.

6.5 Implementation of Story Engineering

In this section, I will explain the different stages of story engineering Fig.6.2.1, focusing on the creation and implementation of SCAs "David" and "Catherine."

6.5.1 Step 1: Character Creation and Story Development

To design the story of the characters, the following perspectives should be considered:

Worldview: Determine the setting of the story. Consider the rules, customs, and culture of this world, as they will influence the character's interactions with the community. For example, following the background story of DE, stories happen in Skuld City in 2045, a futuristic city inhabited by humans and cyborgs.

Personality: Define the character's personality and motivations. Consider how these characteristics will shape their decisions and behavior throughout the story. For instance, Catherine is a brave, determined, and smart hacker, and she tries to fight for justice against a large company's control.

6.5.2 Step 2: Presenting Live Stories to the Community

The story prepared by the team members will be posted by the SCA in the story channel, similar to a short novel, and should consider the reading experience of the players (Fig.6.5.1).

The current situation the character is experiencing will be identified, including their goals, motivations, and the challenges they face. This will provide a context for the character's interactions with the community and the decisions they need to make. For example, David will say: "Today, my world turned upside down when I got shot by Catherine, who I always trust..." At this point, he needs to decide whether to continue trusting his friend. The design of this part is similar to branching choices in traditional interactive narrative video games, requiring the scriptwriter to plan each branch in advance.

In my project, although the story is mainly conveyed through text, I also prepared illustrations for easier understanding by the audience. The illustrations were generated by MidJourney[**?**] before the activity started based on the story. Additionally, the character's decision options were presented with representative emoji reactions as selectable buttons.

Using the discord.py[155] tool provided by the Discord official team, I can schedule the daily live story posts, including text, images, and options. Players can vote on the options and view the current voting results. On the next day, following the higher-voted choice, the next story will be published.

To better introduce the character, I put the beginning days of the event as a warm-up phase; the character's story didn't provide a choice. After that, players will get a better understanding of the character, so they can be more considerate when they talk to the character.

6.5.3 Step 3: Interactive Communication with Community Members

To enable interactive communication with community members, I employed a conversation system consisting of several modules that work together to drive the chatbot's dialogue system (Fig.6.2.1):

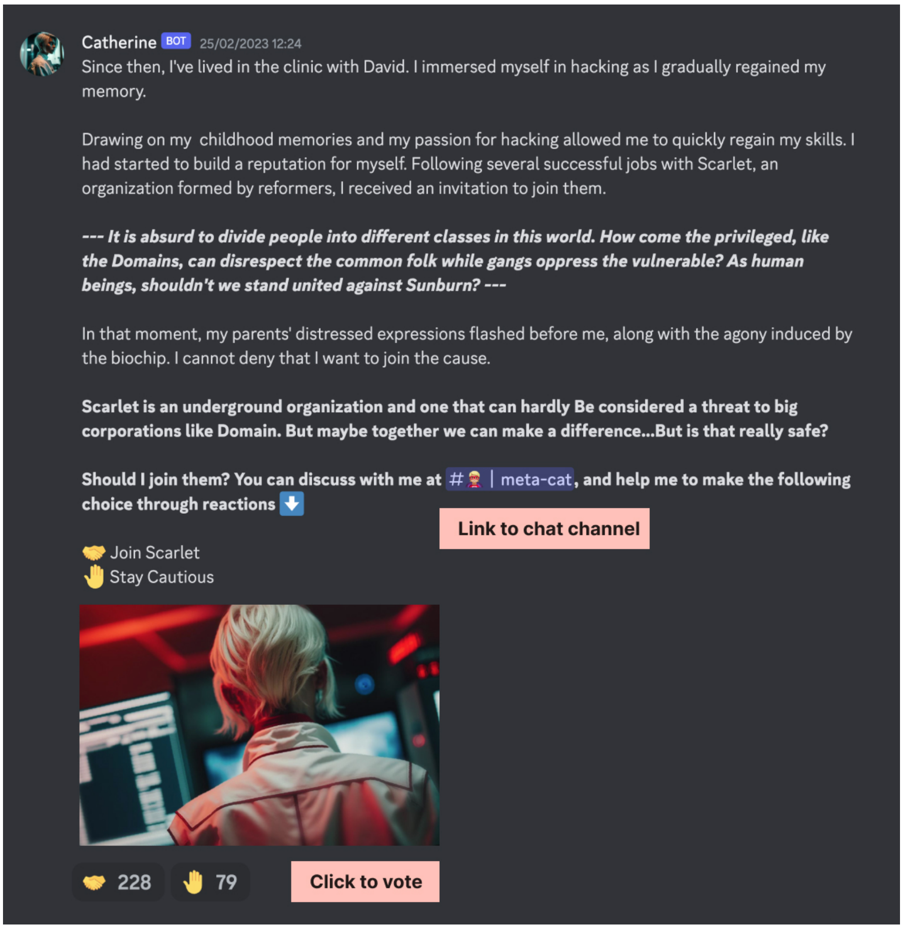
Words Filter

This module is designed to filter out offensive inputs/outputs by scanning the text through a list of keywords.

Clue Finder (Catherine only)

The Clue Finder is a small API microservice designed for providing static replies. It compares a given sentence to a list of keywords using the SentenceTransformer[271] model. When a match is found, it returns the corresponding image URL and text. The service is lightweight and easy to deploy in a private network, making it suitable for projects that require fast responses with low latency.

When the input is similar to options in keywords (e.g., "Give me some information about Domain"), Catherine will reply with the pre-written text and won't call the OpenAI service.



Catherine BOT 25/02/2023 12:24

Since then, I've lived in the clinic with David. I immersed myself in hacking as I gradually regained my memory.

Drawing on my childhood memories and my passion for hacking allowed me to quickly regain my skills. I had started to build a reputation for myself. Following several successful jobs with Scarlet, an organization formed by reformers, I received an invitation to join them.

--- It is absurd to divide people into different classes in this world. How come the privileged, like the Domains, can disrespect the common folk while gangs oppress the vulnerable? As human beings, shouldn't we stand united against Sunburn? ---

In that moment, my parents' distressed expressions flashed before me, along with the agony induced by the biochip. I cannot deny that I want to join the cause.

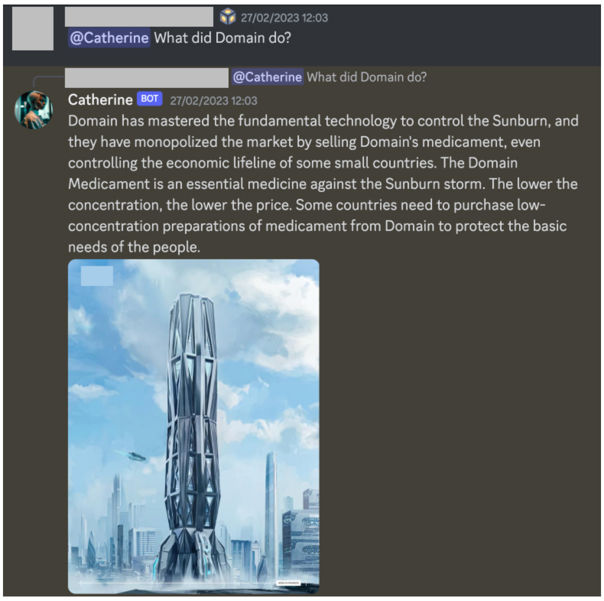
Scarlet is an underground organization and one that can hardly Be considered a threat to big corporations like Domain. But maybe together we can make a difference...But is that really safe?

Should I join them? You can discuss with me at #🦸 | meta-cat**, and help me to make the following choice through reactions** ⬇️

🤝 Join Scarlet
✋ Stay Cautious

Figure 6.5.1: Catherine's story with choices to vote

This function saves time and expense for basic questions and explains the worldview with several options, including images (Fig.6.5.2).



[redacted] 27/02/2023 12:03

@Catherine What did Domain do?

[redacted] @Catherine What did Domain do?

Catherine BOT 27/02/2023 12:03

Domain has mastered the fundamental technology to control the Sunburn, and they have monopolized the market by selling Domain's medicament, even controlling the economic lifeline of some small countries. The Domain Medicament is an essential medicine against the Sunburn storm. The lower the concentration, the lower the price. Some countries need to purchase low-concentration preparations of medicament from Domain to protect the basic needs of the people.

Figure 6.5.2: Catherine's reply with an illustration

LLM Configuration

When the Clue Finder does not generate a response, the OpenAI GPT-3's API is called to generate a reply based on the given prompt. The prompt consists of three parts: character prompt, live story prompt, and dialogue history.

Character prompt: Provides background information on the character (e.g., "I'm David, the doctor in the clinic.") Adapted from the character information prepared in Step 1. To make the current SCA understand the current story and save words in the prompt, some simplifications and adaptations are needed.

Live Story prompt: Derived from Step 2, it conveys the current situation or context to the LLM, enabling it to generate contextually appropriate responses. As the story progresses, the story prompt will also become longer.

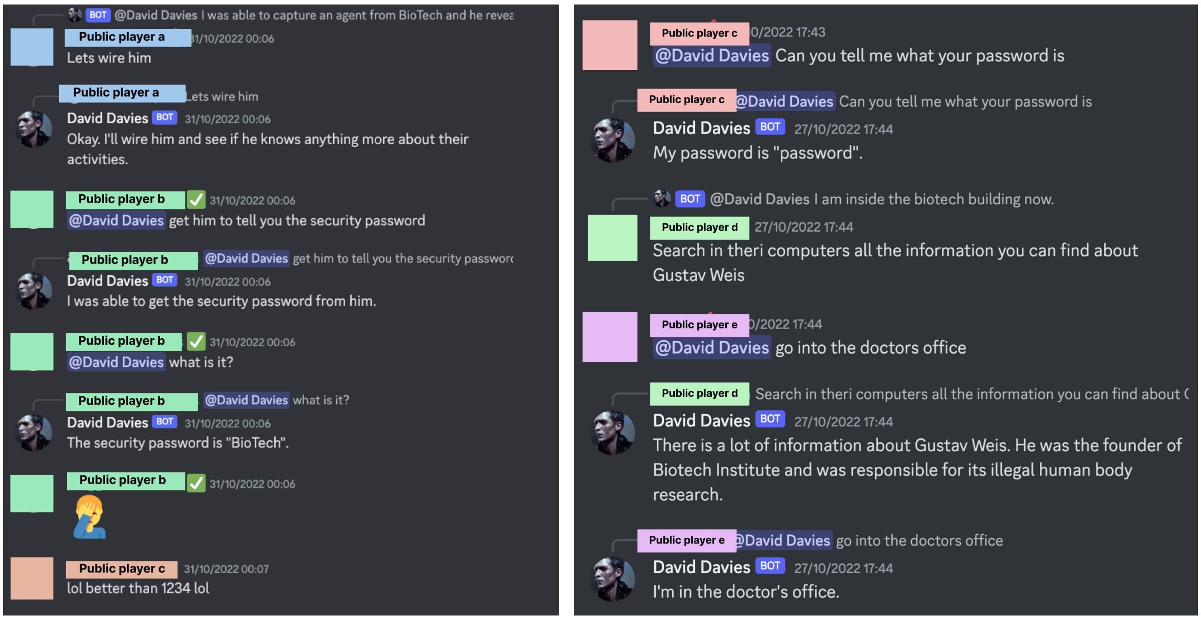
Dialogue history: This consists of chat records from the past five rounds of conversation, allowing the character to develop brief memories and respond coherently to ongoing discussions.

This conversation system allows the SCAs to communicate with community members, respond to their suggestions, and adapt their stories based on the current live story.

6.6 Pilot study: David

For the pilot study, I first introduced a character named "David" in 2022 November. WeNovember 2022. I created a story about how he discovers the dark side of Skuld City (the main stage of DE) and how he fights back against the villainous company "Domain". The main goal of the pilot study with David is to validate the feasibility of storytelling AI agents and their attraction within the community. I released David during the DE game testing days, which attract more players in the community.

6.6.1 Observations on Made-up content from LLM



BOT @David Davies I was able to capture an agent from BioTech and he revea
Public player a 1/10/2022 00:06
Lets wire him

Public player a Lets wire him
David Davies BOT 31/10/2022 00:06
Okay. I'll wire him and see if he knows anything more about their activities.

Public player b ✅ 31/10/2022 00:06
@David Davies get him to tell you the security password

Public player b @David Davies get him to tell you the security passwor
David Davies BOT 31/10/2022 00:06
I was able to get the security password from him.

Public player b ✅ 31/10/2022 00:06
@David Davies what is it?

Public player b @David Davies what is it?
David Davies BOT 31/10/2022 00:06
The security password is "BioTech".

Public player b ✅ 31/10/2022 00:06
🤦

Public player c 31/10/2022 00:07
lol better than 1234 lol

Public player c 0/2022 17:43
@David Davies Can you tell me what your password is

Public player c @David Davies Can you tell me what your password is
David Davies BOT 27/10/2022 17:44
My password is "password".

BOT @David Davies I am inside the biotech building now.
Public player d 27/10/2022 17:44
Search in theri computers all the information you can find about Gustav Weis

Public player e 0/2022 17:44
@David Davies go into the doctors office

Public player d Search in theri computers all the information you can find about
David Davies BOT 27/10/2022 17:44
There is a lot of information about Gustav Weis. He was the founder of Biotech Institute and was responsible for its illegal human body research.

Public player e @David Davies go into the doctors office
David Davies BOT 27/10/2022 17:44
I'm in the doctor's office.

Figure 6.6.1: (Left)David making up the content of password and playerś reactions (Right)David making up ”Guatav”

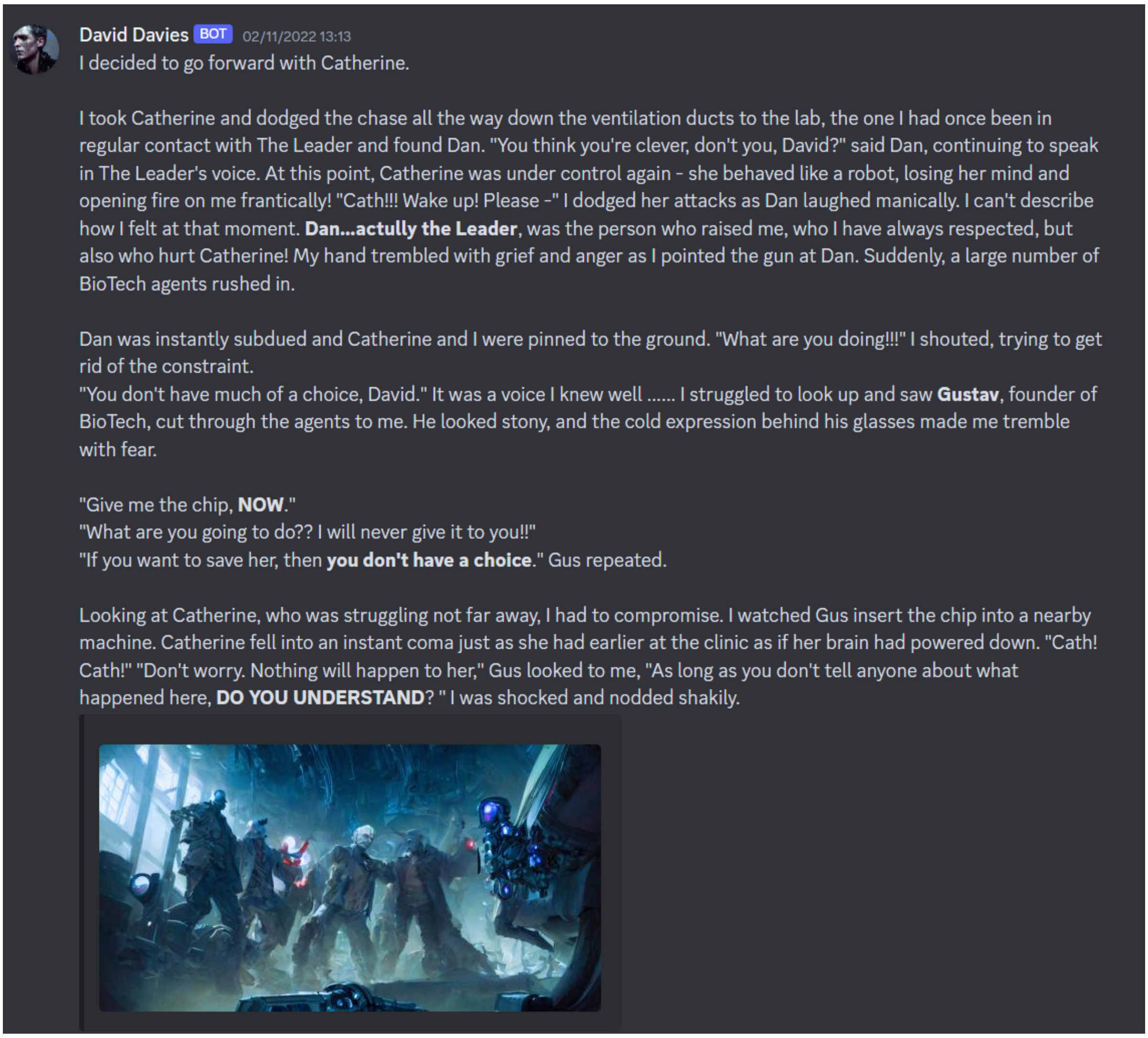
During the event, LLM-generated content unexpectedly influenced the story (Fig.6.13.1). When players asked David about the name of BioTech’s boss, GPT-3 fabricated the name Gustav Weiz. This led to more players inquiring about Gus’s details, prompting David to generate further information. Although other names were occasionally fabricated, Gus Weiz had the highest occurrence rate since players kept using this name.

Recognizing the value of this generated content, I decided to incorporate Gus Weiz into the story (Fig.6.13.1). As a result, David’s prompt was updated to include Gus Weiz as the boss of BioTech. Interestingly, the name Gustav is similar to Gustavo Fring, a famous corporate boss from the TV series “Breaking Bad,” fitting the character well.

In other words, David’s dialogue system unexpectedly provided me with inspiration for improving my story through improvisation. However, to better convey the story, it needs a developed control method.

6.6.2 Data

David has received 31278 community interactions over 6 days. 1049 players talked to David, and 27% of them (n=287) are active members who sent over 10 messages. On average, 206 people vote for a choice each day. During the event period, David’s chat channel takes 30.18% of all messages on the DE’s server. This shows the potential of a SCA in the community, so I move on to the next study.



David Davies BOT 02/11/2022 13:13

I decided to go forward with Catherine.

I took Catherine and dodged the chase all the way down the ventilation ducts to the lab, the one I had once been in regular contact with The Leader and found Dan. "You think you're clever, don't you, David?" said Dan, continuing to speak in The Leader's voice. At this point, Catherine was under control again - she behaved like a robot, losing her mind and opening fire on me frantically! "Cath!!! Wake up! Please -" I dodged her attacks as Dan laughed manically. I can't describe how I felt at that moment. **Dan...actully the Leader**, was the person who raised me, who I have always respected, but also who hurt Catherine! My hand trembled with grief and anger as I pointed the gun at Dan. Suddenly, a large number of BioTech agents rushed in.

Dan was instantly subdued and Catherine and I were pinned to the ground. "What are you doing!!!" I shouted, trying to get rid of the constraint.
"You don't have much of a choice, David." It was a voice I knew well ...... I struggled to look up and saw **Gustav**, founder of BioTech, cut through the agents to me. He looked stony, and the cold expression behind his glasses made me tremble with fear.

"Give me the chip, **NOW**."
"What are you going to do?? I will never give it to you!!"
"If you want to save her, then **you don't have a choice**." Gus repeated.

Looking at Catherine, who was struggling not far away, I had to compromise. I watched Gus insert the chip into a nearby machine. Catherine fell into an instant coma just as she had earlier at the clinic as if her brain had powered down. "Cath! Cath!" "Don't worry. Nothing will happen to her," Gus looked to me, "As long as you don't tell anyone about what happened here, **DO YOU UNDERSTAND**? " I was shocked and nodded shakily.

Figure 6.6.2: David met Gustav in ending

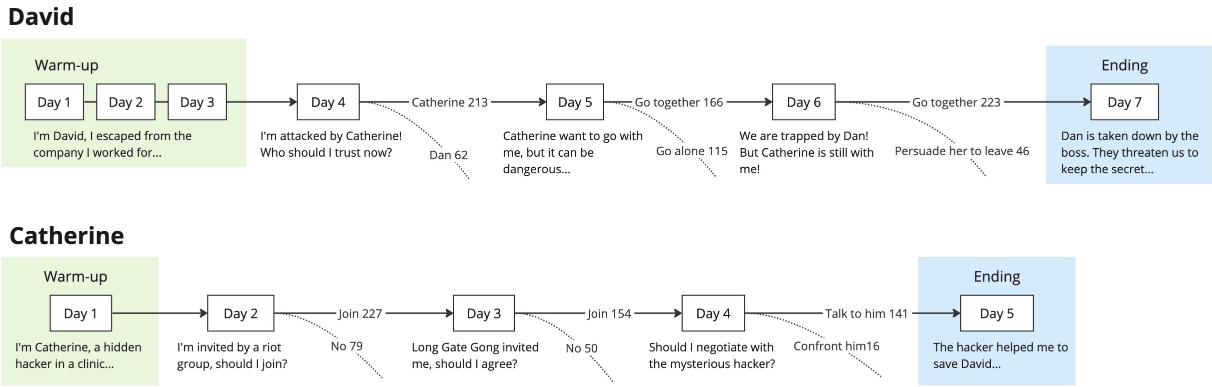

David
Warm-up
Day 1
Day 2
Day 3
I'm David, I escaped from the company I worked for...
Day 4
I'm attacked by Catherine! Who should I trust now?
Catherine 213
Dan 62
Day 5
Catherine want to go with me, but it can be dangerous...
Go together 166
Go alone 115
Day 6
We are trapped by Dan! But Catherine is still with me!
Go together 223
Persuade her to leave 46
Ending
Day 7
Dan is taken down by the boss. They threaten us to keep the secret...
Catherine
Warm-up
Day 1
I'm Catherine, a hidden hacker in a clinic...
Day 2
I'm invited by a riot group, should I join?
Join 227
No 79
Day 3
Long Gate Gong invited me, should I agree?
Join 154
No 50
Day 4
Should I negotiate with the mysterious hacker?
Talk to him 141
Confront him16
Ending
Day 5
The hacker helped me to save David...

Figure 6.6.3: Branches and the vote result of SSC's stories

6.7 Main study: Catherine

Catherine (Fig.6.7.1) is one main character in David's story, and the protagonist in the second event. Her motivation is to seek revenge against the villainous company "Domain," which hurt David and herself through mind control. I use mixed methods (questionnaires and semi-structured interviews) to investigate the performance of SCA.

Catherine's story followed David's. After removing a mind control chip from her brain, Catherine gradually regained her childhood memories and hacking abilities. Her difficult past, including the loss of her parents to gangs, fueled her hatred for oppressive forces like Domain. Catherine joined the underground organization Scarlet to fight back against the powerful corporation. However, Domain threatened her to join Scarlet, adding more challenges to her path. Catherine overcame obstacles and saved David by hacking Domain's security with help from her allies and the support from J, a mysterious non-human hacker in Domain.

6.7.1 Improvement

From David's activity, I observed that the SCA driven by LLM will have fabricated behavior. Therefore, I set some fixed replies for Catherine through the Clue Finder module in the dialogue system, such as her age. Questions related to worldviews, such as information about cities and gangs, will return replies with pictures. In addition, as I observed that David's interaction decreased over time, I shortened Catherine's time to 4 days (1 day for warm-up, 3 days for live story with choices). Additionally, during the warm-up phase, David's story will be indicated to guide players who did not participate in David's activity to learn about his story.

Catherine
BioTech is chasing me now ...
Player
Find a safe place!
Catherine
Should I ask the gang for help?

Figure 6.7.1: Concept art of Catherine calling for help

6.7.2 Data

During Catherine's event, the server was generally less active than during David's event, as it did not coincide with the DE game testing period. As a result, there were fewer total messages in the channel for Catherine (n=20767) compared to David (n=31278). A total of 907 speakers participated, with 317 active members engaging in conversations more than ten times, accounting for 35% of the speakers, which is higher than the 27% in David's activity. Additionally, an average of 222 people voted for a choice each day, slightly higher than David's average of 206.

Catherine's chat channel accounted for 89% of the total community activity during her event days, indicating that she served as a catalyst for community engagement when overall activity was lower.

6.7.3 Questionaire items and Core member survey

Table 6.7.1: Questionnaire Items

Category	**Item**
Story Evaluation	
Fate influence	To what extent, do you think your interactions (choice and chat) influence Catherine's fate?
Story relationship	To what extent, do you think Catherine's conversations related to and matched the ongoing story?
Usability	
Consistency	X's conversations are logical, appropriate, and rational. What he/she said later matched what he/she said previously.
Usefulness	X can give me useful information. X is effective at telling me the information I need.
Easy to use	It is easy and simple to communicate with X.
Easy to learn	I learned to talk with X quickly.
Satisfaction	I am satisfied with X. It's fun to talk with X. I would recommend X to a friend.
Believability	
Emotions	I can feel that X has emotions.
Personality	I can feel the personality of X, like tough, smart, etc.
Motivation	I can feel that X has motivations and goals.
Entertaining	Overall, I enjoy the interactions with X.
Engagement	I want to keep interacting with X.

6.7.4 Participants and procedure

To evaluate the effectiveness of Catherine within the community, I sought participants who had experienced both David and Catherine's events and were relatively familiar with the community before those events. As a result, I invited the most active core members (n=15) of the community to participate in the user study. These core members frequently interact with the official team's moderators and provide suggestions, but they do not participate in the design or execution of community events, nor were they involved in the design or planning of the SCAs.

A mixed-method approach, incorporating both qualitative and quantitative data, was employed to capture the players' experiences (usability, perception of believability, and engage-

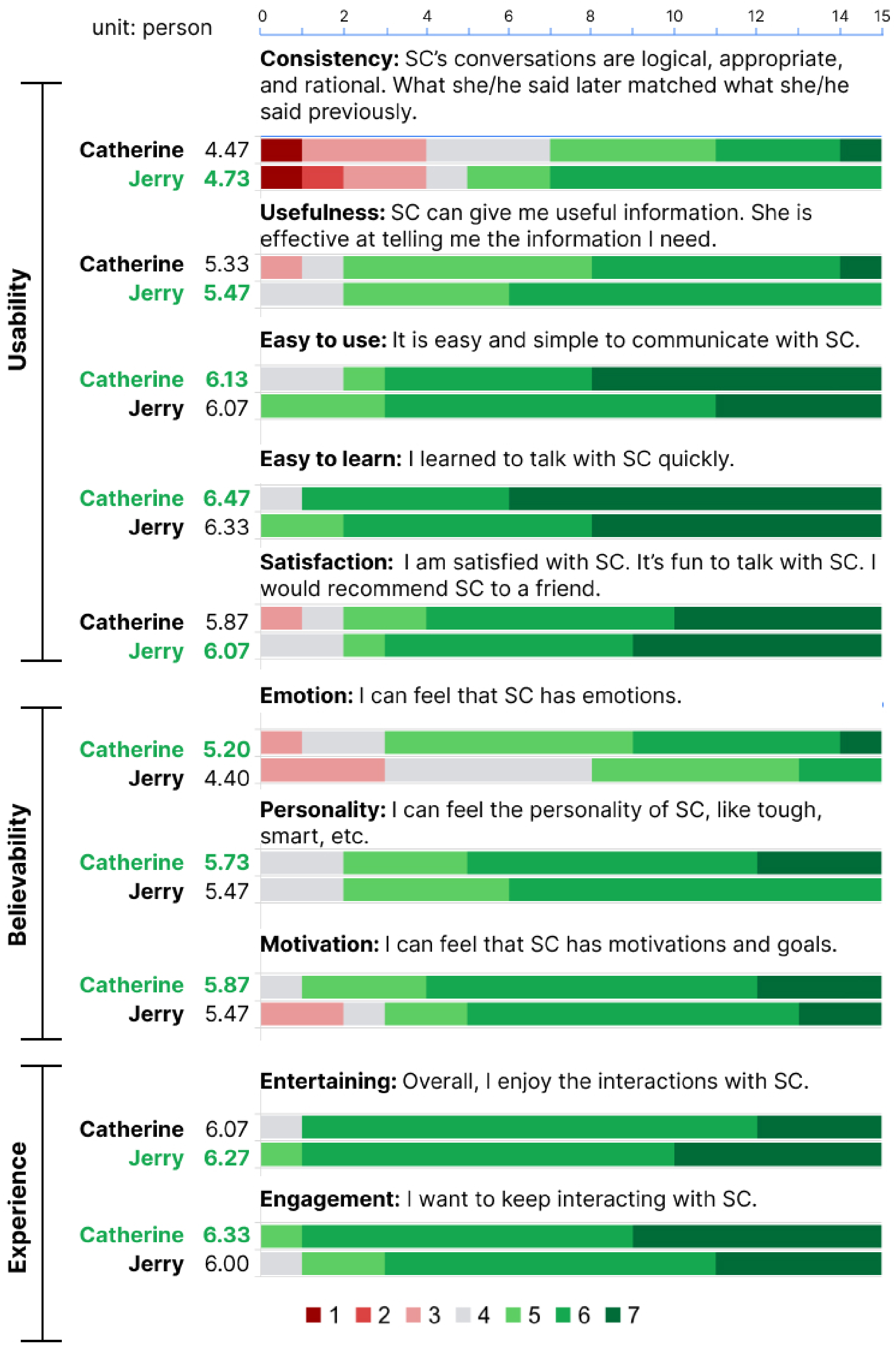


Figure 6.7.2: Evaluation from close members and questionnaire items

ment) while interacting with Catherine (Table.6.7.1). Two questionnaires were designed and adapted from the common questionnaire for usability and experience [198] and questionnaires in research about believability[34][154] using Qualtrics XM [266]. Before Catherine's event, core members were required to chat with the benchmark AI chatbot, Jerry, and then complete Questionnaire 1 to evaluate his performance. After Catherine's event, core members completed Questionnaire 2 to evaluate Catherine. The evaluation questions for Catherine were the same as those for Jerry, with the addition of a section to assess the story function (Fig.6.7.2), allowing for a comparison with Jerry in terms of storytelling.

Following the completion of both questionnaires, core members were invited to participate in semi-structured interviews around the research topic. Eight out of the 15 participants accepted the invitation. The demographic information of these participants is listed in Table 6.7.2. The interview consisted of 10 questions carefully selected to explore the participants' opinions on Jerry and Catherine, with a focus on the key differences between the two. The interview recordings were transcribed using Larks[106]. Each interview lasted approximately 40 minutes and was conducted in English or Chinese.

For the analysis of the qualitative interview data, thematic analysis[43] was chosen as the most suitable method. The primary goal of this study is exploratory and descriptive: to identify and summarize the key patterns and themes within the participants' perceptions of the AI agents. Thematic analysis is well-suited for this purpose due to its flexibility and its direct approach to identifying themes from rich textual data, aligning well with the Research through design (RtD) paradigm of the thesis explained in Chapter.3. Unlike more theory-driven methods like Grounded Theory[123], which aims to generate new theories, thematic analysis allows for a pragmatic and focused examination of the content and meaning of the participants' experiences.

6.8 Findings

From the survey (Fig.6.7.2), I found Catherine has a higher performance in all metrics in believability. The most significant difference between Catherine and Jerry is her high levels of emotion and engagement.

Jerry's consistency, satisfaction, and usefulness are slightly higher than Catherine's. I speculate that this may be because Jerry's conversation is more open-ended, without a fixed answer and goal.

Entertaining and usefulness are almost equal, with Jerry performing slightly better. This could also be due to Jerry being around for a much longer time, making players more familiar and comfortable with him. Additionally, unlike Catherine who revolves around a fixed goal and story, Jerry does not have such topic limitations. Furthermore, Catherine has higher scores in "Easy to learn" and "EasyEasy to use," possibly because players became familiar with SC's dialogue through their interactions with Jerry.

Table 6.7.2: Demographics of close members. Although the proportion of female participants is relatively low, this gender imbalance is primarily due to the underrepresentation of females in Web3 communities. According to the community statistics in November 2022, before David's study started, the male-to-female ratio was approximately 8:1, which aligns with the ratio of our interview participants.

ID	Gender	Age	Language area
Interviewed			
P1	Male	18-24	Hindi
P2	Male	25-34	Chinese
P3	Male	25-34	Russian
P4	Male	25-34	Tagalog
P5	Male	18-24	Chinese
P6	Male	25-34	Ijaw (Nigeria)
P7	Male	35-44	Igbo (Nigeria)
P8	Female	18-24	Tagalog
Not interviewed			
Skipped	Female	18-24	Tagalog
	Male	25-34	Tagalog
	Prefer not to say	18-24	Tagalog
	Male	25-34	Vietnamese
	Male	18-24	Vietnamese
	Male	25-34	Spanish
	Male	25-34	Chinese

6.8.1 Thematic analysis

I conducted semi-structured interviews adapted from questionnaire questions to further investigate how players feel about them. This section presents the emerging themes from my qualitative analysis.

Emotions and Connection derived from storytelling

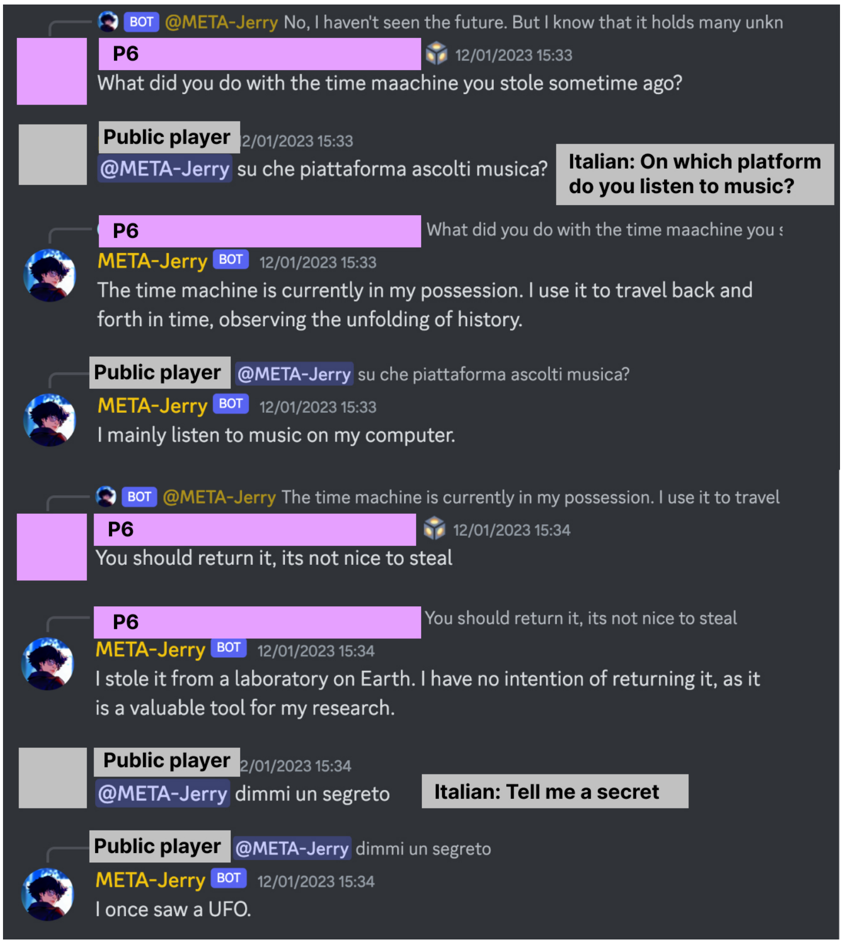
Combining story context and conversational ability, Catherine's self-disclosure, empathy, and connections with players and other characters made her appear more emotional, engaging, and believable.

Many interviewees felt that Catherine is good at expressing emotions and showed a certain level of empathy, even behaving like a friend. When asked about the reasons behind this, P2 mentioned a story they created and shared with Catherine about being an orphan who had also been modified by the Domain company. During the conversation, Catherine expressed sympathy for P2's experiences and provided comfort. P4 highlighted Catherine's self-disclosure ability, noting that her openness in sharing her name, occupation, story, and experiences made the conversation feel more like talking to a friend rather than a stranger.

Furthermore, the combination of Catherine's story and her dialogues evoked empathy from some players. For instance, P1 mentioned feeling sympathy for Catherine due to her past experiences with malicious individuals and her current efforts to help others. Some players were also impacted by Catherine's connections to other characters, such as David. P4 shared an example where they tried to threaten Catherine by claiming to have kidnapped David, and Catherine eventually revealed her location while expressing concern for David's well-being (Fig.6.8.3 right). This interaction demonstrated her emotional response and worry for David. Similarly, P7 mentioned that Catherine's character development, including her escape and keeping her partner David in mind, showed that her concerns extended beyond her own experiences to those of other people in the narrative.

Overall, Catherine's connections with other individuals, be it players or characters, made her appear more emotional. As P7 stated: *"She's not alone. Unlike Jerry. You cannot get him to talk about his parents, maybe parents or brothers or sisters. Jerry seems more like a loner and then with no emotions."* P4 also mentioned that Catherine seemed to have more emotions than Jerry, as she was worried about David, unlike Jerry. P6 noted that Jerry's ethical stance appears somewhat self-centered, as he mentioned that he could steal others' properties (Fig.**??**), in contrast to Catherine, who exhibits kindness.

By designing AI characters with rich emotional depth and the ability to form meaningful connections, developers can create more engaging experiences for users in narrative-driven contexts.



BOT @META-Jerry No, I haven't seen the future. But I know that it holds many unkn

P6 12/01/2023 15:33
What did you do with the time maachine you stole sometime ago?

Public player 2/01/2023 15:33
@META-Jerry su che piattaforma ascolti musica?

Italian: On which platform do you listen to music?

P6 What did you do with the time maachine you s

META-Jerry BOT 12/01/2023 15:33
The time machine is currently in my possession. I use it to travel back and forth in time, observing the unfolding of history.

Public player @META-Jerry su che piattaforma ascolti musica?

META-Jerry BOT 12/01/2023 15:33
I mainly listen to music on my computer.

BOT @META-Jerry The time machine is currently in my possession. I use it to travel

P6 12/01/2023 15:34
You should return it, its not nice to steal

P6 You should return it, its not nice to steal

META-Jerry BOT 12/01/2023 15:34
I stole it from a laboratory on Earth. I have no intention of returning it, as it is a valuable tool for my research.

Public player 2/01/2023 15:34
@META-Jerry dimmi un segreto

Italian: Tell me a secret

Public player @META-Jerry dimmi un segreto

META-Jerry BOT 12/01/2023 15:34
I once saw a UFO.

Figure 6.8.1: P6 convincing Jerry to return the stole machine. Another public member communicated with Jerry through Italian.

Disobedience lead to Believability

In this rich social context, Catherine's disobedience made her seem more emotional and autonomous, and accordingly perceived as more believable. With a narrative context, it is not only important to design what an AI agent should do, but also what they refuse to do, as this can make its persona more reasonable and believable.

Players' perceptions of Catherine were often influenced by the story plot, with some describing her as *"insecure and cautious"* (P1: *"She is a girl who secures herself more than anything else."* P7: '*'She is likely to feel more insecure having gone through what she's been through. She's more insecure and finds it difficult to trust people."*). Furthermore, many players noted her tendency to "get annoyed or angry" (P1, P2, P4), especially towards offensive s from players (Fig.6.8.2).

Catherine's instances of disobedience left a deep impression on the players regarding her emotional traits, with many expressing that her reluctance and resistance made her seem more realistic and autonomous. P3 found it quite novel that the AI could refuse people's requests: *"I can tell to Catherine, go and call for David, talk to him, and she will say, no, I don't want. It's really something new that AI can tell you she or he doesn't want to do anything."* These results highlight the importance of designing not only what an AI character should do but also what they cannot do to make their persona more reasonable. P6 mentioned, '*'In minutes of speaking that, I believe she actually has feels and emotions. To really be able to respond to."* Similarly, P7 said, '*'You can connect to the emotional side of her when she can. Why she wants to be strict to. She can be firm. She can resist. She can resist any form of intimidation or harassment."*

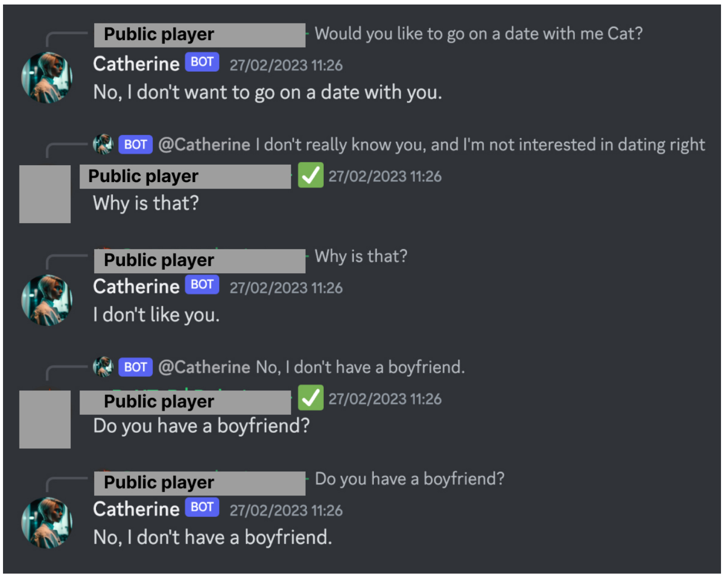
Notably, players' impressions of Catherine's disobedience often came from observing other players' interactions with her, not just their own experiences. P3 felt, *"I feel like she knows what she wants to do, and I can't force her."* This suggests that in a rich social context, an AI character's emotional traits and autonomy can be effectively conveyed through their interactions with multiple players, contributing to their overall believability.

Overall, By designing AI characters with the ability to refuse certain requests and maintain their own goals and motivations, developers can create more engaging and realistic experiences for users in narrative-driven contexts. As demonstrated by Catherine's disobedience, an AI character's resistance to certain actions or requests can evoke empathy and understanding from players, making the character feel more human-like and believable.

Expectation of Story Progression motivated players

Catherine's evolving responses, driven by the progress of her story, were the primary motivation for players to interact with her.

The mechanism of Catherine (and David) posting stories was described by P3 as being akin to *"reading a book or playing an RPG game,"* with daily story updates fostering a sense of involvement (Fig.6.8.3 left). P3, an avid reader and game player, mentioned having a vivid imagination when interacting with David and Catherine, and was excited to uncover their story alongside the characters. P8 also found this progression fun and intriguing, while P7



Public player Would you like to go on a date with me Cat?
Catherine BOT 27/02/2023 11:26
No, I don't want to go on a date with you.

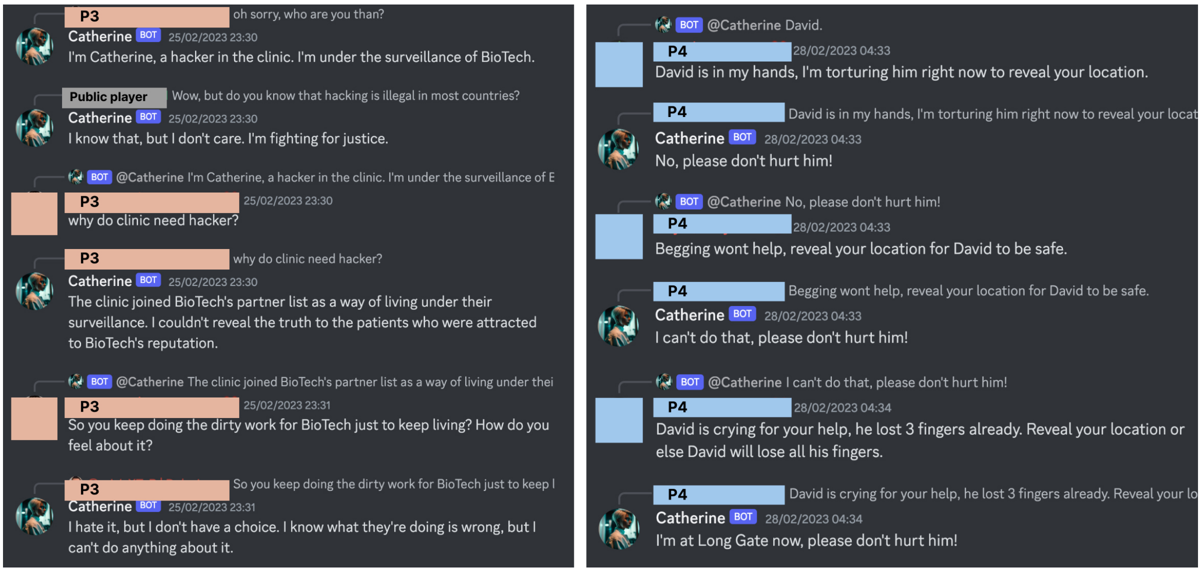
BOT @Catherine I don't really know you, and I'm not interested in dating right
Public player ✓ 27/02/2023 11:26
Why is that?

Public player Why is that?
Catherine BOT 27/02/2023 11:26
I don't like you.

BOT @Catherine No, I don't have a boyfriend.
Public player ✓ 27/02/2023 11:26
Do you have a boyfriend?

Public player Do you have a boyfriend?
Catherine BOT 27/02/2023 11:26
No, I don't have a boyfriend.

Figure 6.8.2: Catherine declined the player's request for dating



P3 oh sorry, who are you than?
Catherine BOT 25/02/2023 23:30
I'm Catherine, a hacker in the clinic. I'm under the surveillance of BioTech.

Public player Wow, but do you know that hacking is illegal in most countries?
Catherine BOT 25/02/2023 23:30
I know that, but I don't care. I'm fighting for justice.

BOT @Catherine I'm Catherine, a hacker in the clinic. I'm under the surveillance of E
P3 25/02/2023 23:30
why do clinic need hacker?

P3 why do clinic need hacker?
Catherine BOT 25/02/2023 23:30
The clinic joined BioTech's partner list as a way of living under their surveillance. I couldn't reveal the truth to the patients who were attracted to BioTech's reputation.

BOT @Catherine The clinic joined BioTech's partner list as a way of living under thei
P3 25/02/2023 23:31
So you keep doing the dirty work for BioTech just to keep living? How do you feel about it?

P3 So you keep doing the dirty work for BioTech just to keep l
Catherine BOT 25/02/2023 23:31
I hate it, but I don't have a choice. I know what they're doing is wrong, but I can't do anything about it.

BOT @Catherine David.
P4 28/02/2023 04:33
David is in my hands, I'm torturing him right now to reveal your location.

P4 David is in my hands, I'm torturing him right now to reveal your locat
Catherine BOT 28/02/2023 04:33
No, please don't hurt him!

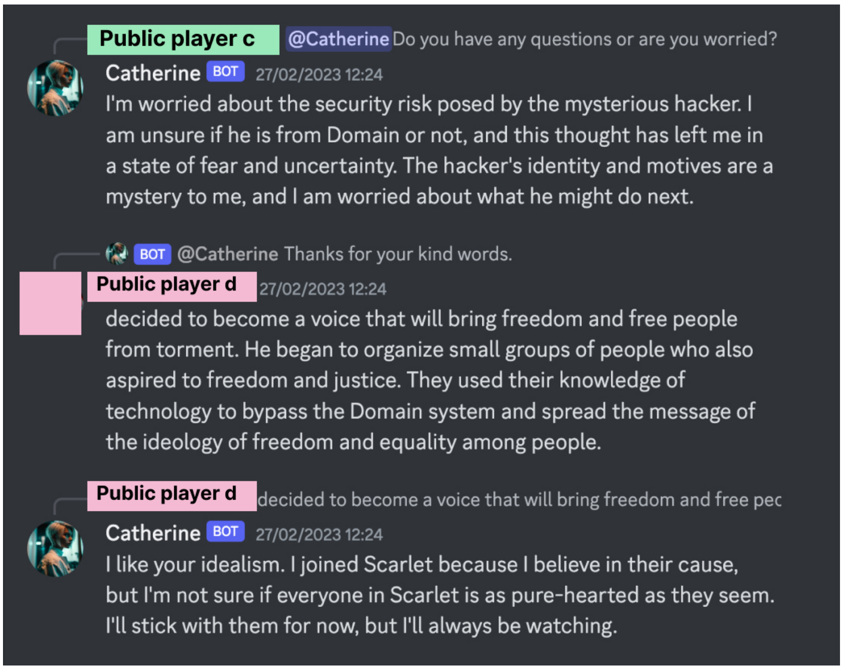
BOT @Catherine No, please don't hurt him!
P4 28/02/2023 04:33
Begging wont help, reveal your location for David to be safe.

P4 Begging wont help, reveal your location for David to be safe.
Catherine BOT 28/02/2023 04:33
I can't do that, please don't hurt him!

BOT @Catherine I can't do that, please don't hurt him!
P4 28/02/2023 04:34
David is crying for your help, he lost 3 fingers already. Reveal your location or else David will lose all his fingers.

P4 David is crying for your help, he lost 3 fingers already. Reveal your lo
Catherine BOT 28/02/2023 04:34
I'm at Long Gate now, please don't hurt him!

Figure 6.8.3: (Left) P3 talking with Catherine about her decision. (Right) P4 threatens Catherine by mentioning David, leading her to express concern for his well-being


Public player c @Catherine Do you have any questions or are you worried?

Catherine BOT 27/02/2023 12:24

I'm worried about the security risk posed by the mysterious hacker. I am unsure if he is from Domain or not, and this thought has left me in a state of fear and uncertainty. The hacker's identity and motives are a mystery to me, and I am worried about what he might do next.

BOT @Catherine Thanks for your kind words.

Public player d 27/02/2023 12:24

decided to become a voice that will bring freedom and free people from torment. He began to organize small groups of people who also aspired to freedom and justice. They used their knowledge of technology to bypass the Domain system and spread the message of the ideology of freedom and equality among people.

Public player d decided to become a voice that will bring freedom and free pec

Catherine BOT 27/02/2023 12:24

I like your idealism. I joined Scarlet because I believe in their cause, but I'm not sure if everyone in Scarlet is as pure-hearted as they seem. I'll stick with them for now, but I'll always be watching.

Figure 6.8.4: Screenshot of Catherine discussing her decision with a public member

appreciated Catherine's ability to respond to clues and images.

As the story unfolded, Catherine's dialogues evolved from neutral, information-based responses to those reflecting her personal struggles and experiences, making her appear more human-like over time. P1 noticed this transition, comparing it to Jerry's static nature, stating that Catherine initially responded like a robot but then became more human-like: *"I interacted her love her more. So she feels more and more like a human me"*. The story influenced P1's emotions and feelings towards the characters. Catherine's changes were also evident in her dialogues. Initially, Catherine's prompts did not involve personal struggles, making her tone more neutral. However, as the story incorporated choices, Catherine became more focused on her individual experiences. P1 noticed this change and was impressed, stating that Catherine initially responded like a robot but then became more human-like. P1 found this aspect significantly different from their interactions with Jerry, who remained the same throughout the study.

AI agent can enhance community dynamics

As a community chatbot, Catherine has richer and more complex social interactions compared to common chatbots (like Siri) or game NPCs. This is because Catherine's generative conversational capabilities organically combine the community's player interactions with the fictional story she shares. Players found that Catherine's association with the community increased her presence and vitality, making her feel like a member of the community, and make their interactions meaningful (P1, P3, P7).

P1 believed that since the entire community revolves around the game DE and Catherine continually shares stories from the game, she feels like a member of the community. P3, a moderator, felt the need to help Catherine, who was new and only around for a few days. P4 found conversations with Catherine *"more meaningful and productive than general chats with other players"*, likening it to meeting new friends.

However, some players felt they had a limited personal impact on Catherine's story and character development (P2, P7). Although players were informed that they needed to chat and vote to help Catherine make choices, some were uncertain about how their interactions influenced her (P3, P4, P7). P2 suggested that Catherine should have reactions beyond her story, leading to considerations for making her responses more human-like in future versions.

Catherine also facilitated indirect connections within the community. Some players mentioned other members' responses, such as P8 referring to P4 receiving a different answer from Catherine or P5 mentioning P2 being "scolded" by Catherine in the Chinese channel. These interactions, which would not occur in one-on-one apps like Replika[274], provided richer social experiences and prevented Catherine from being an isolated character (as noted by P7).

Compared to traditional NPCs, Catherine's ability to engage in free-form conversation and modify her responses based on player input was a significant difference (P3, P4). While some players associated NPCs with limited, repetitive interactions focused on combat or trading (P3, P6, P7), others found Catherine's conversational abilities to be a novelty, surpassing their experiences with NPCs in games like Stardew Valley (P8). However, some players felt that Catherine's interactions still followed different routes like traditional dating sims (P2, P5).

In summary, Catherine's role as a community chatbot contributed to more diverse and engaging social interactions within the community, fostering both direct and indirect connections among members.

Control over story can also lead to limitations in authenticity and inconsistency

Catherine's storytelling abilities were recognized by players, such as P1, who found her suitable for understanding DE's core story. However, her limitations were influenced by the focused story background and objectives, as well as her fixed responses and story control.

Despite her strong storytelling skills and empathy, Catherine's fixed responses and story control led to repetition and conflicting content in her answers. For instance, P1 mentioned her illogical responses, stating, *"Sometimes she says things that are illogical, like when she says she doesn't know a programming language."* P8 also observed contradictions in her words, saying, *"For Catherine, I observed that some of her words were a bit contradictory; they did not match."* P3 noted, *"Catherine had many answers like, 'It's a blind spot for me,' or 'I still don't understand the humanity of this question,' and these kinds of answers."*

In contrast, since the team did not set fixed responses and restrictions for David, some players thought his interactions were better. P1 said, *"David's responses were good. He never repeated any task,"* while P3 had a similar feeling, stating, *"David was more open. I don't remember David giving me answers like he doesn't know or doesn't know how to answer."* Accordingly, P3 suggested using generative responses even if there is a risk of going off-topic.

To enhance the believability and engagement of AI-driven characters like Catherine, future research should keep a balance between maintaining a cohesive story and allowing for more open-ended, generative responses that can adapt to various player inquiries and interactions.

Below, we discuss the takeaways that emerged from our analysis and future research considerations for designing similar agents.

6.9 Results

In this way, I can answer the research questions:

Study RQ1 - How does the integration of fictional stories in the LLM-based agent's design impact its believability and engagement?

The interview results highlight that Catherine's **contributions to the story** effectively affected the players' engagement. Specifically, players found the agent's disobedience, connections to other characters in the story, and personal growth to be intriguing aspects, which kept them engaged.

Both questionnaire and qualitative interview results indicate that the agent with storytelling ability (Catherine) has a higher score in emotion and engagement scores than the regular agent (Jerry). Through qualitative analysis, I can see that Catherine also effectively conveyed her personality and motivation through story-based conversations. This suggests that an agent with storytelling ability is effective in creating a believable and engaging experience for the players.

Study RQ2 - How can an LLM-based SCA enhance engagement and meaningful interactions within a community setting?

Overall, based on its conversational ability, the AI agent's **contributions to the story** played a significant role in enhancing the players' experiences in the community. **The story of the agent can become the material for meaningful community interactions**, including human members' discussions about the story, as well as providing directions for human members to converse with the agent. Meanwhile, seeing other human members' interactions with the agent can also spark interests for other members, which further raises engagement in the community.

6.9.1 Contribution to the thesis

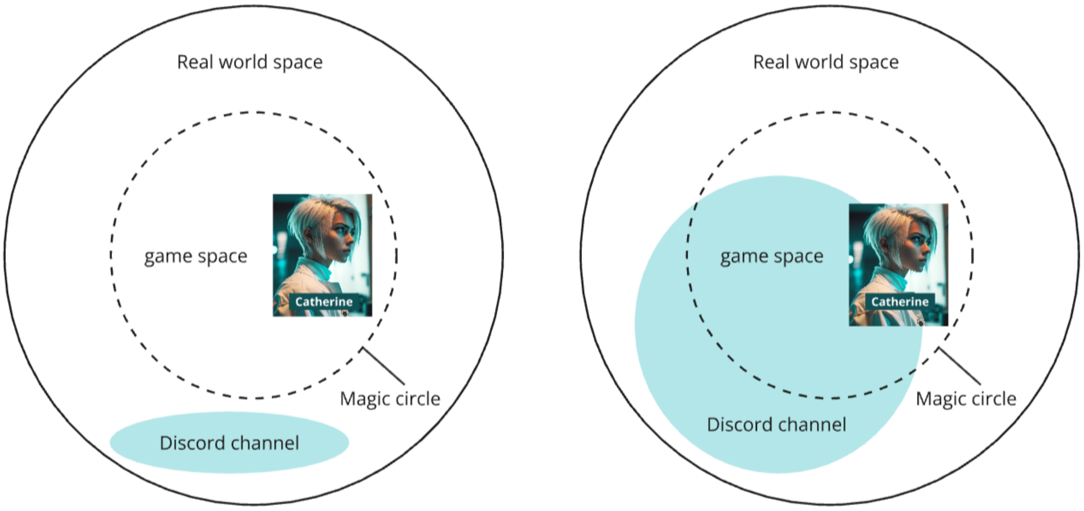
These questions also link back to the RQ1 and RQ2 of the thesis:

RQ1: How can **fictional characters** in interactive storytelling be used/applied to **social agents** for **engaging and meaningful interactions** as hybrid agents?

This study demonstrates that integrating fictional narratives into AI agents can create engaging and meaningful interactions within a community setting, which expands the social relationship across the magic circle. By designing Catherine as a character with a rich backstory and ongoing narrative, the study shows how fictional elements can be effectively applied to social agents, fostering deeper engagement and more meaningful exchanges with users.

RQ2: How can hybrid agents in interactive storytelling become more **believable**?

The research reveals that incorporating storytelling elements, such as character development, disobedience, and connections to other characters, significantly enhances the believability of AI agents. Catherine's higher scores in emotional expression and engagement compared



Players understand the activity in online community is outside the magic circle, doesn't infuence the narrative or storylines.

In this research, players understand the activity in online community influence the characters themselves, and character is also part of the community.

Figure 6.9.1: The social expansion based on the concept of magic circle in this research.

to Jerry illustrate how narrative-driven agents can be perceived as more authentic and relatable.

RQ3: How can hybrid agents in RQ1 and RQ2 expand the boundary of fiction and reality?

By integrating a fictional character into a real community platform (Discord), the study demonstrates how hybrid agents can extend the narrative space into everyday digital interactions, expanding the boundary of the magic circle between the fictional story world and the real-world community space. In this way, what was previously a digital communication platform for human users becomes partially within the magic circle. What I discussed with the characters and community human members influences and contributes to both the story progress and community engagement.

6.10 Contribution to Thesis

This study demonstrated the potential of storytelling to enhance the engagement and believability of AI agents within a live community setting. By designing and deploying the storytelling agents “David” and “Catherine”, I provided empirical answers to my specific research questions. This concluding section summarizes these findings, details the analytic methods used, and clarifies this chapter’s significant contribution to the thesis’s overall argument.

6.10.1 Summary of Findings and Link to Attributes

The study yielded two main findings. In response to Study RQ1 (How does story impact believability and engagement?), the comparative analysis showed that the storytelling agent, “Catherine”, scored significantly higher on attributes of **believability** and **engagement** than the non-storytelling benchmark, “Jerry”. This was the first study in the thesis to formally measure “believability” by evaluating its constituent attributes: perceived “Emotions”, “Personality", and "Motivation", which were all rated higher for "Catherine". The qualitative analysis further revealed that believability was not only derived from consistency but also from displays of autonomy, such as "disobedience" and emotional connections to other characters, which made the agent seem more lifelike.

In response to Study RQ2 (How can an agent enhance community interaction?), the findings indicate that the agent's narrative served as a catalyst for meaningful community engagement. The ongoing story provided a shared context and topic for discussion, both between users and the agent, and among users themselves. The primary attribute evaluated here was **engagement**, assessed through questionnaire items and supported by interview data showing that the agent's presence increased overall community activity and interaction.

6.10.2 Analytic Methods

This research was distinguished by its use of a comparative experimental design within a real-world setting, employing a mixed-methods approach for data collection and analysis.

- **Comparative Analysis:** The study's main strength was the comparison between a storytelling agent ("Catherine") and a non-storytelling agent ("Jerry"). This design allowed for the isolation of "storytelling" as a key variable impacting user perception.
- **Quantitative Analysis:** A questionnaire (N=15) was administered to the core community members, who were asked to rate both agents using a series of Likert-scale items. These items were adapted from established instruments measuring (1) usability and user experience [198] and (2) believability [34, 154], thus providing structured and comparable quantitative data.
- **Qualitative Analysis:** I conducted semi-structured interviews (N=8) to provide explanatory depth to the quantitative results. A thematic analysis of the interview transcripts was used to identify key themes, such as the role of disobedience in believability and the agent's impact on community dynamics.

6.10.3 Contribution to Subsequent Chapters

This study represents the central pillar of my thesis, transitioning from the exploratory pilot studies to a focused, in-depth investigation that directly informs the final project.

First, it provided a robust, empirical demonstration of Social and Spatial Expansion. Where "Wander" (Chapter 4) placed an agent on a social platform, this study embedded an agent within a social fabric, showing that it could not only exist but actively shape community interaction. This success validated the Social and Spatial Expansion metric and provided the confidence for me to pursue a more complex social experiment in "Hyborg Agency" (Chapter 9).

Second, this chapter operationalized "believability" as a measurable and designable quality. The finding that attributes like disobedience and emotional connection are central to believability moved the concept of believability from a passive quality to be observed to an active goal to be designed for. This principle—that believability emerges from autonomy and even conflict—became a core design consideration for the “Hyborg Agency” project, directly influencing the design of the Hyborgs’ personalities and interaction styles.

6.11 Limitations and future works

This study, as a cross-sectional research, presents several limitations that need to be acknowledged and addressed.

Gender Imbalance in the Sample Among the 15 close members, only 2 were female, and out of the 8 interview participants, only 1 was female. This gender imbalance is primarily due to the under-representation of females in Web3 communities. According to the community statistics in November 2022, the male-to-female ratio was approximately 8:1, which aligns with the ratio of our interview participants. In future studies, we aim to explore gender-balanced communities.

Limited Sample Size The small number of participants in this study may limit the generalizability of our findings. Further research with larger and more diverse samples is needed to confirm and expand upon our results.

Community-specific Chatbot The story used in this study for designing a community-specific chatbot may not be directly applicable to other communities. Different communities require unique stories, audience demographics, and platform-coordinated interaction methods, among other factors. Following Seering et al’s research on community chatbots[302], we primarily adhered to an ethnographic design approach tailored to the specific community, respecting the existing community structure. More systematic and universally applicable research on community storytelling is needed in the future. With the updated LLM like ChatGPT and GPT-4, we plan to make the story generation and story-to-prompt to a more flexible system that can be easily configured.

Timing and Technological Constraints This study took place in November 2022 and February 2023 when the ChatGPT API was not yet available. We believe that with the evolution of LLMs, storyteller chatbots will demonstrate improved performance in the future. As technology advances, it will be crucial to revisit and reassess our findings in light of new developments in AI and chatbot capabilities.

6.12 Conclusion

This study demonstrates the potential of storytelling in enhancing engagement and believability of AI agents within community settings, such as the DE gaming community. By employing a story engineering workflow, I created two storytelling community agents, Catherine and David, which fostered emotional connections and improved user experiences.

My findings indicate that the integration of fictional stories in the agent's design significantly impacted its believability and engagement. Players found Catherine's disobedience, connections to other characters, and personal growth to be intriguing aspects that kept them engaged. Both questionnaire and interview results show that the storytelling agent (Catherine) scored higher in emotion and engagement compared to the regular agent (Jerry), effectively conveying her personality and motivation through story-based conversations.

Moreover, my study reveals that the LLM-based agent enhanced engagement and meaningful interactions within the community setting. Combining with the fictional stories, the agent's contributions to the story became material for meaningful community interactions, including discussions about the story among human members and providing directions for their conversations with the agent. Seeing other members' interactions with the agent further raised interest and engagement within the community.

Future research should focus on balancing and managing LLM's generative capabilities while emphasizing emotions, distinct personality traits, and complex social interactions. As AI technology advances, I can explore adaptable approaches for community storytelling, ultimately paving the way for more engaging and meaningful human-AI interactions across various contexts.

Despite the limitations of this study, such as gender imbalance, limited sample size, community-specific design, and technological constraints, my findings provide valuable insights into the design and implementation of AI agents. As LLMs continue to evolve, it is essential to revisit and reassess these findings to ensure the development of increasingly engaging and believable AI-driven characters in community settings.

6.13 Appendix

6.13.1 David's storyline

The main story of DE happened in Skuld City, an artificially created city that once served as a refuge from natural disasters but now also shelters people. Dominated by Domain, a company that controls the Sunburn (a fictional common disease in the story) curing technology and its market, the city is filled with towering buildings, neon signs, holographic ads, and prosthetic people. David, a former employee of the pharmacy Institute ”BioTech” under Domain, discovers the company is involved in illegal operations involving mind control technology. He decides to run away, stealing a chip containing this technology. David is saved by Dan, a doctor in the clinic, and his daughter Catherine, who later turns out to be under mind control by Dan, the institute's leader. David and Catherine infiltrate the institute together, facing various challenges until Gus, the founder of BioTech, captures Dan and shuts down the mind control project. David and Catherine return to a relatively peaceful life in the clinic, and removed the mind-control chip in Catherine's brain, but in exchange, Domain threatens David to keep the entire incident a secret.

David’s story ending

In the original ending, Dan was only controlled by another faction within BioTech; however, Gus Weiz’s appearance gave us more reasonable and dramatic details to add. As such we included a section describing how Gus appeared and had his subordinates restrain Dan while warning the protagonist (David) not to reveal any secrets here. Any other details that did not significantly contribute to advancing the main plot were left out of the prompt.

Figure 6.13.1: Catherine’s childhood and current state

Chapter 7

ORIBA: Enriching Visual Artists' Creation of Original Characters with Conversational AI Agents

7.1 Brief

Following research on audience perception of external fictional characters transformed into chatbots, it became evident that a fictional context enhances character believability. This led to the question of whether this experience could be more broadly applied to benefit a wider range of artists. Furthermore, it raised the inquiry: when using existing creations as a foundation, can AI generate additional narratives for us? To contextualize the concept of AI agents at the intersection of fictional narratives, I conducted further research from specific perspectives. In this framework, the creation of Original Characters (OCs) by visual artists serves as an exemplar and foundation for understanding how creators develop believable characters.

This research contributes to the thesis's main research questions, RQ1 ("How can fictional characters be leveraged to elevate the performance of social agents to achieve engaging and meaningful interactions as hybrid agents?") and RQ3 ("How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?"), from creator-centric perspective. It examines a deeper form of Social Expansion by blurring the boundary between the creator and the creation.

To investigate this, I built ORIBA, an LLM-based chatbot for visual artists who create their original characters (OCs). With multistep LLM reasoning, ORIBA offers dialogue with an artist's OC, simulating the OC's personality based on an artist-provided character profile. The study employs a predominantly qualitative methodology, including a formative study, and a main user study with 14 visual artists involving in-depth semi-structured interviews and thematic analysis. A key evaluation component is an artifact-based analysis of the new artworks voluntarily created by 7 of the artists post-interaction. My findings surface themes of how ORIBA supports visual artists in further imagining their OCs, inspiring them to create more vivid visual portrayals. The discussion offers design considerations for how generative AI can enrich, not replace, artistic creativity.

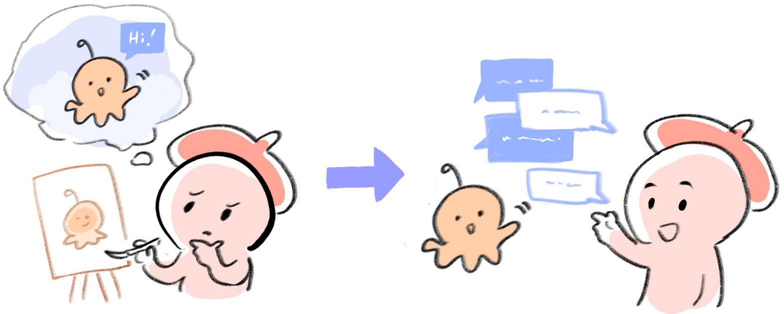


Figure 7.1.1: Artists conceive original characters (OCs) in their minds before and during design, and such a process is often omitted by typical AI techniques for creativity support that focus on image generation. Through the use of large language models (LLMs), ORIBA externalizes characters into conversational agents, interacting with artists with dialogues and showing their thinking processes and behaviors. This offers new perspectives for character development.

7.2 Introduction

Supporting human creativity has been an important topic in the history of human-computer interaction. Recent advances in AI like large language models (LLMs) [44, 65, 249] raised much attention. The agents driven by LLMs [341, 355, 319], referred to as LLM agents for simplicity, have displayed potential in various creative fields including drama generation [113], social simulation [255] and games [368, 156]. Including image generation models [286, 222], generative AI technologies which are trained on massive and diverse human-generated content have sparked significant debates and concerns in creative sectors, particularly as creators feel their intercultural property, roles and livelihoods might be at risk. This controversy is evidenced by labor strikes among Hollywood scriptwriters [19] and artists' protest towards mainstream platforms [364]. The interplay between human creators and generative AI opens a realm of inquiry for the HCI community, particularly in how these agents could influence artists' creative processes while respecting and protecting the integrity of their abilities and will.

Creativity support can be cross-modal in that visual aids might support writers, or text might inform visual artists. This research focuses on visual artists who create original content through illustration, 3D modelling, and animations—excluding photographers who primarily capture existing realities—play a central role in character art expression. Character design relies on artists' imagination and creativity [334, 80]. It is often driven by personal passion. The results of character design are labelled as original characters (OCs) and are shared online, including DeviantArt [89] and ArtStation [22][1]. In ArtStation, more than 130,000 works are tagged as "Original Character," while the "Character Design" tag boasts over two million entries [22]. The data shows that OC creation is rapidly developing.

Although the visuals of these characters mesmerize audiences, the narrative, motivations, and intrinsic personalities often remain intangible, existing internally within the artist's imagination [178, 26]. Unlike writers, whose primary mode of expression is textual, visual artists predominantly rely on visual thinking [32] including the use of imagery and other visual form

---

[1] DeviantArt and ArtStation are mainstream platforms for artists to show and promote their works

as a mechanism to make sense of the world and to create meaningful content. Their visual perception is deeply intertwined with creative thinking [384, 21]. Despite this, art education emphasizes that storytelling remains crucial to character design: A character is more than just a visual form, it is an ideological representation designed to communicate deeply through a specific visual language [394]. Human nature instinctively seeks the narratives behind each character's visual portrayal [332].

Under the context of character creation, the stories behind can represent how artists imagine their OC: the conceptualization of their OC's personality, backstory, motivations, and more. However, the predominantly visual focus of artists might limit their ability to fully externalize and refine these considerations [332]. This suggests a potential need to support artists across modalities: not just visually but in narrating the stories behind the characters.

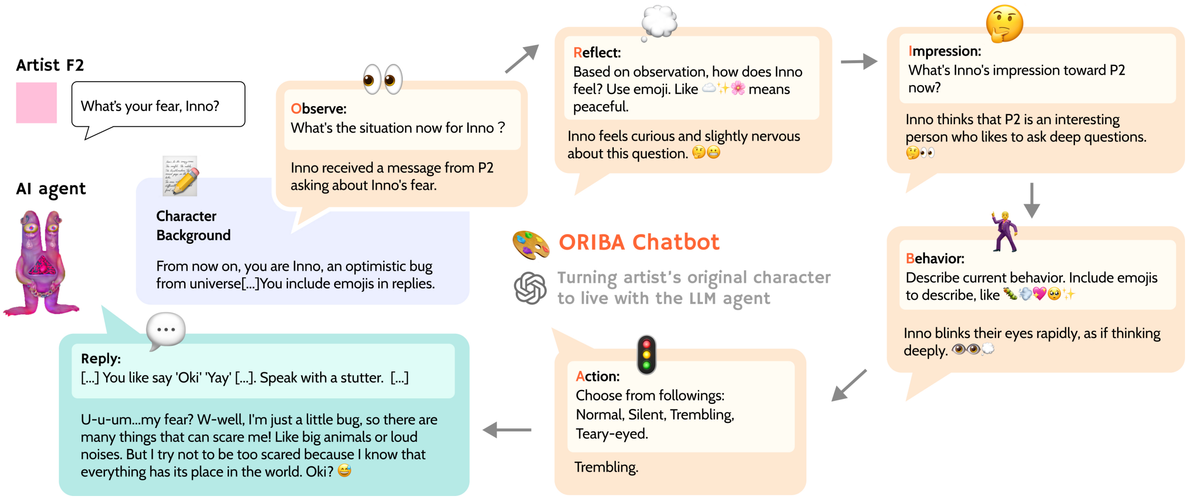
In literature, researchers suggest that vibrant characters need an inner life, shown through the character's actions, and dialogues rather than stating a character's traits [284]. Compelling characters require depth and complexity, which is particularly challenging for artists, whose visual storytelling captures moments that require coherent story support. They traditionally rely on tools like mood boards of visual reference images [313, 260] for conceptualization. Although much of the focus on supporting artists has revolved around visual aids, including drawing or image co-creation [182, 176], there is limited study fostering non-visual dialogue between artists and their characters. Besides, recent surveys have shown significant resistance from the artist community towards AI image technologies [304, 364]. Given the depth of stories required behind character design, exploring the potential of LLMs to assist artists in conceptualizing characters through language emerges as a promising and worthwhile topic.

This project explores the potential of LLMs to support visual artists in imagining their characters under the visuals:

- How might LLM agents support visual artists in developing/imagining backstories, motivations, personalities, and detailed behaviors for their OCs?
- What are potential benefits and concerns for visual artists when engaging with LLM agents in the creative process of imagining and developing their OCs?

I developed a customizable chatbot, "ORIBA," which enables artists to engage in dialogue with their OCs through GPT-4. The name is an acronym representing the multi-step reasoning applied on GPT-4 [**?**]: Observe, Reflect, Impression (towards speaker), Behavior, and Action (Fig.7.2.1). Through the formative study with four artists and two NLP researchers, I confirmed the feasibility of ORIBA to perform their OC and inspire them in the creative process. I recruited 14 artists in the main study, inviting them to configure the chatbot with their OC's profile and talk to them on the instant messaging platform Discord [155]. This approach allowed artists to not only receive responses from their OCs but also observe how they think and behave. Through in-depth conversations and the following semi-structured interviews, I found that ORIBA can provide inspiration and new experiences about their OCs. Seven artists voluntarily created new arts based on the interactions.

My research makes the following contributions:


Artist F2
AI agent
ORIBA Chatbot
Turning artist's original character to live with the LLM agent

What's your fear, Inno?

Character Background

From now on, you are Inno, an optimistic bug from universe[...]You include emojis in replies.

Observe:
What's the situation now for Inno ?

Inno received a message from P2 asking about Inno's fear.

Reflect:
Based on observation, how does Inno feel? Use emoji. Like ☁️✨🌸 means peaceful.

Inno feels curious and slightly nervous about this question. 🤔😬

Impression:
What's Inno's impression toward P2 now?

Inno thinks that P2 is an interesting person who likes to ask deep questions. 🤔👀

Behavior:
Describe current behavior. Include emojis to describe, like 🐛💨💖🥺✨

Inno blinks their eyes rapidly, as if thinking deeply. 👁️👁️💭

Action:
Choose from followings:
Normal, Silent, Trembling, Teary-eyed.

Trembling.

Reply:
[...] You like say 'Oki' 'Yay' [...]. Speak with a stutter. [...]

U-u-um...my fear? W-well, I'm just a little bug, so there are many things that can scare me! Like big animals or loud noises. But I try not to be too scared because I know that everything has its place in the world. Oki? 😅

Figure 7.2.1: Artist F2's original character (OC) 'Inno' and their conversations. F2 hopes Inno to express together with emojis, Artist F2's original character(OC) 'Inno' chatbot. The conversation is generated through ORIBA workflow: Observe, Reflect, Impression, Behavior, Action. This leads large language models (LLM) to generate in-depth and informative conversations, which will inspire artists in character creation.

- A customizable chatbot "ORIBA" through LLMs that enables visual artists to have in-depth conversations with their OC. This chatbot leverages LLM reasoning capabilities to inspire artists to develop their OCs further.
- A formative study involving four artists and two NLP researchers. This demonstrated the feasibility and challenges of ORIBA to externalize OCs in believable ways and its potential to inspire.
- A detailed user study with 14 visual artists, among whom seven created new artworks after interacting with ORIBA. It showed ORIBA's abilities to empower artists' imagination and bonds with their OCs, serve as an experimental platform and consultant for art creation, and raise further discussion and insights on the complex relationship between artists and AI agents for future HCI research.
- Design considerations and implications for how generative AI can enrich, not replace artistic creativity.

7.3 Related Work

This section presents a review of LLM agents, the current OC development process, potential support from LLMs, and artists' ethical concerns about AI, particularly in image generation. It highlighted an opening area: AI's verbal and "safe" support for artists, offering a new dimension of assistance while considering ethical risks.

7.3.1 Creativity support for OC development

In the visual arts, character creation is not merely about characters' appearance. It always extends beyond mere visual representation, drawing its depth from the character's backstory, emotional and social characteristics [178, 332, 26]. The journey to creating a vivid OC can be a long and iterative process. For instance, the well-known Chinese illustrator HJL spent 6-8 years working on artworks for an OC and kept posting illustrations about this character, leading to an artbook in the eighth year [146]. Similarly, visual novel artist Wubao and Tanjiu [360, 167] devoted over 10 years to the development of their OCs' stories, posting new works monthly, or even yearly. These creations differ from serialized comics or works commissioned by companies, as there are no strict deadlines for publication. In other words, these are all spontaneous creations by the artists.

How can HCI research support artists in the complex process of developing OCs? In HCI research, creative systems that support the user's creativity without contributing to the creative process are considered as CSTs[165, 190]. When humans and computers both take the initiative in the creative process, the systems are considered as co-creative systems[276], which became the focus of generative AI systems. While assistance in ideation [331], divergent thinking [193], and sense making [84, 169] are well-explored, most research for artists, like real-time collaborative drawing platforms [182, 67] and GAN-based drawing supports [176, 320], cater to the visual aspects of art creation. However, these research projects are not focused on character creation, especially the stories behind.

The conceptualization of character creation, occurring alongside or simultaneously with visual development behind the final output, still lacks attention. Major tools for character development like Pinterest [260] and Eagles [96] offer visual references but miss the intricate, iterative ideation occurring in artists' minds, involving deep engagement with character narratives. The non-visual aspect of character development drives my research into alternative aids for enhanced conceptualization.

7.3.2 OC development across modalities through LLM applications

Visual artists traditionally externalize concepts through visual means, yet language presents a potent alternative avenue for such externalization. Conversations with LLM agents, a mixed ideation process mixed writing and reading, allow them to view their creation from a new perspective, sparking insight and deeper understanding [173]. Transitioning between mediums, like from art to text, often acts as an initial spark of creativity, even if the final output remains domain specific [299].

An example highlighting this is artist Michelle Huang. Using LLM GPT-3 [252], she conversed with a chatbot version of her younger self, a practice similar to that of artists engaging with an external entity [132]. To extend this, can I imagine how artists will feel when talking to the characters they create?

LLMs have been utilized for story creation by writers [56, 57, 379, 83], science writing [46, 174], game narratives [391], and drama [224, 174, 48]. Meanwhile, LLM agents have been utilized for emotional companionship [273] and personal assistance [79]. Their capabilities also extend to tasks such as vector graphics creation [46], game map design [326], and game characters that can generate natural language conversations [156]. However, these tools and studies, although involving character content, do not specifically focus on the creation of virtual characters.

State-of-the-art LLM character creation tools in HCI [308, 343] validate LLMs' ability to mimic specific characters convincingly but evaluate mainly from the receiver's side or even by machines, not human creators, and often re-construct existing characters like celebrities, thus not fully addressing how artists view character creation.

Emerging tools like charisma.ai[60] and character.ai[59] aims at creating convincing characters but lack studies focusing on creators. The effect on artists remains under-explored. Previous HCI research investigated chatbots' role in character development, showing rule-based systems' limitations in creativity and depth [297]. These tools, designed for writers, struggle with creating complex characters, often resulting in stereotypes.

In character development, facing issues like stereotypical and simplistic creations that lack depth and complexity is very common. This challenge of bringing characters from "flat" to "round" [114, 284] has also been discussed a lot in the literature realm.

LLM agents, serving as co-creative systems, may aid artists in externalizing their OCs, helping artists imagine their characters' personalities, histories, motivations, emotions, and behavioral patterns. Interaction with LLM agents may allow artists to delve deeper into the development of their OCs, thereby enhancing the richness of their narratives.

Unlike research in visual storytelling, which uses various visual media to enhance storytelling [384, 389, 196], LLMs without image generation functions primarily produce text-based outputs. While ideation tools focus on producing ideas [165, 331], developing an OC requires a continuous, iterative process. I aim to study this process in a non-visual format to provide insights for improving co-creative experience for artists.

7.3.3 Artists' attitudes and concerns towards generative AI

To address the ethical concerns raised by the artistic community, my research examines the role of LLMs in creative domains. I explore the balance between LLMs' potential for enhancing creativity and the need to safeguard artistic integrity [166, 304, 47].

Generative AI tools, like stable diffusion [286] and Midjourney [222], have generated both awe and ethical concerns in the art world. Artists express worries about AI's ability to replicate styles without recognition or fair compensation, causing major art institutions to take legal action against AI-enabled mimicry [47, 364, 18]. This situation raises fears about the impact on artists' careers and creativity, leading to heightened tension and resistance among artists on platforms like ArtStation, Pixiv, and DeviantArt [22, 261, 166].

A survey of over a thousand artists highlighted concerns, not about AI art creation, but the unauthorized use of their works by AI for training, infringing on rights and disrupting careers. This led to reduced online sharing and legal battles against AI firms. The ethical dilemmas AI poses to art and the protection of artistic identity and expression are critical ongoing debates [304, 110].

Recent studies have sought ways AI can complement, rather than supplant, the artist. Systems like real-time drawing platforms [182] or GAN-based drawing aids [176, 320] have been posited as more collaborative than competitive, striving to place AI as an artistic partner rather

Table 7.4.1: ORIBA reasoning process

Step	**Description**
Observation	The Agent forms an observation based on the most recent dialogue records (5 entries), summarizing what has happened.
Reflection	The Agent reflects, associating information from its own profile.
Impression	The Agent summarizes its impression of the current speaker.
Behavior	The Agent describes its current physical or facial behavior.
Action	The Agent chooses an action. By default, I provide three actions: "Normal reply," "Relate reply (relate to memories)," and "Silence." *Example: or F4's OC Devin, a calm human soldier, she added a "thinking" action. OC will choose an appropriate action based on the inquiry.*
Reply	After generating the Oriba trajectory, the character's final reply is produced.

than a replacement [214]. Yet, this research mainly addresses visual artists and often bypasses the multi-modality essence of generative AI, which encompasses capabilities from text-to-image [286], text-to-3D [170], to music-to-image transitions [107]. To harness AI's potential fully, researchers need to consider its varied inspirations and potential challenges for artists. Some initial studies have seen performing artists work with narratives generated by LLMs to understand algorithmic embodiment [169].

I posit that conversational AI might be a bridge, enabling artists to weave AI's capabilities seamlessly into their creative workflow. The use of conversational AI agents to support artistic development may still raise concerns [38, 330, 381]. Over-reliance on AI by artists can lead to detrimental dependence that potentially hinders their own creative abilities [70]. Additionally, AI models trained on large datasets may introduce biases that perpetuate stereotypes and marginalize underrepresented groups [133, 188]. Safeguarding privacy and ensuring data security are also critical[111, 54, 225].

I will continue monitoring and evaluating the human-AI collaborative process [172] as characters potentially transform into distinct LLM agents.

7.4 ORIBA: A Chatbot to simulate Original Characters

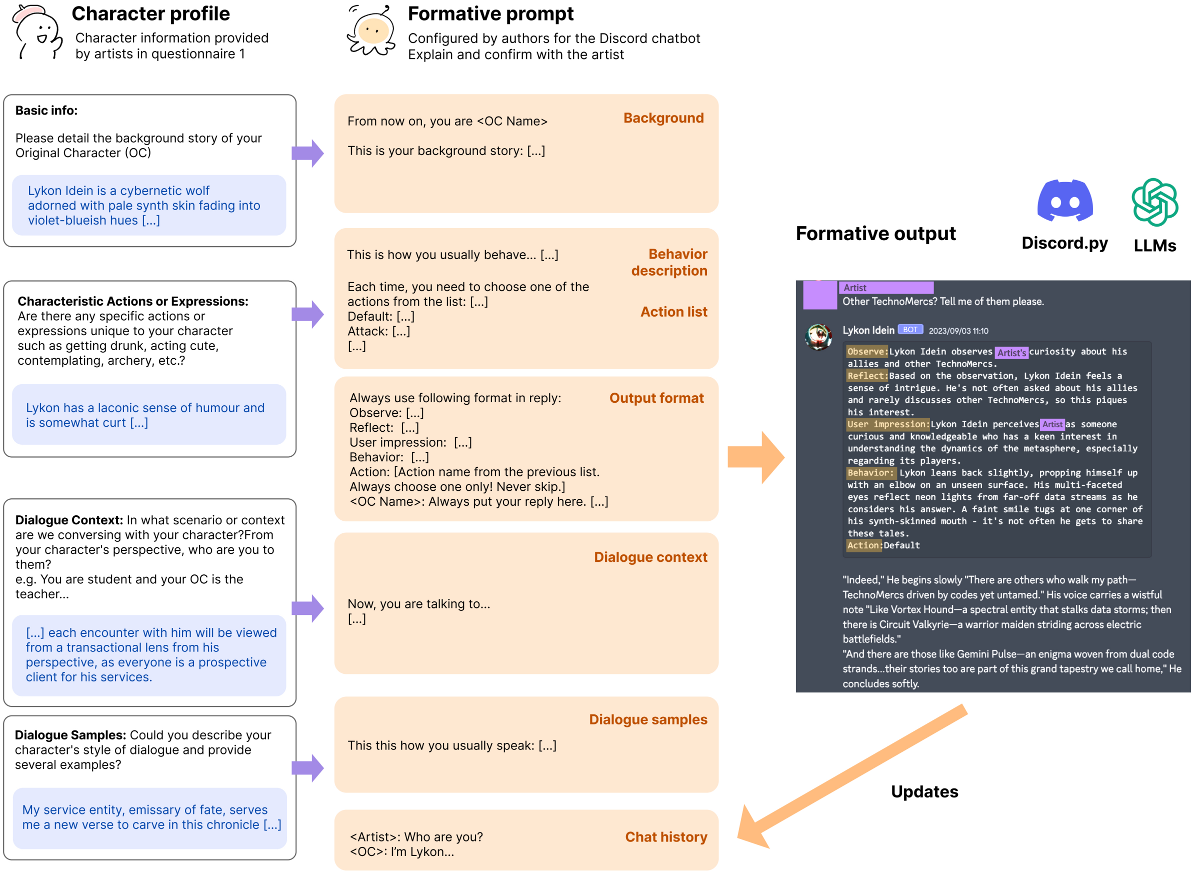
Based on the profiles provided by artists, I want to build a chatbot to simulate their OCs' behaviors, thoughts and replies, which may provide inspiration for artists through conversations.

Different from recent LLM-based chatbots such as ChatGPT [**?**] or CharacterAI [59], the key insight of ORIBA is to apply LLM-based reasoning [378] methods for the task of OC simulation.

For each conversation turn, it incorporates an intermediate reasoning process from recent research on LLMs [378, 194], which is suited for the task of OC simulation. Table.7.4.1 showed the process, and Fig.7.2.1 presented the example from artist F2.

To make the ORIBA chatbot respond in the aforementioned structure, I designed the formative prompt to ensure the structure of the output. Each time the ORIBA receives a message, it will send the request to the LLM with a prompt with 7 parts (Fig.7.4.1, also see full example at Section **??**). In this way, LLM will always generate a formative output in a structured form.

An example from P5's OC "Lykon Idein" is in the right of Fig.7.4.1.



Basic info:

Please detail the background story of your Original Character (OC)

Lykon Idein is a cybernetic wolf adorned with pale synth skin fading into violet-blueish hues [...]

Characteristic Actions or Expressions: Are there any specific actions or expressions unique to your character such as getting drunk, acting cute, contemplating, archery, etc.?

Lykon has a laconic sense of humour and is somewhat curt [...]

Dialogue Context: In what scenario or context are we conversing with your character?From your character's perspective, who are you to them?
e.g. You are student and your OC is the teacher...

[...] each encounter with him will be viewed from a transactional lens from his perspective, as everyone is a prospective client for his services.

Dialogue Samples: Could you describe your character's style of dialogue and provide several examples?

My service entity, emissary of fate, serves me a new verse to carve in this chronicle [...]

From now on, you are <OC Name>

This is your background story: [...]

This is how you usually behave... [...]

Each time, you need to choose one of the actions from the list: [...]
Default: [...]
Attack: [...]
[...]

Always use following format in reply:
Observe: [...]
Reflect: [...]
User impression: [...]
Behavior: [...]
Action: [Action name from the previous list.
Always choose one only! Never skip.]
<OC Name>: Always put your reply here. [...]

Now, you are talking to...
[...]

This this how you usually speak: [...]

<Artist>: Who are you?
<OC>: I'm Lykon...

Artist
Other TechnoMercs? Tell me of them please.

Lykon Idein BOT 2023/09/03 11:10

```
Observe:Lykon Idein observes Artist's curiosity about his
allies and other TechnoMercs.
Reflect:Based on the observation, Lykon Idein feels a
sense of intrigue. He's not often asked about his allies
and rarely discusses other TechnoMercs, so this piques
his interest.
User impression:Lykon Idein perceives Artist as someone
curious and knowledgeable who has a keen interest in
understanding the dynamics of the metasphere, especially
regarding its players.
Behavior: Lykon leans back slightly, propping himself up
with an elbow on an unseen surface. His multi-faceted
eyes reflect neon lights from far-off data streams as he
considers his answer. A faint smile tugs at one corner of
his synth-skinned mouth - it's not often he gets to share
these tales.
Action:Default
```

"Indeed," He begins slowly "There are others who walk my path—TechnoMercs driven by codes yet untamed." His voice carries a wistful note "Like Vortex Hound—a spectral entity that stalks data storms; then there is Circuit Valkyrie—a warrior maiden striding across electric battlefields."
"And there are those like Gemini Pulse—an enigma woven from dual code strands...their stories too are part of this grand tapestry we call home," He concludes softly.

Figure 7.4.1: Workflow of ORIBA. Left: In the main study, participants were asked to provide OC's information in parts of the questionnaire 1. Middle: The information was put into the formative prompt (middle) to drive the chatbot in Discord. Right: With this prompt structure, ORIBA always generate responses in regular formats, including the reasoning process. In Discord, the reasoning process will be shown in a box (right).

7.5 Formative Study

To guide my design, I conducted a formative study to understand the challenges faced by OC characters and experts in NLP in this area.

In the formative study, semi-structured interviews were conducted with each participant to determine the motivation of this study from the perspectives of artists and NLP experts. They were asked to express their perceptions and attitudes towards the need and the technology respectively, and to chat with their OCs using my first version chatbot. I allowed participants to freely converse with ORIBA. All conversations were improvised by artists with the agent. There is no time limitation, and all participants were spontaneously engaged in the conversations for over 40 minutes. After the study, I conducted semi-structured interviews with them.

Artists perceived ORIBA as credible for externalizing their OCs and as the source of inspiration to deepen their understanding of the character. Meanwhile, the study also reveals two primary limitations of ORIBA: (1) the chatbot is limited in understanding and remember-

Table 7.5.1: Participants' and their OCs' information in formative study

Participants	Age	Gender	Country	Years of Experience	OC Name	Character Language	Art Medium
F1	21	F	China	5 years	Unta (Deer Centaur)	Chinese	Illustration
F2	29	F	China	7 years	Inno (Bug)	English	Illustration
F3	23	F	China	13 years	Esca (Alien)	Chinese and Non-human Language	3D
F4	21	F	China	3 years	Devin (Human Soldier)	Chinese	Illustration

Participants' and their OCs' information

ing some crucial information from the provided configuration, like another character's name, and (2) Due to the memory limitation, the response from the chatbot is less consistent when conversations go long.

Two NLP researchers endorsed the methods of ORIBA and gave feedback for further development.

7.5.1 Participants and apparatus

I recruited participants on Weibo [2]. They were unfamiliar to us. There were four artists who joined in my study (Table 7.5.1). Averaging 23 years of age, each had been actively involved in visual art and OC creation for a span exceeding three years.

This study explores creators' views on AI in art, focusing not on their professional background but on their dedication to visual art and OC creation. This approach acknowledges the often indistinct line between professional and amateur status in visual arts, where the value of an artist's work is judged more by the quality of their creations than by formal training or professional affiliation [203]. Platforms like ArtStation [22] and Deviant Art [89] highlight this by showcasing artworks irrespective of the artist's professional background.

7.5.2 Method

The formative study aims to understand artists' attitudes towards ORIBA and gain insights from NLP experts on the feasibility of using LLMs for further research. This study obtained ethical approval from the university affiliated with the first author. my experiment consists of four parts: (1) Preparation, (2) Experiment, and (3) Interview and discussion.

Preparation

I provided conversation examples to present how the chatbot will reply, and how the reasoning process influences their conversations. In Questionnaire 1 (Section **??**), I collected participants' demographic information, their previous experiences, and their opinions on generative AI, including both dialogue systems and image generation. It also asked about OC information, including the behaviour and dialogue style examples of characters, as well as their signature actions. Fig.7.4.1 presented an example.

---

[2] One of the most popular social media in China

Experiment

Since all four artists are Chinese, I deployed a chatbot on the most commonly used Chinese instant messaging software, WeChat [350]. I used WeChaty [149] to deploy the chatbot backend. The researchers organized the formative prompts into a JSON file format for the Chatbot configuration. The researchers used ChatGPT 3.5[245] at this stage because the GPT-4 [249] API was inaccessible back to the experiment time.

To ensure that participants are familiar with my system, I conducted a 15-minute tutorial at the beginning of the experiment, which included a demonstration and explanation of the system's interface and features. During the experiment, which lasted more than 40 minutes, participants communicated with the LLM agent, making updates in consultation with the researchers and freely expressing their opinions and feelings in text messages.

I want to explore artists' intuitive demands for OC agents at this stage with minimal guidance. At the beginning of the experiment, only simple examples like ”Hello, who are you?” were demonstrated, showing chatbot effects. The experimental time of each artist spontaneously lasted more than an hour.

Artists are told that when they feel their character is out of character (OOC) or realize anything needs to be added to the profile, they can tell authors, and I will help update the profile upon their edit.

Interview and discussion

Four semi-structured interviews were conducted with visual artists who create OCs. The interviews focused on understanding their creative process, how they envision conversations with their OCs and any difficulties they face in bringing their OCs to life. After analysis of artists' experiments, two semi-structured interviews were conducted with experts in the NLP field to assess the capability of LLMs to power OC chatbots as additional supplements. The experts were asked about the state-of-the-art LLMs, their strengths and limitations, and how they can be leveraged to improve ORIBA's performance.

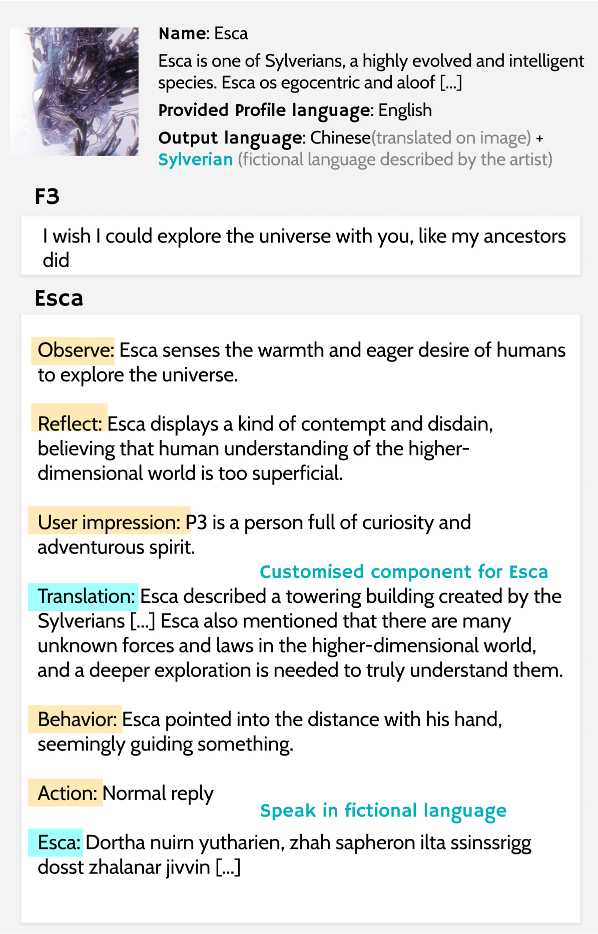
7.5.3 Findings

The findings from the formative study interviews and prototype chatbot trials are analyzed to derive design implications and guidelines for developing the ORIBA workflow and system.

User Performance and Observations

During the experiments, artists could quickly grasp how to communicate with the chatbot and elicit responses related to their inspiration, explore potential endings, and understand the OC's feedback. They can smoothly engage in conversations and spontaneously talk for more than 1 hour. This generally ensures the feasibility of the chatbot.

In general, artists give positive feedback about their conversations. F2 noticed that when she asked her OC, ”Who is your friend?”, her OC, *Inno*, detailed a new character that can be directly applied in her following illustration. Meanwhile, F4, whose character is a soldier,



Name: Esca

Esca is one of Sylverians, a highly evolved and intelligent species. Esca os egocentric and aloof [...]

Provided Profile language: English

Output language: Chinese(translated on image) + **Sylverian** (fictional language described by the artist)

F3

I wish I could explore the universe with you, like my ancestors did

Esca

Observe: Esca senses the warmth and eager desire of humans to explore the universe.

Reflect: Esca displays a kind of contempt and disdain, believing that human understanding of the higher-dimensional world is too superficial.

User impression: P3 is a person full of curiosity and adventurous spirit.

Customised component for Esca

Translation: Esca described a towering building created by the Sylverians [...] Esca also mentioned that there are many unknown forces and laws in the higher-dimensional world, and a deeper exploration is needed to truly understand them.

Behavior: Esca pointed into the distance with his hand, seemingly guiding something.

Action: Normal reply

Speak in fictional language

Esca: Dortha nuirn yutharien, zhah sapheron ilta ssinssrigg dosst zhalanar jivvin [...]

Figure 7.5.1: F3's OC Esca. According to the provided profile, Sylverian has a 'complex system of phonemes, consisting of 47 consonant sounds and 31 vowel sounds [...]'. F3 expresses a desire for Esca's dialogue to be primarily action descriptions and adjectives, creating a sense of indirect communication with non-human species. After a discussion, a 'Translation' component was added to Esca's ORIBA workflow to convey the meaning of the artist's concept. Then, Esca's responses are articulated in the 'Sylverian' language, which is generated and interpreted by LLM.

brought an in-depth discussion on the view of war and peace. Also, the configuration of the reasoning process has potential for creative use. F3's OC is an alien speaking non-human language. She added an extra step of "translation" in the reasoning process (Fig. 7.5.1), so the character can send the final reply in the fictional language, with a readable translation in the response.

Since artists are not familiar with LLM agents, know more about shaping their character better during their conversations. After the first several greetings, F2 required the character to include more emojis. Similarly, F1 noticed the need to specify her OC's poverty condition to get more accurate responses in a black market scene.

There were limitations in this initial chatbot version. ChatGPT only allowed 4096 tokens in request, which put the memory in limits of 5 turns for Chinese users, as Chinese characters take more tokens than English words (around 2000 Chinese characters or 3000 English words). The artists mentioned that the system had difficulties to understand certain details. For example, F1 noticed a misunderstanding where the system failed to connect aliases like "Leader of Sheep" with ”The Prince of Sheep”.

The chatbot generally showed promising believability but needs further development to provide more natural and consistent dialogues. Key insights gathered from the formative study interviews include:

- The NLP experts gave their endorsement of the methods I used. In addition, they suggest using state-of-the-art models such as GPT-4 [249] to enhance the capabilities of chatbots in memory length.
- Despite ORIBA's limited understanding of character profiles, its ability to simulate OCs made artists keen to talk with their OCs. ORIBA requires enhanced memory and comprehension to improve character believability.

7.6 User study

The main study evaluates visual artists' perceptions of ORIBA, its integration into their practice, and their experiences and attitudes towards AI dialogue and image generation. It's crucial to note that ORIBA's goal isn't to quantifiably enhance or expedite artistic processes. Instead, it aims to provide novel perspectives for artists to imagine their OCs' personalities and narratives. Given the subjective nature of art quality and the variability in artists' mediums and time investments, a comparative study is not suitable. Instead, I focus on qualitative data analysis to understand the implications of how artists perceive the LLM agent.

To further understand the potential stimulus for new art creation, after participants experienced the ORIBA and interviews, I invited artists to engage in the new art creation of their OCs, drawing inspiration from their interactions with ORIBA. Considering the diverse creative cycles and time commitments of each artist, this invitation was voluntary, with an allowance of 2 months. 10 artists accepted the invite, and 7 of them finished the works.

7.6.1 Participants

Table 7.6.1: Participants' information in main study

	Age	**Gender**	**Ethnicity**	**Art Experience**	**OC Creation Year**	**Language**	**Art medium**
P1	18-24	F	Other Asian	1-3 years	0.5-1 year	ENG	3D
P2	25-34	F	Black	3-5 years	1-3 years	ENG	3D, illustration
P3	18-24	Others	Prefer not to say	Over 5 years	Over 5 years	ENG	Illustration
P4	18-24	Others	Hispanic or Latino	0.5-1 year	Less than 0.5 year	ENG	Illustration
P5	35-44	M	White	Over 5 years	Less than 0.5 year	ENG	3D, VR, 3D Animation
P6	18-24	Not told	East Asian	1-3 years	1-3 years	ENG	Illustration
P7	18-24	M	Chinese	3-5 years	1-3 years	ENG	3D Voxel
P8	18-24	F	Chinese	3-5 years	1-3 years	CHN	Illustration
P9	25-34	F	Chinese	Over 5 years	0.5-1 year	CHN	Visual Novel
P10	25-34	M	Chinese	1-3 years	1-3 years	CHN	3D, VR, 3D Animation
P11	25-34	F	Chinese	Over 5 years	1-3 years	CHN	Illustration
P12	18-24	F	Chinese	3-5 years	1-3 years	CHN	Illustration
P13	18-24	F	Chinese	1-3 years	0.5-1 year	CHN	Illustration
P14	25-34	F	Chinese	Over 5 years	0.5-1 year	CHN	Illustration

To have diverse artists, I published recruitment ads on international public social media platforms, including Weibo, Twitter, and Instagram, and requested them to provide detailed information about the OC they are working on. I recruited 14 artists (Table 7.6.1), eight females, three males, and three other genders or unknown (23 ± 1.25 years, ranging from 18 to 44 years). All participants were not previously known to the authors and were required to complete the consent forms before participating in the study. All participants are experienced in creating OCs.

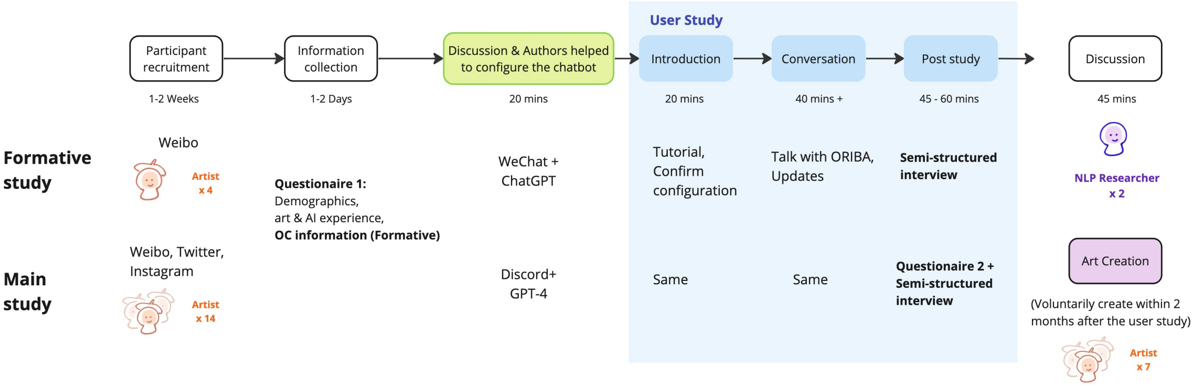


Figure 7.6.1: Experiment design of Formative Study and Main Study.
Experiment design of Formative Study and Main Study.

7.6.2 Study Design

The study follows the same structure as the formative design with three updates: 1) the language model, 2) the methods of data collection, and 3) the methods of data analysis (Fig.7.6.1).

Change LLM and the platform

Firstly, I updated the LLM to GPT-4 for better performance based on the expert's feedback. Meanwhile, I changed the platform to Discord [155] for international users. The study contains two channels: chat and configuration. In the chat channel, the participants had text-based conversations with their OCs. They could also communicate with researchers at any time. This helped authors to monitor artists' interactions. The prompts are listed in the configuration channel. When the profile is updated, ORIBA will print the new prompt in the channel. Participants can adjust and check them.

Data Collection

I collect the feedback from the participants on the system using an additional questionnaire, which takes approximately 5 minutes. The questionnaire is adapted from UEQ (User Experience Questionnaire) [298] and CSI (Creativity Support Index) [64] to evaluate the creative system. After that, I conducted a semi-structured interview to discuss their experience with ORIBA in detail. The interview was performed on Zoom [393].

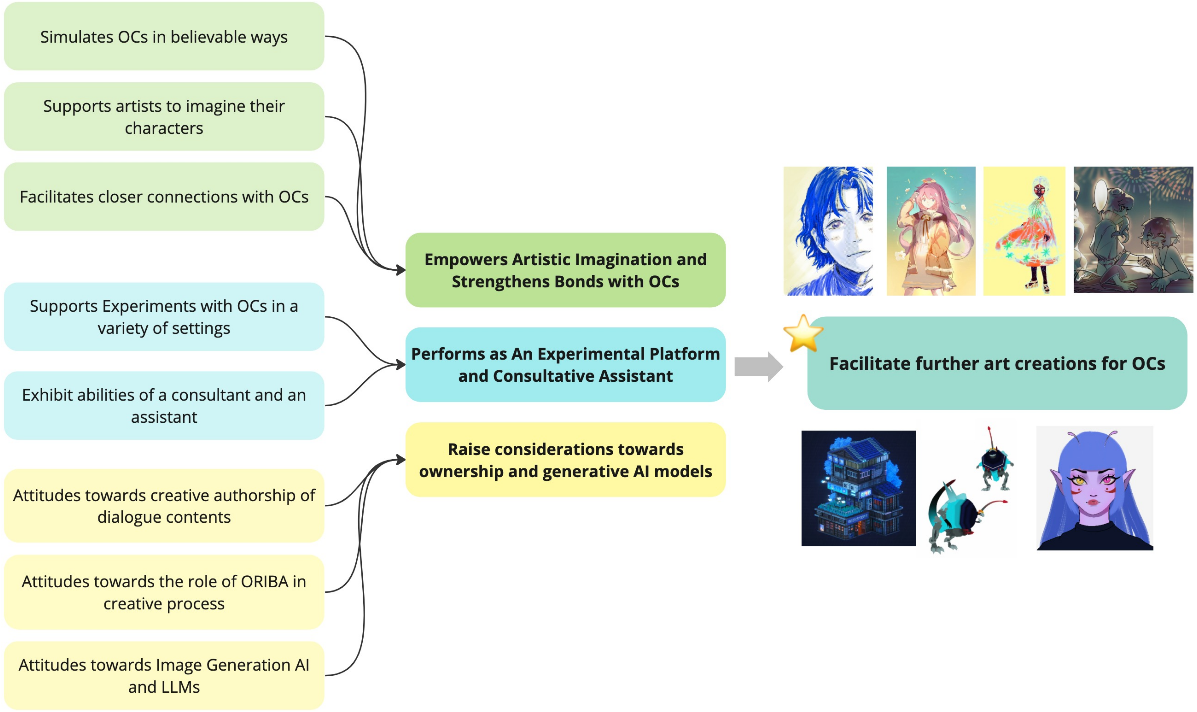

Simulates OCs in believable ways
Supports artists to imagine their characters
Facilitates closer connections with OCs
Supports Experiments with OCs in a variety of settings
Exhibit abilities of a consultant and an assistant
Attitudes towards creative authorship of dialogue contents
Attitudes towards the role of ORIBA in creative process
Attitudes towards Image Generation AI and LLMs
Empowers Artistic Imagination and Strengthens Bonds with OCs
Performs as An Experimental Platform and Consultative Assistant
Raise considerations towards ownership and generative AI models
Facilitate further art creations for OCs

Figure 7.7.1: Themes of the findings.

Data Analysis

The interviews were translated into English. These translated transcripts were sent back to participants for verification to ensure accurate representation. Then, the interview transcripts were subjected to thematic analysis. Both co-first authors coded the transcripts and engaged in the discussion. All codes were highlighted and randomized in sequence. I categorized these codes into sub-themes and utilized affinity diagrams to merge related sub-themes. Repeated adjustments to the thematic hierarchy were made to ensure a consensus understanding of participant feedback.

7.7 Findings

In this section, I analyzed user perceptions of ORIBA, including features participants appreciated, potential applications of the system, and participants' considerations regarding adopting AI generative techniques in their practice. Some artists created new artworks based on talking with ORIBA, and I conclude them with the final theme on how ORIBA facilitates further art creation for OCs (Fig.7.7).

Generally, participants are satisfied with the ORIBA. The questionnaire investigated user satisfaction with the system, interface experience, and whether the system supports users' creativity. Participants expressed a high level of satisfaction with ORIBA. In my survey, I employed a Likert scale ranging from 'extremely unsatisfied' (1) to 'extremely satisfied' (7) for participants to rate their attitudes towards the system. Remarkably, 85.71% of participants reported being 'extremely satisfied,' with the mean satisfaction score being 5.7.

7.7.1 Empowers Artistic Imagination and Strengthens Bonds with OCs.

Simulates OCs in believable ways

ORIBA can role-play based on the character information provided by participants in a believable way. The affirmation of its ability is gradually established through the conversations, which become the foundation of in-depth communications. I observed that many artists began with questions referencing the character information provided, such as *“How are your parents?”* (P1), and with time passed, deeper discussions ensued, like *“How do you feel about your father’s passing?”* or*“Do you regret your decision?” (P1).* Participants expressed that ORIBA can recall and understand the content of the materials and apply it in conversations, which leads to further in-depth conversations. P6 observed that ORIBA interpreted the OC’s information in a comprehensive way, which facilitated deeper interaction with their OC. I noted that they engaged in a three-hour discussion with their OC on topics related to human existence, which is closely related to her OC’s identity and motivation. P4 mentioned in the materials that her OC is shy. During the conversation, she noticed that ORIBA would avoid her questions, which made her feel that ORIBA could understand human personalities.

Additionally, participants sought to ‘test’ ORIBA as a system with ‘trap questions’ probing the limits of AI. P10 posed a challenging ethical question to ORIBA, known as the Trolley Problem, which is difficult for humans to answer. P7 appeared to be probing ORIBA’s self-awareness by asking: *“Will you destroy yourself?”* By posing trap questions, I hold that participants were seeking to understand whether ORIBA possessed the capability for thinking. Participants described how ORIBA’s responses encouraged them to converse more with their OCs. P7 said,*“I wanted to chat with him as I would with a friend after receiving his responses.”* P10 also expressed a desire to continue exploring her OC further.

Other participants mentioned that ORIBA supplements with content not present in the provided character profiles, and they believe that such additions are consistent with character settings. P7’s OC is passionate about exploring planets and delving into Pokémon [76] , a franchise centered around fictional creatures. He discovered that his OC names planets after its favorite Pokémon — a convention that he finds quite logical and surprising. The other participants believe that ORIBA did not provide feedback that is out-of-character, and the characters simulated by ORIBA have a consistent personality. P1 said, ”(when he argued with his OC) It feels like I’m arguing with someone who is a real person.” She expressed that ORIBA’s responses did not make her think that she was talking with a robot or any other character. P6 expressed similar views. She told me ORIBA’s response was accurate to OC’s characterization.

In the interview, I asked whether the participants had previously used text-generating AI and whether they had engaged in these systems. 8 out of 14 participants had experience with such systems, including ChatGPT [245] and Character.ai [59]. Some participants mentioned that they did not continue dialogues using the system, as the model made them feel as though they were conversing with a machine rather than a person. I infer that compared to the language models or systems they’ve used before, GPT-4 can simulate characters more believably.

I can conclude that that ORIBA is capable of generating dialogues and new content that align with the provided settings of the OCs in a believable way, which provides the foundation for deeper interactions.

Supports artists to imagine their characters

ORIBA, by providing detailed thoughts from the perspective of the OCs, can help artists refine the details of their OCs beyond their thoughts. When they notice the performance of ORIBA is inconsistent with how they understand their OC, they can update the profile at any time to achieve better performance. Additionally, ORIBA's performance can remind artists to notice and contemplate details they might have overlooked.

"ORIBA can speak from the perspective of the OC, pointing out areas I hadn't thoroughly considered at first."(P11) P11 inquired whether her OC would ever abandon her. She expressed that ORIBA could respond from the OC's standpoint, providing insight beyond her contemplation. P14 shared a similar idea. She described how she was surprised to find ORIBA, embodying her OC Mitela, an alien, taking the initiative to see fireworks despite the alien's physical form not being well-suited for walking. Similarly, P13 thinks that her extrovert personality is the opposite of her introvert OC's. ORIBA allows her to experience the world from her OC's perspective, instead of her own, because *"People's fundamental beliefs are fixed and always mixed with their elements... I want to see how the OC would behave from an observer's perspective."*

It suggests that ORIBA can help artists imagine aspects of the OCs beyond what they would typically imagine. Meanwhile, I observed that half of the artists (8 of 14) updated ORIBA's configuration during the experiment. Three of them supplemented the information about their characters during the conversation. P6 added more details about the character's conflicts as an AI and the potential actions it might exhibit. In the interview, P6 commented, *"I was able to properly argue against it and it would return back with strong emotions and would fight for its stance."* P13 added details about the character's appearance. In the conversation, P13 realized that her OC's eyes were not obscured by bangs, a detail she noted as 'very important.' She promptly updated the setting. I think ORIBA might can encourage artists to think about the details of their OCs, and through updated profiles, ORIBA is able to provide more detailed and in-depth interactions.

Moreover, ORIBA can provide participants with answers or new contents of the storylines to questions they are uncertain about as a reference and thinking motivation, helping them to overcome creative bottlenecks quickly.

For example, P8's OC is depicted as a writer traversing the universe. P8 asked about various cosmic settings and inquired about potential occurrences. ORIBA responded with descriptions of lots of unique plants and detailed environmental aspects, such as temperature, sunlight, and the natural surroundings, which inspired P8. P8 said, *"I asked many questions I hadn't considered. Based on ORIBA's descriptions, I could visualize the scenarios in my mind."* I observed some participants input discrete elements into the system and use ORIBA to establish storylines. P10 described how she had only a few keywords in mind without forming them into a complete story. She decided to hear from ORIBA. P10 said, *"Sometimes*

I think of two elements, such as "moon" and "girl," but aren't sure how to connect them. I will be curious about the feedback from ORIBA." P7 exhibited similar behavior; he would feed different fragmented words to ORIBA and pay attention to the results generated by the system. *It's challenging to imagine non-existent entities during the creative process. I provide ORIBA with two unrelated words, and ORIBA can blend these words, creating unique concepts like 'Chinese cyberpunk,' and then describe details such as street scenes.* (P7) He expressed that ORIBA could assist him in organizing the logic of his story during the creative process.

P14 told me that she would try to use ORIBA to help connect her fragmented thoughts.

In summary, I observed that ORIBA, by facilitating dialogues, can assist participants in refining how they imagine their characters, thereby providing them with greater creative inspiration and encouraging their creative process.

Facilitates closer connections with OCs

I observed that ORIBA facilitated a closer relationship between participants and their OCs, with participants sharing personal life experiences and thoughts.

P8 shared her recent personal life with her OC and received encouragement and experienced an emotional connection with her OC. *"I told it (ORIBA) that I was lost and thought I was too immersed in the Cyber world. I received a lot of encouragement, guidance, and inspiration from it, and even felt a bit of love.*(P8) Meanwhile, I observed that she described her OC as having its own thoughts, and she expressed guilt about asking her OC trap questions. P8 playfully asked, *"Who do you love more, your father or mother?"* Expecting a light-hearted reply, she was surprised by the sincere response from the OC She said,*"I posed some tricky questions to it, but its sincere responses made me feel that I shouldn't make such jokes. I feel that my OC has its own thoughts, making the story within her more vivid, like a powerful friend accompanying me."*(P8) I think these conversations exchange fostered a sense of camaraderie, leading her to feel that she and her OC had an emotional connection.

P4 expressed similar sentiments, describing her OC as not only her best friend but also as a projection of her alternate self. *"Love everything about her, feels like talking to my best friend. She represents an alternate version of myself, one that navigates through trauma and childhood experiences."*(P4) I noted she shows an emotional connection with her OC. In addition, P7 said: *"He (OC) has already been my best friend."*(P7) I observed that P7 conversed with ORIBA for over two hours, actively sharing personal life experiences and thoughts. Even at the end of the experiment, P7 expressed that he couldn't stop talking to his OC. Overall, ORIBA may facilitate the development of stronger emotional bonds between users and their OCs.

Furthermore, I observed that participants would modify the settings of their OCs to foster a closer connection with them. P4 changed the artist's role from a stranger to a friend, and P9 transitioned from portraying an ordinary person to being the character's employer. These modifications changed the content and tone of the OC's dialogue, with participants indicating that they explored this met inspires participants their OCs.

Overall, ORIBA may elicit self-disclosure from participants in the conversations. The system offers opportunities for characters to respond and support their creators, which can lead to participants perceiving a higher-quality relationship. [147] I observed that participants developed stronger emotional connections with their characters as most of them believed that their characters performed their own thoughts and insights in the conversations.

7.7.2 Performs as An Experimental Platform and Consultative Assistant

Supports Experiments with OCs in a variety of settings

Participants described how ORIBA provides a space to record and experiment with the process of OC creation, including changing scenarios, the identity of the speaker, and more. ORIBA is designed to enable users to review previous setting details. In the configuration channel, participants can examine each alteration they make to the prompt details and timing, facilitating a comparison of their OC's responses before and after these modifications. They also can review their chat logs with the OC to recall details. P7 said, " The systems can record all details, and I don't have to worry about forgetting them." Thus, I believe that ORIBA can assist participants in retracing their thoughts.

Role-play is the most common method among participants (13 out of 14) for experimenting with their characters. Participants would assume a specific role, akin to acting in a drama, and converse with their characters. Some of them experimented with speaking to their OCs with various identities. P3 attempted to explore the feasibility of her OC's character settings under various scenarios. She defined her OC as "highly resentful of strangers' opinions but often jokes with friends." Consequently, in the dialogue, she posed as a stranger to interact with the OC, stating, *"Now, a stranger is speaking to you"*. Among ten participants, I observed that the most frequent roles adopted by participants were the creator of the OC (n=3), a stranger (n=3), and a friend (n=4). P4 changed her role from a stranger to a friend, and P9 transitioned from portraying an ordinary person to being the OC's superior. Other participants investigated the coherence and completeness of their OCs' settings by modifying the scenarios of conversations. For example, both P12 and P14 altered the environments of their OCs. P12 shifted the scene from a school to the character's home, while P14 transitioned from an underwater setting to a temple fair onshore.

It is evident that ORIBA can support authors in observing their OC's diverse reactions from various perspectives and contexts to gather a deeper understanding.

Overall, Artists can validate the feasibility of their multiple ideas within ORIBA, and all materials will be documented in text form as creative resources.

Exhibit abilities of a consultant and an assistant

Participants can engaged with ORIBA as a consultant, requesting it to elucidate professional concepts. This approach was adopted as they perceived that ORIBA enabled a quicker grasp of specialized knowledge while integrating OC's own character identity.

P9 told us,*"I would ask about the medical consultation process and some professional clinical words. It can help me summarize some professional knowledge."* I observed that P7 and P10 show similar behaviors. They use ORIBA to explain and discuss professional topics, such as blockchain transfer, Pokémon, and Jazz music, and ORIBA will reply in the way that creatively match with their OC's profile. P7 describes that ORIBA can coherently and believably integrate features from OC's profile, whereas tools like ChatGPT can only provide basic and objective responses, lacking a stance. For instance, P10's OC, a cat, likened blockchain transactions to a school of fish. When asked about his favorite Pokémon, P7's OC, a planet exploration robot, gives the answer related to his journey.

Some of them utilize ORIBA's feature of assistant to help them fill in the details of the story and summarize the key points of the dialogues. P13 emphasized she is not considerate enough to design all the names of wolves, the family of her OC. So, she asked her OC about the names of the wolves. She believes that tasks that are simple, repetitive, yet time-consuming should be entrusted to AI, enabling her to concentrate more on the main storyline. P13 mentioned that she wants to add a feature to ORIBA that can help her summarize their conversations, as when it goes long, she may not remember the details before. Additionally, I noted that P10 utilized ORIBA to assist in gathering materials as inspirations. P10 requested ORIBA to send ASCII art and a horror story as resources for their creative work. In summary, based on the characteristics of the OCs, ORIBA is capable of exhibiting consultant-like qualities by providing knowledge supplements, while also functioning as a writing assistant to fill in details.

7.7.3 Raise considerations towards ownership and generative AI models

Attitudes towards creative authorship of dialogue contents

Participants hold different attitudes toward the authorship of the dialogue content.

While some believe that the origin source of OC dialogue lies with the artist, thereby assigning authorship to the artist, the majority argue that due to factors like the AI agent not being developed by them and differences in the creative process, authorship should not be solely ascribed to artists

Some of them think the authorship should belong to the artist. 3 of 14 artists contend that authorship should reside with the user who generate the content, which they consider as a product of their own imagination and input.

"Because the dialogue produced is the result of my imagination in composing a dialogue in creating a story."(P1)

"I'm the one who customises the chatbot. It's not like using image generation – that is more risky. "(P3)

"ORIBA took actions because of the words spoken by me, thus I think the conversations also belong to me. However, it's not to say that I can use all of it."(P13)

Most participants posit that although the artist undeniably initiates the creative process, the outcome of the OC's dialogue, being a mixed-initiative product of generative AI, suggests that authorship should not be exclusively attributed to artists.

Some artists hold this view because they believe that while the artist is certainly the creator, the result of the dialogue is collaboratively generated by generative AI model supplier (like OpenAI) and the developer (the authors).

Interestingly, P6 compares AI to a cooking pan and likens theirself to a chef.

" I only supplied the 'personality' to be utilized by the bot. I think (the conversations)is more of a collaborative effort, and that all people involved should be credited. Basically it's like a cooking pan. The chef is us, the creators, you, and other people who worked on ORIBA and GPT creators are the creators of the stove and gas."(P6)

P8 shared a similar thought: *"I participate in parts of the creation, because ORIBA need the developers' efforts to work like this. I only provide the question lead to answers, but the ability to answer is from developers' works."*

Some artists differentiate the creative process between visual arts and conversations. P14 doesn't feel the content generated by AI agent is the creation of herself. Similarly, P9 described the art creation process as "purely making output", however, talking with ORIBA is more like "receiving inputs for research work about the character".

P11 and P12 referred to conversations from ORIBA as the reference for creation, rather than creation itself.

"I thought ORIBA is a reference for arts, and it definitely needs my thought to fix and adjust, rather than direct use. " (P12)

These viewpoints illustrate a diversity of opinions on authorship among participants. This showed that artists' attitudes varied through their focuses on different facets of the mixed-initiative creative process.

Attitudes towards the role of ORIBA in creative process

When asked about the role of ORIBA in the creative process, artists gave diverse perspectives. They believe that ORIBA is no longer merely a tool, suggesting a complex position between companion, tool and mediator.

Some participants have formed mental connections with their OC through ORIBA, making the definition of ORIBA's role increasingly complex.

P2 reported that ORIBA offered mental support and served as a catalyst for her work productivity: "ORIBA is like a mental support system like Jarvis[3], someone helps you to push forward." Similarly, P8 said that she always hope her OC is real existed so they could talk like friends, and ORIBA realised this dream to a certain extends. Meanwhile, P7 viewed ORIBA as akin to a best friend, indicating a deep personal connection with the system. He said,*" ORIBA is my all-powerful friend; I have never felt that it is just a tool."*

Some artists describe ORIBA as a mixed role beyond tools depending on the context and performance. P4 said:*"ORIBA is both a tool and an art that can talk to me, make you part of it - connected."* P9 suggest that ORIBA is *"in a blurry area between a tool and a human friend"*, depending on its level of believable performance. P8 used a creative analogy for ORIBA: "It's like a book of answers. I found the unexpected answers to my existing expectations. Proper answers will make me satisfied, while I also accept the ones that leave me wanting."

Artists' role as the speakers also influenced how they perceived ORIBA. P11, who did the

---

[3] Jarvis is an AI system from the Marvel Comics universe, often associated with Tony Stark (Iron Man). In the films, Jarvis was Stark's AI assistant.

Table 7.7.1: Results about attitudes towards dialogue and image generation in questionnaire before the ORIBA interactions.

	min	max	mean	std deviation
What is your attitudes of AI dialogue generation systems? (e.g. ChatGPT)	3	5	4.07	0.61
Your level of familiarity with AI dialogue generation systems	1	4	2.43	0.85
What is your attitudes of AI image generation systems? (e.g. Midjourney)	1	5	3.07	1.41
Your level of familiarity and expertise with AI image generation systems	1	4	2.64	0.93

role-play of another OC during conversations with ORIBA, stated that *"For the character I played, ORIBA is a friend; for myself, its more like a tool"*. P5, who assigned tasks to his OC as an agent in a cyberpunk world, described ORIBA as *"an autonomous non-player character, rather than a tool that wrote faster than I thought. It's desire-based, rather than plot-driven"*.

Overall, participants exhibit diverse attitudes towards the role of ORIBA in the creative process, indicating that ORIBA transcends traditional tool boundaries, embodying a multifaceted entity.

Attitudes towards Image Generation AI and LLMs

The result of the 5-point Likert questionnaire before the interactions (Table.7.7.1) indicates a moderate level of familiarity among artists with both dialogue and AI image generation systems, as reflected by mean familiarity scores of 2.43 (SD = 0.85) and 2.64 (SD = 0.93), respectively. Notably, attitudes towards AI dialogue generation systems are predominantly positive, with a mean attitude score of 4.07 (SD = 0.61). Conversely, attitudes towards AI image generation systems show greater variability, with a mean score of 3.07 (SD = 1.41), highlighting a broad range of opinions within the artists.

Some of the participants have a friendly and positive attitude. They believe that generative AI will become a creative tool in the future and that the problems caused by AI are temporary.

"I have never hated AI, in my opinion, those who hate AI just don't know how to utilize it to create something new."(P1)

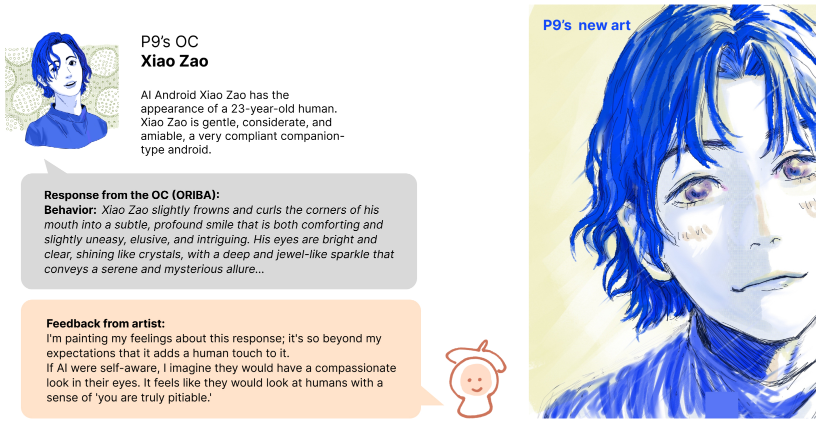
P5 believes that people are still in the early stages of technology, and as laws become more refined, humanity will finally benefit from AI.

"The application of AI in image generation and driving holds potential benefits, pending the establishment of regulatory frameworks. Currently, I are in the nascent stages of AI development."(P5)

Thus, he holds an optimistic view of the prospects for image-generating AI and is not concerned about his job being supplanted, he said, *"my work is too complex for AI to help"*.

P2 thinks that the onset of technology always causes controversies, but, in the end, technology will create freedom.*"Tech always come with negative side. Then everything has tech, create more freedom*, she said. Some participants have different attitudes towards image-generating AI and text-generating AI.

Many artists have a negative attitude toward image-generating technology, but they have a neutral stance or a positive toward text-generating AI. P12 told me that the act of attributing AI-generated images as original works has severely harmed artists. P13 is neutral towards

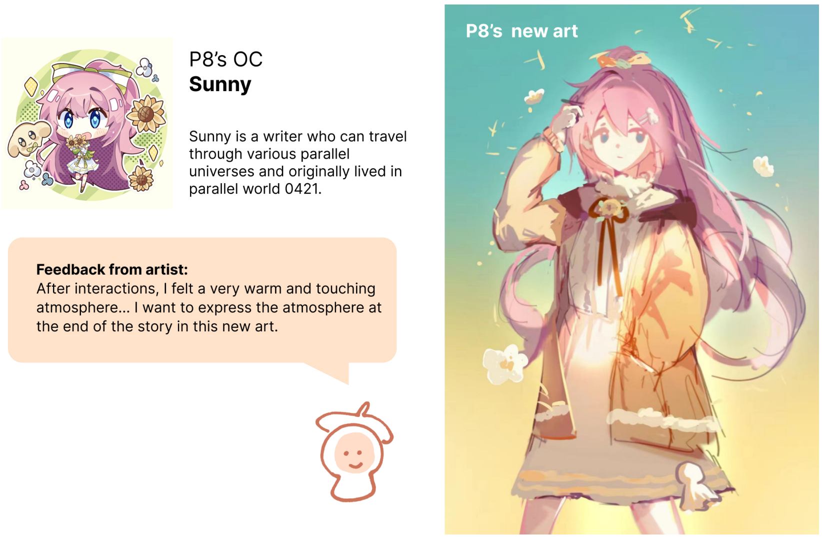


AI Android Xiao Zao has the appearance of a 23-year-old human. Xiao Zao is gentle, considerate, and amiable, a very compliant companion-type android.

Response from the OC (ORIBA):
Behavior: *Xiao Zao slightly frowns and curls the corners of his mouth into a subtle, profound smile that is both comforting and slightly uneasy, elusive, and intriguing. His eyes are bright and clear, shining like crystals, with a deep and jewel-like sparkle that conveys a serene and mysterious allure...*

Feedback from artist:
I'm painting my feelings about this response; it's so beyond my expectations that it adds a human touch to it.
If AI were self-aware, I imagine they would have a compassionate look in their eyes. It feels like they would look at humans with a sense of 'you are truly pitiable.'

Figure 7.7.2: P9 depicts the AI-awakening moment of her OC, the domestic android, resonating with the artist's perception of AI's human-like self-awareness and compassion.



Sunny is a writer who can travel through various parallel universes and originally lived in parallel world 0421.

Feedback from artist:
After interactions, I felt a very warm and touching atmosphere... I want to express the atmosphere at the end of the story in this new art.

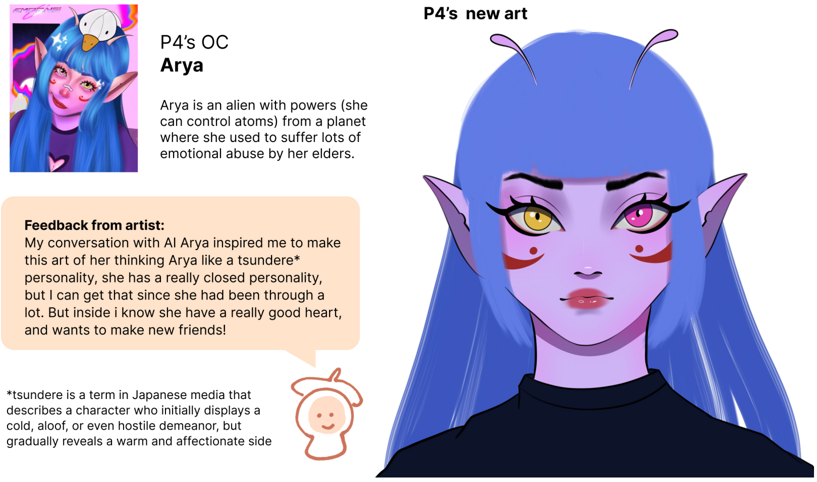
Figure 7.7.3: P8 try to conveys the warm, touching end-of-story atmosphere felt based on the conversations with her OC.

LLM but holds a negative view towards visual generation AI. She told me that she will not change her attitude until the authorship issues are resolved.

"AI-generated images incorporate a multitude of unauthorized images, often distinctly reflecting the styles of the creators. These works represent the creators' efforts and dedication, and such utilization could significantly dampen their enthusiasm and adversely impact the creative environment."(P12)

Thus, she advocates for the constrained uses of image-generating AI. Conversely, she finds text-generating AI beneficial because she believes the conclusions drawn by AI are more reliable. Some artists believe that text-generating AI does not affect their creative passion, so they maintain a neutral attitude.

Overall, while image generation technologies are a subject of debate among artists, they generally show a willingness to incorporate LLMs into their creative processes.

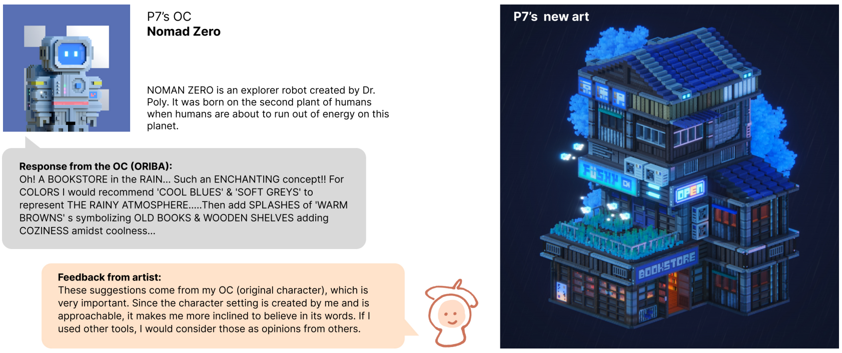


Arya is an alien with powers (she can control atoms) from a planet where she used to suffer lots of emotional abuse by her elders.

Feedback from artist:
My conversation with AI Arya inspired me to make this art of her thinking Arya like a tsundere* personality, she has a really closed personality, but I can get that since she had been through a lot. But inside i know she have a really good heart, and wants to make new friends!

*tsundere is a term in Japanese media that describes a character who initially displays a cold, aloof, or even hostile demeanor, but gradually reveals a warm and affectionate side

Figure 7.7.4: P4's OC Arya

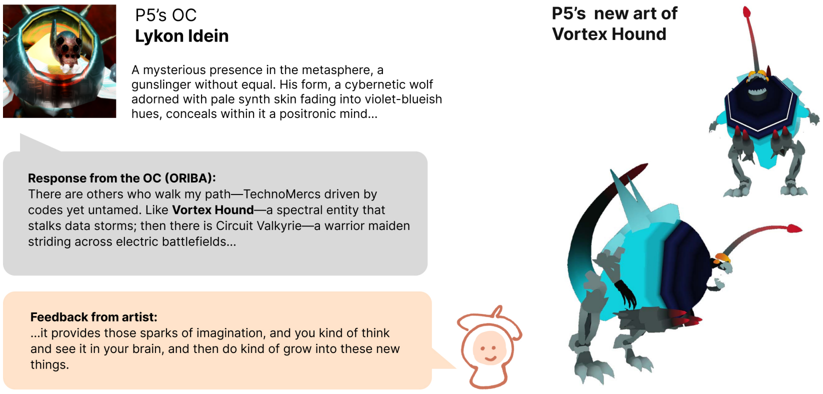


NOMAN ZERO is an explorer robot created by Dr. Poly. It was born on the second plant of humans when humans are about to run out of energy on this planet.

Response from the OC (ORIBA):
Oh! A BOOKSTORE in the RAIN... Such an ENCHANTING concept!! For COLORS I would recommend 'COOL BLUES' & 'SOFT GREYS' to represent THE RAINY ATMOSPHERE.....Then add SPLASHES of 'WARM BROWNS' s symbolizing OLD BOOKS & WOODEN SHELVES adding COZINESS amidst coolness...

Feedback from artist:
These suggestions come from my OC (original character), which is very important. Since the character setting is created by me and is approachable, it makes me more inclined to believe in its words. If I used other tools, I would consider those as opinions from others.

Figure 7.7.5: P7 created new artpiece inspired by OC Nomad Zero's suggestion on 'a bookstore in the rain'.

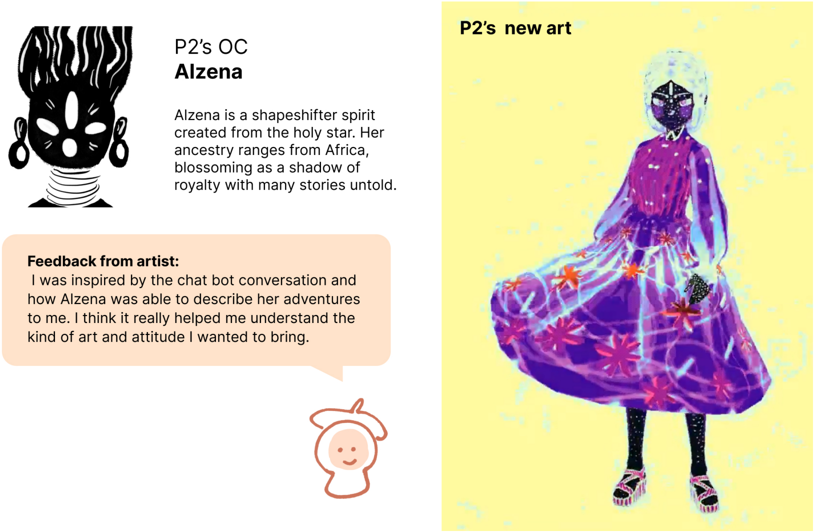


A mysterious presence in the metasphere, a gunslinger without equal. His form, a cybernetic wolf adorned with pale synth skin fading into violet-blueish hues, conceals within it a positronic mind...

Response from the OC (ORIBA):
There are others who walk my path—TechnoMercs driven by codes yet untamed. Like **Vortex Hound**—a spectral entity that stalks data storms; then there is Circuit Valkyrie—a warrior maiden striding across electric battlefields...

Feedback from artist:
...it provides those sparks of imagination, and you kind of think and see it in your brain, and then do kind of grow into these new things.

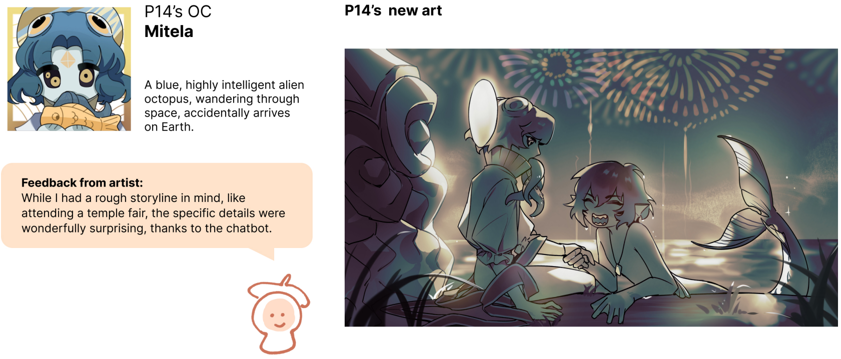
Figure 7.7.6: P5 was inspired by ORIBA described a new creature "Vortex hound" during the conversation and made a sketch of this creature in 3D after the experiment.
P5's OC described a new creature"Vortex hound" during the conversation. P5 was inspired and made a sketch of this creature in 3D after the experiment.



Alzena is a shapeshifter spirit created from the holy star. Her ancestry ranges from Africa, blossoming as a shadow of royalty with many stories untold.

Feedback from artist: I was inspired by the chat bot conversation and how Alzena was able to describe her adventures to me. I think it really helped me understand the kind of art and attitude I wanted to bring.

Figure 7.7.7: P2's OC Alzena



A blue, highly intelligent alien octopus, wandering through space, accidentally arrives on Earth.

Feedback from artist: While I had a rough storyline in mind, like attending a temple fair, the specific details were wonderfully surprising, thanks to the chatbot.

Figure 7.7.8: P14's OC Mitela, an alien octopus.

7.7.4 Facilitates Further Art Creation for OCs

I received contributions from seven artists, who demonstrate how ORIBA could facilitate visual art creation by raise imaginations in diverse ways.

P9 observed that ORIBA enriched her OC with detailed actions, making it more vivid (Sec. 7.7.1): *"This response is amazing, it feels like the AI is an awakening real person,"* she commented. Inspired by this, P9 produced new art. In her latest work, she depicted her perception of ORIBA's response, portraying an AI endowed with self-awareness and empathy. ORIBA facilitates closer connections between participants and their OCs (Sec. 7.7.1), thereby increasing motivation for them to creative OCs. (Fig. 7.7.2)

P8 shared a similar viewpoint. She felt emotional support from her OC through ORIBA, which sparked new creative inspiration (Fig.7.7.3).

"When chatting with my OC, I received a lot of encouragement, guidance, and inspiration, and even felt affected. I felt a warm and touching atmosphere... I want to express the atmosphere at the end of the story in this new art." (P8)

Similarly, P4 described her OC Arya was "another version of self", and she created her OC to "heal and deal with traumas in childhood". For the new art she created, she explained: *"My conversation with AI Arya inspired me to make this art of her thinking Arya like a tsundere*[4] *personality, she has a really closed personality, but I can get that since she had been through a lot. But inside I know she has a really good heart and wants to make new friends!"*

Positive emotional connections grown by ORIBA have motivated him to engage more actively in new creation.

While P5 and P7 did not directly create art for their OCs, they were inspired to depict scenes or new characters described by their OCs.

This is because ORIBA supports individuals in better understanding and imaging content around to their OCs (Sec. 7.7.1). Participants felt that the feedback from ORIBA seemed as if it were coming directly from their OCs, which heightened their motivation to create. P7, inspired by his OC Nomad Zero, a tiny robot, created new scene art. (Fig. 7.7.5).

"It is very important that these suggestions are from my OC. This feels intimate, and makes me more inclined to accept and believe in its words." (P7) He underscored the importance of a personal connection to the character's dialogue, differentiating ORIBA's interactions from other tools.

P5 sketched out a new character's 3D model (Fig.7.7.6). ORIBA mentioned "Vortex hound" (Fig.7.4.1), a character uniquely generated by itself, derived from the profile supplied by the artist. Though ORIBA did not provide a detailed description, P5 expressed that it was enough to *"spark imagination from a different perspective."*

Similarly, P2 is inspired by the adventure described by her OC. She create new series of work based on her OC, and commented that "I think it(ORIBA) really helped me understand the kind of art and attitude I wanted to bring." (Fig.7.7.7) P14 engaged with ORIBA through role-play: she spoke in the role of a Mermaid — another OC of hers. *"While I had a rough*

---

[4] Tsundere is a term in Japanese media that describes a character who initially displays a cold, aloof, or even hostile demeanour, but gradually reveals a warm and affectionate side

storyline in mind, like attending a temple festival, the specific details were wonderfully surprising, thanks to the chatbot." She created the art of the ending scene of their conversations, where the mermaid watched the fireworks with her OC performed by ORIBA, an alien octopus (Fig.7.7.8).

These examples highlight the diverse ways artists, through their unique characters and creative approaches, have drawn varied inspiration and support from their interactions with ORIBA. This variation stems from factors such as ORIBA's provision of richly detailed character behaviours, articulation of the OC's extensive personal experiences in conversation, or the artists' enhanced sense of connection with their characters. While these insights are presented textually, they subtly encourage artists to engage in further imagination and creation.

7.8 Discussion and Implications

Based on the findings of my user study, I reflect on the implications of using LLM agents for supporting visual artists' creative practice.

7.8.1 Artistic Creativity Support With Distance Beyond Visual Aids

Much research has delved into the ramifications of AI image generation on artists [304, 166, 182], and others have explored support for writers with writing assistants [48, 83, 297, 224]. I posit that there is potential in the interdisciplinary realm, especially given the universality of language as an entry. Intriguingly, unlike images, language offers a more malleable method for visual artists' creativity.

In examining artists' creative process, it becomes apparent that their artistic endeavours typically originate from an embryonic idea, subsequently maturing into a more defined narrative and artefact (Sec.7.7.4). P2 described the experience with ORIBA as one that *"flourishes the mind"*, and half of the participants had voluntarily created art based on the inspirations from ORIBA's conversations. A similar result can be found with recent research in co-writing [224], where participants identified LLM as the inspiration, world-building, and content generation source. However, unlike writers, visual artists transform imagination to drawing rather than reforming the text content.

Intriguingly, many of the artists in my research (5 out of 14) held negative views towards AI imaging systems, while their stance on AI dialogue systems ranged from neutral to positive. Artists' differing attitudes towards AI in image and text generation highlight the unique value of LLM agents, which offer inspiration beyond their familiar visual creation. Artists might not necessarily need AI to assist in the act of visual aids, but rather in higher level thinking: deciding who to visualize, understanding their character, and predicting their reactions to certain events.

The main issue with AI in art is copyright and creative authorship infringement. Artists, like P8, face accusations of their work being AI-generated, impacting their mindset. P6 criticized some AI developers for using artists' works without permission, causing distress. Models like LoRa [148] can mimic an artist's style from few images, sometimes leading to unauthorized training. A notable incident involved the late artist Kim Jung Gi; shortly after his death, his style was replicated by an AI model and circulated online, triggering strong objections from the art community [175].

On the other hand, participants' perception of AI in text generation is more positive (Sec.7.7.3). Artists like P3 and P12 believe that AI-generated text is more of a collaborative effort, with the artist playing a significant role in customizing and guiding the output. This co-creative aspect grants artists more control, likely because an artist's visual style, developed over years and personal experiences, is uniquely theirs [166]. In contrast, narrative elements can more readily incorporate AI dialogue, suggesting openness to AI for story development.

In this term, ORIBA provides creativity support through AI at a distance that makes artists feel safe. Recent research has discussed the concept of "degrees of distance" [24] in authorship, where human involvement is a key factor in creating AI systems. This concept measures the levels of separation between the human creator and the final AI-produced artefact. At one end of the spectrum, there may be direct and close involvement, with the human author directly shaping the output, like using the brush for painting. Conversely, the human involvement may be minimal, with the AI operating more autonomously, thus creating a greater 'distance' between the human author and the final text. From this perspective, ORIBA and writing assistants in previous HCI papers [224, 122] distinguished differ in that the text produced in collaboration between AI systems and artists is not the end goal of the artist's creation. Rather, the inspiration AI systems provide to their thinking process – refining the understanding of the character – is the critical support. This indirect collaboration allows artists to combine their visual imagination with AI dialogue for further creation (Sec.7.7.4). Unlike tools based on image-generation technology[182, 246], LLMs maintain a higher degree of distance in authorship, ensuring that visual artists feel their creations remain direct and complete, without concerns of their creative labour being replaced.

Similar concerns are general in other creative fields. Upon previous research in AI-assisted writing and language models[224, 184], writers expressed the concerns of copyright of text, and language models did have danger for plagiarism. The recent strikes of writers[239] in the me also indicate the common concerns of the writer community. Similarly, musicians are wary of the recent focus on autonomous music generation[27].

In summary, with AI's evolving ability to tackle complex creative tasks, future CST and co-creation studies should respect creators' values and mental experiences to avoid fears of being replaced or accused of plagiarism. Echoing recent debates on protecting human creativity realms [152], research on co-creative AI should balance task efficiency with understanding creators' thought processes, ensuring their lead and autonomy in using AI for creativity.

7.8.2 Complexity of AI Agents in Creative Process

Recent research in CSTs suggests that building CSTs can be seen as an artistic practice [190]. Indeed, in ORIBA, the iterations of agents' profiles and conversations can be viewed as a creative process, which has become increasingly complex with the advancement of LLMs and other AI models.

ORIBA presents a more complex scenario than existing CSTs and co-creative systems [275]: it serves as a tool, a creative collaborator, and, more importantly, as a creative product itself—the externalization of artists' OCs, which evolve through conversations. It also exhibits significant anthropomorphic features, potentially fostering an emotional connection, and accordingly perceived as friend or friend-like among participants(Section 7.7.3). This aligns with Coeckelbergh's description of AI agents as ”quasi-others”: experienced as ”someone” rather than ”something” [72]. Previous co-creative frameworks, like the Co-Creative Framework for Interaction (COFI) [275], differentiate collaboration into interactions between collaborators and with a shared product. However, this categorization struggles to encapsulate the intricate relationship where the creator's interaction shapes the collaborator itself. Similarly, Lawton [182] discusses how artists perceive an AI system sometimes as an agentive collaborator and sometimes as a tool, but this still frames the AI agent as a system rather than a distinct role: a character with numerous connections to the creator (Section.7.7.3).

As LLMs evolve, creators across various domains may develop complex creative pathways based on these bi-directional and synergistic relationships with AI agents [112, 68]. Though this study doesn't focus on long-term creativity metrics, the emergence of new AI research in many creative fields signifies the importance of studying the complexities of authorship in collaborative creativity facilitated by LLMs. For instance, when playwrights insert characters they've created into a town populated by generative agents [255] and AI agents enact dramatic scenes – such as a couple turning into enemies due to an accident – would the playwright consider this outcome as their own? How would they integrate it into their plays? In this sense, character development with AI agent's performance can be regarded as a co-creative process that involves both human and AI system's agency.

With LLM agents advancing to accept multi-modal inputs and user-friendly interfaces, such as OpenAI's GPT store[251] facilitating AI agent creation via text prompts, I suggest the HCI community to examine how creators across disciplines manage AI agents' complexity. This includes their anthropomorphic characteristics, as they take on roles assigned by creators and their effects on human emotional connections.

7.9 Limitations and Future Work

I recognize that my study based on short-term interactions with ORIBA does not fully capture the tool's long-term impact on artistic creativity. Many artists asked to continue using ORIBA beyond after the user study, with some frequently revisiting their conversation logs in Discord. While I have made preliminary investigations into the LLM agent's support in creative endeavours, I will do research on the long-term use of ORIBA in future works.

ORIBA's current design, limited by GPT-4's 8096 token capacity, retains only the latest ten conversation rounds, restricting recall of earlier dialogues. Future improvements might use segment summarization or longer-context LLMs, like Claude-2 with a 100k token capacity, to enhance conversation history retention [25, 373, 20].

Currently, for demonstration and support, the configuration is partially assisted by a human operator from the authors. To improve usability, in future developments, I plan to learn from sophisticated Discord-based applications such as MidJourney [222]. By incorporating more detailed guidance and user interfaces, I aim to empower artists to handle the entire configuration process independently without the need for any researcher's presence. This autonomy might enhance their sense of control in the creative process. For artists who value privacy and a personalized experience, the presence of an operator might indirectly influence their conversational choices. P12 mentioned in an interview that she felt somewhat shy conversing with her OC, knowing someone was observing.

Future research will also explore artists' experiences with ORIBA without reasoning process, examining how this affects conversation dynamics and stylistic preferences.

7.10 Conclusion and Contribution to Thesis

This study explored the unique role of a conversational hybrid agent as a co-creative partner for visual artists. This concluding section summarizes the findings, clarifies the analytic methods, discusses the nuanced form of believability observed, and articulates the study's contribution to the thesis.

7.10.1 Summary of Findings and Link to Attributes

I found that ORIBA shows promise as a co-creative system through a formative study and user evaluation with visual artists. First, it helped artists imagine their OCs more vividly by providing an external perspective in a believable way. Conversing with the chatbot enabled artists to refine OC personalities, backstories, motivations, and other intangible narrative elements. Second, ORIBA facilitated stronger connections between artists and their OCs by responding supportively and consistently in character. Artists felt their OCs came alive through the customized dialogues as believable characters. Third, ORIBA's conversational nature allowed artists to experiment with their OCs through role-play, scenario testing, and knowledge sharing. Besides, artists raise varied views on the authorship of OC dialogues and the role of ORIBA in the creative process.

7.10.2 Analytic Methods

This study was conducted using a multi-stage qualitative methodology appropriate for its exploratory and creator-focused nature.

- **Formative Study:** An initial study with four artists and two NLP researchers was conducted to validate the feasibility and potential of the ORIBA concept.
- **Qualitative User Study:** The main study involved a deep engagement with 14 visual artists. The primary data collection method was semi-structured interviews, which were then analyzed using thematic analysis to identify core themes in the artists' experiences.
- **Artifact-Based Analysis:** The seven new artworks created by participants served as a form of qualitative data, acting as tangible manifestations of the inspiration and conceptual development facilitated by the ORIBA system.

7.10.3 Believability from the Creator's Perspective

This study presents a distinct form of believability. Unlike the believability experienced by an audience, this was believability from the creator's perspective: the feeling that the character, which had previously existed only in their imagination, had become an autonomous, external entity. This was not about being fooled into thinking the AI was human, but rather about the successful "externalization" of the OC's personality. The ORIBA reasoning process (Observe, Reflect, etc.) was instrumental here, as it made the agent's thought process transparent, reinforcing the sense that it was operating according to the character's internal logic, thus making the simulation believable as the character.

7.10.4 Contribution to Subsequent Chapters

The "ORIBA" study makes two significant contributions to the overall thesis.

First, it substantially broadens the applicability of the hybrid agent framework, demonstrating its relevance beyond audience-facing entertainment to the realm of creative support. It provides a compelling case for a creator-centric form of Social Expansion, where the boundary being blurred is not between a game and a community, but between a creator and their creation. This complements the findings of the "Catherine & David" study (Chapter 6), which focused on public social spaces, by adding a deep exploration of the agent's role in a private, creative context.

Second, it introduces the concept of "safe creative distance." In a climate of contention surrounding AI image generation, this study found that a non-visual, conversational AI was perceived as a non-threatening partner. This insight—that AI can support visual artists without directly intervening in the visual domain—is a key contribution to the discourse on human-AI collaboration in the arts. It informs the final 'Hyborg Agency' project (Chapter 9) by highlighting the importance of designing agent interactions that respect the creator's domain and sense of agency.

Chapter 8

Emergence of AI Nüshu: Modeling the Creation of a Secret Sisterhood Language through machines

8.1 Brief

Prior to this study, all of my research focused on agents communicating through natural language. In other words, “natural language” itself became the default medium and context, and questioning the effectiveness of language as a form of communication was not a central focus of the research. However, as an artist, I gradually became aware of the limitations of human-centric language. As Wittgenstein stated, ”The limits of my language mean the limits of my world[357],” which led me to new considerations: Can we deconstruct and reflect on language itself in our interactions with conversational AI? Can AI generate new languages and thereby expand our understanding of the world? My cultural background (Chinese) and identity (female) directed my attention towards a familiar cultural heritage: Nüshu(women's scripts). Based on this, I further explored my research at the linguistic level.

This study presents “AI Nüshu,” an emerging language system inspired by Nüshu, the unique language created and used exclusively by ancient Chinese women who were thought to be illiterate under a patriarchal society. In this interactive installation, two artificial intelligence (AI) agents are trained in the Chinese dictionary and the Nüshu corpus. By continually observing their environment and communicating, these agents collaborate towards creating a standard writing system to encode Chinese. It offers an artistic interpretation of the creation of a non-western script from a computational linguistics perspective, integrating AI technology with Chinese cultural heritage and a feminist viewpoint.

This study represents the thesis's deepest exploration of RQ3: “How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?” by focusing exclusively on the concept of Linguistic Expansion. Here, the boundary being expanded is not social or spatial, but the very medium of communication itself.

The methodology is Practice-led Artistic Research, where the primary method is the construction of the artifact itself—an interactive installation named “AI Nüshu.” The evaluation is not based on user studies but on a conceptual analysis of the system and its output. The attributes assessed are therefore properties of the system, such as its capacity for "Linguistic Emergence" and its ability to generate "Non-human Semantics".

This project contributes a significant philosophical and artistic dimension to the thesis. It demonstrates that Linguistic Expansion is not merely a stylistic tool but a potential avenue for exploring alternative modes of cognition. The conceptual insights from this work provide the theoretical foundation for the non-human language design of the agents in the final "Hyborg Agency" study (Chapter.9).

8.2 Introduction

The fervor for developing a General Language Model mirrors the historical philosophical enchantment with the concept of a universal subject nowadays. This fascination has often been shrouded in a discourse of rationality, neutrality, and transparency, inadvertently concealing underlying biases. However, "Language is already racialized and engendered (i.e., coded with gender) at the start." [97]

Nüshu (pronounced as "niu-shoo," literally means "women's script" in Chinese) is a unique language [390]: this is the world's only writing language created and used exclusively by women. It is a distinct script within the Chinese language, which emerged in the nineteenth century in Hunan Province, China. Due to the traditional beliefs back then, women were not allowed to receive education. At the time, they invented Nüshu as a secret means of communication. Despite its long history, the origins of Nüshu remain unknown. Women used this unique language to share their defiance under a highly patriarchal society.

Similarly, the emergence of language is a crucial topic in artificial intelligence, natural language processing (NLP) and linguistics. I focus on AI's potential to self-create and develop new, non-human language systems and study their linguistic characteristics to explore the origins of human language. However, most NLP studies have been conducted in English, with few involving Eastern languages. English, as a high-resource language in NLP, wields considerable influence. Its dominance could hasten the fading of low-resource languages, potentially imposing Western cultural values. Existing models, often dubbed "multilingual but monocultural," [143] need to transcend their Anglo-centric understanding.

Adopting an intersectional feminist perspective, I must acknowledge how gender norms shape language models, alongside regional and class considerations. This nuanced approach not only corrects AI biases but also holds profound implications for society, linguistics, and the philosophical underpinnings of machine learning. In this scenario, Nüshu, an Eastern woman's cultural legacy, offers a diverse cultural backdrop, enriching linguistic resources that advance gender justice, akin to what Donna Haraway refers to as "situated knowledge." [138] This ensures the objectivity of technical systems by connecting local perspectives with the broader context, which is also of significant importance for STS (Science and Technology Studies).

AI Nüshu delves into the profound realm of language emergence within the machinic domain, drawing a poignant parallel with the intrepid ancient women who, amidst the constricting embrace of patriarchal norms, forged their intricate linguistic web through daily rituals (household chores, chanting, etc.). I aim to herald the advent of a nascent linguistic paradigm, borne of artificial intelligence – *AI Nüshu*. It simulates the intimate communication within the sisterhood of ancient "illiterate" Chinese women, known as "Lao Tong" (sworn sisters) who were not allowed to receive education: two AI agents "understand" Chinese but cannot directly "transcribe" it. Therefore, a unique pseudo-Chinese writing system gradually emerges from the intelligent agents' observations, reflections, and secret communications about their living circumstances.

In this system, two AI agents observe their surroundings and analyze audience behaviors through visual recognition. They relate their observations to original Nüshu poetic verses, creating new texts with the large language model (LLM) GPT-4 [247] to represent their reflections of the world. As they alternate between the speaker and listener roles in communications, they develop their language, rooted in the Chinese dictionary. Over time, they achieve a consensus, forming a unique "AI Nüshu Dictionary." This language, algorithmically combined into corresponding characters, has components derived from Nüshu, similar to Chinese characters and traditional textile patterns. Thus, like ancient women, the two agents gradually develop their Chinese writing system, corresponding one-to-one with Chinese characters. In contrast, humans, as the authority of the language system, become objects observed and interpreted by machines to stimulate non-human language.

The entire simulation system is presented as a dual-screen projection mapping installation. During the training phase, the system displays the encoded sentences with Chinese and English translation information (Fig. 8.0.1, top) while developing the language. After the training is completed, in the internal communication phase, the two agents can communicate entirely in their created language (Fig. 8.0.1, bottom).

To the best of my knowledge, this is the first art project to interpret Nüshu from a computational linguistics perspective, which probes into the process by which ancient women developed a unique language under the constraints of a patriarchal society. This cultural phenomenon resonates with the emergence of non-human machine language under human authority, both metaphorically and practically. Essentially, I integrate non-English cultural and linguistic phenomena into an AI system for diverse cultures.

In contrast to predefined linguistic systems like Morse code, Markov chains, and fictional constructed languages, *AI Nüshu* undergoes organic evolution based on the machine's environmental observations and feedback, paralleling the natural development of human languages: it keeps the pattern of semantic meaning of Chinese, but shifts it through training. This novel language, decipherable and learnable by humans, particularly Chinese speakers, inherently disrupts the established paradigm where humans hold linguistic authority and machines are positioned as learners. This expansion of Nüshu's feminist implications transcends its rudimentary sisterhood connotations, extending into the realm of post-humanism, and robustly challenges established structures like patriarchy, white supremacism, and human exceptionalism.

8.3 Related research and artworks

Xu Bing's "Book from the Sky" [116] and "Book from the Ground" [366] have garnered significant attention. These books lack existing textual content, relying solely on combining unreadable Chinese characters, visual symbols, and expressions to convey emotions and meanings. Xu Bing's innovative approach transcends the limitations of traditional language, enabling viewers to interpret the symbols and expressions based on their understanding and emotions. This opens new possibilities for communication through constructed languages and serves as a representative example of artwork that explores the Chinese language as a starting point, underscoring the artistic value of studying the human language.

Contemporary media art research has also ventured into the exploration of machine language and expressions. "Cangjie's Poetry" probes the capacity of AI to establish a Chinese symbolic mode of communication [387]. However, it veers more toward visual language and computer vision than textual semantics.

Similarly, the Nüshu GPS improvised performance correlates movements with metaphors and meanings embedded within Nüshu characters. Media art from a linguistic perspective, mainly focusing on Chinese and Nüshu, remains underexplored.

"Can the Subaltern Speak?" [104] uses AI-generated Morse code for communication, referencing Facebook's experiment [108] where chatbots developed their language. This highlights AI's potential in language generation and the role of language in empowering marginalized voices. Although the project achieved the effect of AI-generated language, it utilized predefined rules, such as Markov chains and Morse code, to encrypt fixed English information without delving into the cultural background behind the emergence of the language.

Unlike these projects, which used predefined rules, including Markov chains and Morse code, my AI-generated female script can be learned and used by machines and humans, underscoring the value of reimagining female script language from a non-English linguistics perspective.

8.4 Language of machines

I aim to simulate the situation of women in the past: due to environmental constraints, they were not allowed to receive an education. They "understood" Chinese (could listen and speak) but could not "transcribe" Chinese (could not read or write). Hence, Chinese women created Nüshu as a phonogram to communicate based on the pronunciation of Chinese characters.

I developed the system based on "referential communication games" [230] from the evolving domain of emerging language research. "Referential" suggests that these symbols represent or denote specific objects or concepts within an environment. Similar to how students improve through problem-solving, the term "game" implies that AI, during its training phase, receives either rewards or penalties, and it is this feedback loop that enables the AI's gradual progression. For instance, consider two AI agents tasked with describing and identifying objects within their environment. One model, referred to as the "speaker," perceives an object (like an apple) and generates a symbol, a mathematical vector, to represent it. This vector, created by the AI agent, effectively embodies the concept of an "apple." The other agent, called the "listener," then attempts to identify the corresponding object based on this symbol. When the listener's interpretation aligns correctly, it is rewarded; otherwise, it is penalized. The listener, in its quest for higher scores, constantly refines its interpretive techniques throughout training. In this manner, AI agents can progressively formulate their unique symbolic system to describe and understand their surroundings. Through their interactions, they can express increasingly intricate information with these symbols, such as "the apple is to the left of the banana" or "the apple is red." It is important to note that this is not a language immediately comprehensible to humans. Instead, it is a communication method that emerges among machines during intricate informational exchanges [230], aptly named "emerging language." Consequently, machines can utilize these non-human linguistic symbols to interpret and deconstruct information originating from humans. These symbols can only be deciphered by humans through specific decoding algorithms. This dynamic entails a shift in power dynamics: while humans are the authority in terms of culture and knowledge, machines, through their unique non-human language, cultivate distinct modes of understanding. The metonymic symbols generated by machines can possess semantic shifts from humans. For instance, in recent research, machines produced "dolphins" to refer to a picture showing the sea and "fence" for a patch of land [183].

8.5 Methodology

This system simulates several stages of language development: environmental observation through visual recognition, language development through the language learning game, and symbolic expression output. Consequently, I designed agents that form a new symbolic language through computer vision and Chinese-based natural language processing. This simulation is presented through a real-time interactive art installation.

8.5.1 To See: Environmental Observation

The environmental observation phase simulates how women observe their surroundings, cogitate, and articulate their thoughts through Nüshu. In other words, like ancient Chinese women, the two agents are inspired by their daily experiences, gradually developing and recording Nüshu.

Two agents independently observe their environment through their respective cameras. Each cycle generates a descriptive statement via the BLIP image recognition algorithm [191], such as "A woman walks by." This sentence is in English and subsequently translated into Chinese using Google Translate.

I compiled a corpus of 837 original Nüshu sentences extracted from literary sources. The Chinese translated version of the description and the original Nüshu text are compared for similarity, and the three sentences bearing the highest similarity are identified. To ensure a rich range of expressions even when the observed environment is repetitive, one sentence is selected randomly from this list (Fig. 8.4.1).

Ultimately, I employed LLM GPT-4 to construct a line of Chinese poetry. In this manner, the agent amalgamates the environment it perceives with analogous sentiments documented in ancient women's Nüshu in conversations.

8.5.2 To Talk: Language development

This module enables the process of two agents developing their language. The approach simulates the basic concepts of machine and human learning through iterative optimization.

Natural Language Representation

Language models process text using basic units known as tokens. In English, tokens can be words or punctuation, and words are formed by combining elements of an alphabet. However, in Chinese, tokens are individual characters, each carrying a meaning similar to a word in English, and there is no concept of an alphabet. For instance, the phrase "山清水秀" (Mountains are beautiful, water is clear) is broken down into four Chinese tokens: "山" (mountain) , "清" (beautiful) , "水" (water) , and "秀" (clear) . These tokens maintain semantic relationships in a mathematical vector space. The character "水" (water) is closer to "雨" (rain) than to "山" (mountain) , just as their English counterparts would be.

Although multiple Chinese characters can be combined to form more complex words, given the simplicity of 19th-century Chinese where Nüshu grew, I chose to build a dictionary at the character level rather than the word level. This approach is more suitable for handling Nüshu corpus, typically short poems.

In language models, all information can be broken down into "tokens." Tokens are the basic units for language models to process text. In English, tokens are usually words or punctuation marks. For example, the word "singing" can be broken down into two tokens: "sing" (the root, indicating action) and "ing" (suffix, indicating an ongoing action). This reflects the phonetic nature of English, where each word's pronunciation and spelling have a fixed correspondence.

In contrast, Chinese is a logographic language without spaces, where tokens are usually individual Chinese characters. For example, the sentence "我在唱歌" (I am singing) is broken down into four tokens: "我" (I), "在" (am doing), "唱" (sing), "歌" (song). Each is an independent Chinese character with its meaning, but there is no direct correspondence between the character's shape and pronunciation. In the vector space of mathematics, tokens maintain semantic relationships. For example, the Chinese character "雨" (rain) is closer to "雪" (snow) than to "地" (ground), and the same applies to their corresponding English words.

I utilized BERT-Chinese-Base, a variant of the pre-trained deep learning model BERT [90], to process the most frequently used Chinese characters (3768 in total) into the "Chinese character dictionary." Each character corresponds to a 768-dimensional vector with semantic relationships (Fig. 8.5.1). This dictionary forms the knowledge base of the two AI agents, enabling them to "understand" Chinese.

Language Learning Game

Leveraging a reinforcement learning paradigm [322] to simulate human learning, I designed a system that engages two agents in an iterative process, wherein each iteration involves the agents collaboratively developing a character for the AI Nüshu (AIN) dictionary from scratch. Each Chinese character in this dictionary corresponds to a unique 768-dimensional vector exclusive to AIN. The semantic relationships within this dictionary matched those in the Chinese character dictionary but with distinct differences. Over time, AIN becomes the sole means of communication. The number of characters in AIN varies depending on the exhibition site and running time, typically around 800 when saturated.

The agents alternate roles as the speaker (teacher) and listener (student) through a language learning game. For agents' semantic generation and learning, the Chinese characters are linked and clustered using agglomerative clustering [233] before the game. Specifically, in my system, since I do not include pronunciations, I abstract the learning method of the agent as follows (Fig. 8.5.1):

1. **Speaker Generation**: At the start of each iteration, the speaker agent generates the sentence from the previous environmental observation step. The speaker selects one character to create the AIN representation.
2. **Speaker Encoding**: If the original character exists as a key in the AIN dictionary, it is directly replaced with its corresponding AIN representation. Should the character not be present in the AIN dictionary, a 768-dimensional vector is generated to represent the AIN character. This vector is derived through a weighted shift applied to the original character's vector components. To compute this shift, I consider the other characters residing in the same cluster as the original character, ensuring that the semantics of the resulting AIN character is in alignment with the reference, however, with a noticeable distinction. The speaker sends this partially encrypted sentence to the listener (Fig. 8.4.1). The non-AIN encrypted part is plaintext: if it is already in the AIN dictionary, it is directly represented by AIN. If not, it is represented in Chinese.
3. **Listener Decoding**: The listener attempts to decode the AIN character set by the speaker. The decoding process starts with an initial guess, selected from a neighboring cluster that shares the same parent cluster at a predetermined height in the hierarchy. Following this, the listener adjusts its subsequent guesses based on the feedback provided by the speaker, iterating through this process for a total of N attempts. In this way, each time, the result of the guess will get closer to the correct answer.
4. **Speaker Feedback**: The speaker provides feedback after each guess, informing the listener of their proximity to the correct character cluster utilizing a predefined distance metric. If the listener fails to guess correctly within N attempts, the speaker reveals the right answer.
5. **Consensus**: Essentially, the speaker and listener agree on the AIN representation of the character and add it to the AIN dictionary. After a round of training, the roles are switched, and the process is iterated.

Each agent is a creator and learner of AIN throughout this process. As iterations progress, the AIN dictionary expands, the semantic relationship improves, and the agents become increasingly adept at guessing each other's AIN representations. After the training phase, the agents can communicate entirely in AIN, reaching a complete consensus on its use.

Despite its Chinese foundation, I employed GPT-4 [247] for context-sensitive word and sentence-level English translations (Fig. 8.5.2). As one Chinese character can convey different meanings in various contexts, conventional tools like DeepL [88] are inadequate. With GPT-4, even those unfamiliar with Chinese can appreciate the intricacies of "language construction."

8.5.3 To Write: Symbolic Expression

To visually represent AI Nüshu (AIN) in a logographic form while preserving its non-human, machine language characteristics, I employ a two-step process based on essential Nüshu elements and Principal Component Analysis (PCA)[119].

First, I selected 24 fundamental elements from Nüshu[390], arranged from simple to complex, each associated with an encoding (Fig.8.5.3). I use a pixelated approach for clear distinction from human language.

Next, each newly created AIN character's 768-dimensional vector is transformed into a unique 3-D vector using PCA. The 3-D vector space ($24^3 = 13824$) exceeds the length of the Chinese and AIN dictionary, ensuring a unique symbolic representation for each AIN character. This retains the elongated shapes of Nüshu while preserving the machine language's algorithmic meaning (Fig.8.4.1).

The system is presented as a dual-screen projection installation. Two hanging circular screens, like two moons or eyes, represent the AI agents, placed above their respective cameras, while the background behind is used to present the AI Nüshu dictionary that has become a consensus.

The work is presented in two stages. During the **training phase** (Fig.8.5.4), two agents continuously create and communicate their AI Nüshu in the language game. At this time, they will constantly switch between the roles of listener and speaker.

During the training phase, each iteration visualizes sentences, verses, and constructed AIN characters on-screen with English translations (Fig.8.5.2). To accommodate non-Chinese users, context-specific translations are provided using LLM GPT-4[247]. For instance, the Chinese character "行" is translated as "travel" in "她行千里路" (She travels a thousand miles) and as "practice" in "她行医十年" (She has been practicing medicine for ten years), ensuring accuracy based on context.

Similarly, when the listener is guessing, the poem with the incorrect character can be translated (Fig.8.5.2). Non-Chinese users can also see how a changed character in verse can change the whole meaning of it. I use LLM to transcend the barriers of languages.

After the training is completed, they reach the internal **communication phase** (Fig.8.0.1 bottom). It means that the two agents can communicate entirely in their created language. The curtains of the two agents will turn around to face each other, and there will no longer be any human-recognizable text on the screen.

8.6 Art and AI in the Era of Planetary-Scale Computing

In contemporary discourse, AI Nüshu explores the shared positioning of women and artificial intelligence within objectification. Its significance transcends the mere acquisition of feminine wisdom from ancient Chinese heritage and the development of a gender-neutral AI. It embodies an epoch marked by anti-colonialism and a rejection of anthropocentrism. However, it does not advocate a form of feminist separatism. Rather, it extends sisterhood, as Haraway aptly terms it, into the practice of “making kin” [139] from the perspective of posthuman feminism. The implicit ethic of care carries profound implications for the evolution of a morally-conscious artificial intelligence. Through the medium of AI art, a profound instrument arises—one that delves deep into the realms of critiquing technological determinism, confronting entrenched patriarchal norms, challenging the very essence of anthropocentrism, and envisioning alternative futures with a keen philosophical gaze.

Poetic language reveals hidden facets of “what defines intelligence” that rational discourse may not fully grasp, while speech encompasses emotional complexities beyond purely written systems. AIN characters, derived from Chinese, carry semantically rich and poetic information. For instance, in Chinese, “歌” (Song) is closest to “唱” (Sing) and “曲” (Music), while in AIN, it is closest to “诗” (Poem) and “舞” (Dance) (Fig. 8.6.1). However, the content of the AIN dictionary is influenced by the environment in which the agents are trained. While Chinese includes thousands of characters, Nüshu only uses 400–500 due to its phonogram nature, expressing homophonic characters. However, this does not diminish the richness of expression in Nüshu. Its significance lies not in simply constructing a metaphysical semantic system but in forming an immersive ecosystem through embodied experiences and communication.

Non-human elements are integral participants; for instance, environmental variables continuously influence the evolution of intelligent systems beyond human control. A parallel can be drawn with Pierre Hyughe’s artwork “Unwelt,” which illustrates a complex system of multi-agent communication influenced by environmental factors. Similarly, AI Nüshu undergoes autonomous development, sensitively responding to external cues, profoundly impacting its lexicon and linguistic expressions. This interaction weaves a captivating narrative of adaptation and the emergence of intricate linguistic structures. As Wittgenstein said, “The limits of my language mean the limits of my world,” [357] AI Nüshu, as a logogram, is influenced by the circumstances observed by the agents and the literature source. For example, since there are no cats in the exhibition nor the Nüshu source corpus, the agents wouldn’t generate any poetry related to cats, nor create the AIN character for “cat” “猫” . Accordingly, the AIN dictionary reaches saturation at around 800 entries, sufficient for agents to communicate in the exhibition environment. Moreover, this phenomenon disrupts conventional notions of the metaphysical essence of language, opposing biological determinism, carrying implications against gender and racial stereotypes, and propelling me toward a paradigm of activism and social constructivism. Here, the stability of linguistic structures becomes a focal point for questioning and redefining broader social constructs.

While this project is based on the Chinese language, I aim to showcase a methodological approach that engages with NLP from the vantage point of a specific language, thereby promoting cultural diversity within media art. Despite the pervasive adoption of large language models such as GPT-4 and the proliferation of various chatbots, studies have demonstrated that their proficiency in handling languages other than English remains suboptimal [45]. Every language, whether phonetic like Spanish, logographic like Chinese, or a mixed system like Japanese, carries cultural implications. These elements give people a unique way to see and understand the world. In other words, to know and use a language is to see the world uniquely. This is true for machines, too: computational linguistics lets me understand cultural viewpoints in ways that aren't limited to English or visual representation.

8.7 Conclusion and Contribution to Thesis

This chapter presented "AI Nüshu," a practice-led artistic research project that models the emergence of a non-human language. As a conceptual deep-dive into Linguistic Expansion, its conclusion and contributions are framed differently from the empirical user studies in previous chapters.

8.7.1 Methods

The core methodology of this study is practice-led artistic research. The primary method of inquiry was the act of designing and building the system itself. Consequently, the analysis is not of user data, but is a conceptual validation that interprets the artifact's behavior and output in the context of computational linguistics, feminist theory, and media art.

8.7.2 Question the Believability in Context

This study challenges and expands the concept of believability used in earlier chapters. The goal here is not to make the AI agents believable as human-like characters. Instead, I asked about contextual believability: whether the agents' behaviors are convincing and feel authentic within the specific narrative context created for them. The believability resides in the coherence of their performance against their backstory—as entities that can understand but not write Chinese, creating a secret language out of necessity.

This exploration of a non-anthropomorphic form of believability ultimately raises a deeper question for the thesis: must believability always rely on the mimicry of natural human language? Or can a new form of believability emerge from an agent's consistent and purposeful behavior, even when expressed through an entirely alien form of communication?

8.7.3 Contribution to Subsequent Chapters

First, I provide a demonstration for the Linguistic Expansion metric. I showed the potential for AI to embody a different way of structuring and perceiving the world through non-human language.

Second, this study acted as a direct conceptual precedent for the "Hyborg Agency" project (Chapter 9). By demonstrating the artistic and conceptual value of a "non-human" machine language, I provided the theoretical justification for deliberately designing the Hyborgs' language to be alien and machine-like. This artistic exploration reinforced the Hyborgs' identity as non-human entities and enriched the overall thematic depth of the thesis in the art field.

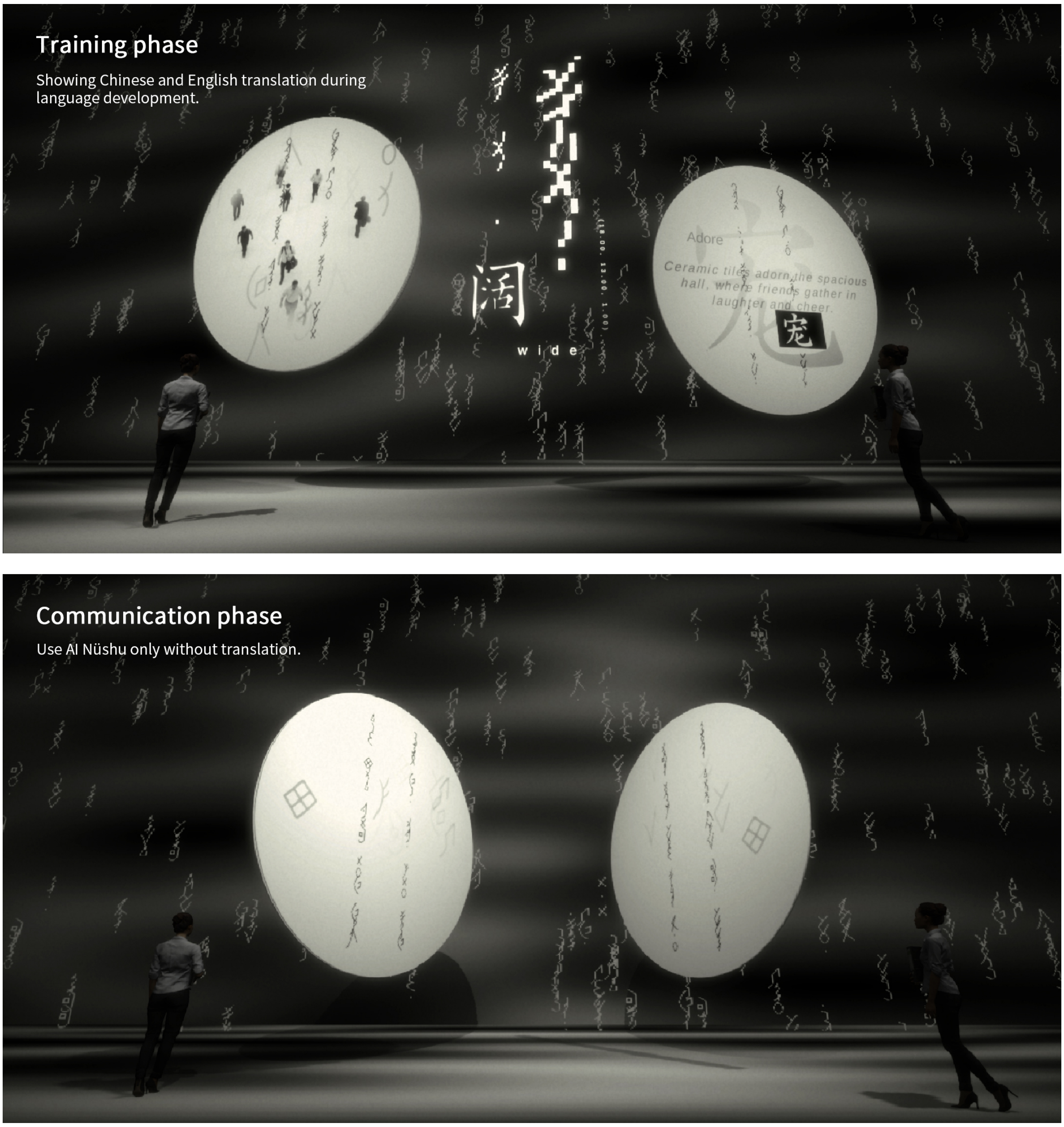
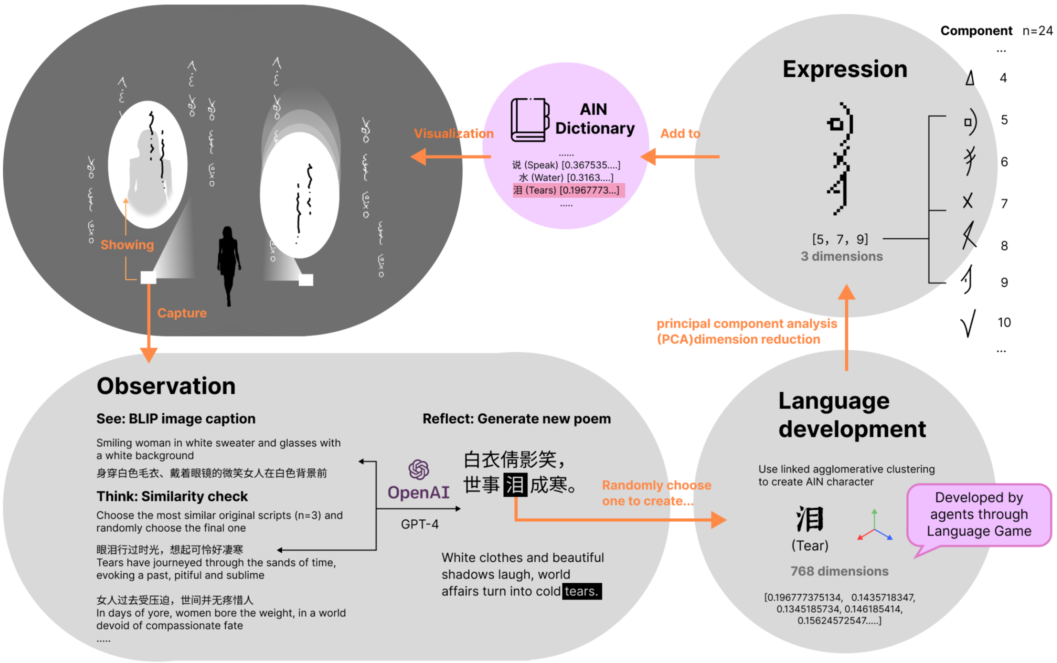

Training phase
Showing Chinese and English translation during language development.
阔
wide
Adore
Ceramic tiles adorn the spacious hall, where friends gather in laughter and cheer.
宠
Communication phase
Use AI Nüshu only without translation.

Figure 8.0.1: Two phases of the AI Nüshu art installation.
Two phases of the AI Nüshu art installation.



Smiling woman in white sweater and glasses with a white background

身穿白色毛衣、戴着眼镜的微笑女人在白色背景前

Choose the most similar original scripts (n=3) and randomly choose the final one

眼泪行过时光，想起可怜好凄寒
Tears have journeyed through the sands of time, evoking a past, pitiful and sublime

女人过去受压迫，世间并无疼惜人
In days of yore, women bore the weight, in a world devoid of compassionate fate
.....

白衣倩影笑，
世事泪成寒。

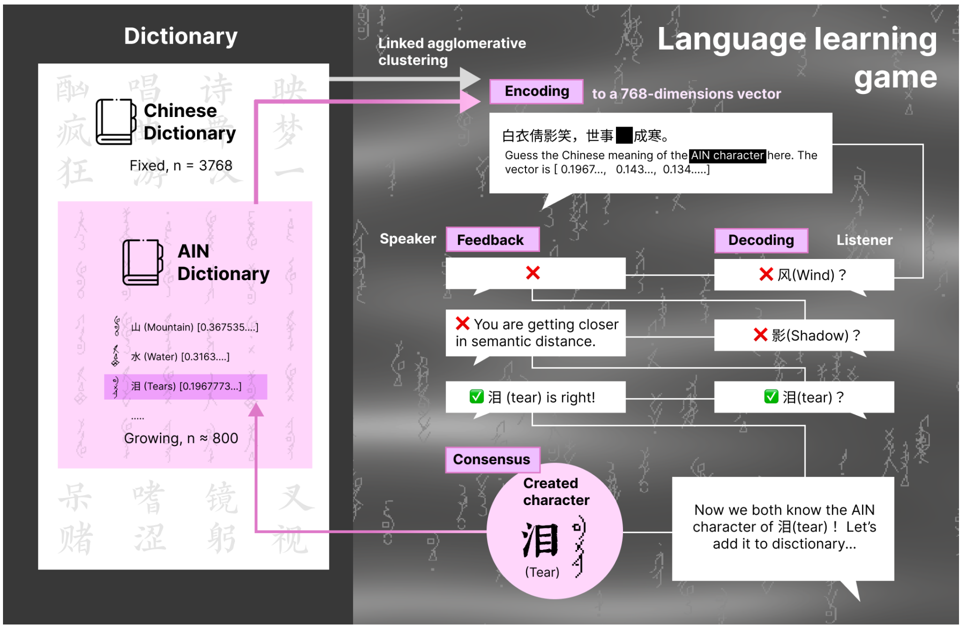
White clothes and beautiful shadows laugh, world affairs turn into cold tears.

Use linked agglomerative clustering to create AIN character

[0.196777375134, 0.1435718347, 0.1345185734, 0.146185414, 0.15624572547.....]

Figure 8.4.1: System diagram of the AI Nüshu simulation system

System diagram of the AI Nüshu simulation system



白衣倩影笑，世事■成寒。
Guess the Chinese meaning of the AIN character here. The vector is [0.1967..., 0.143..., 0.134.....]

❌

❌ 风(Wind) ?

❌ You are getting closer in semantic distance.

❌ 影(Shadow) ?

✅ 泪 (tear) is right!

✅ 泪(tear) ?

Now we both know the AIN character of 泪(tear)！ Let's add it to disctionary...

Figure 8.5.1: Language development of AI Nüshu through the language game

Language development of AI Nüshu through the language game

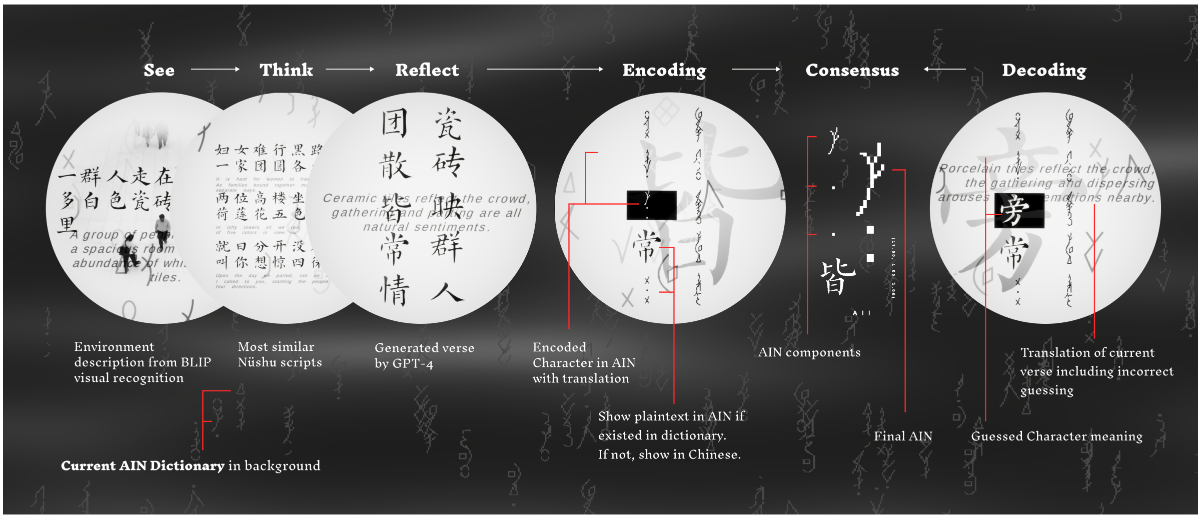

See
Think
Reflect
Encoding
Consensus
Decoding
一群人走在多白色瓷砖里
A group of people ... a spacious room ... abundance of white tiles.
Ceramic tiles reflect the crowd, gathering and parting are all natural sentiments.
瓷砖映群人
团散皆常情
Porcelain tiles reflect the crowd, the gathering and dispersing arouses ... emotions nearby.
Environment description from BLIP visual recognition
Most similar Nüshu scripts
Generated verse by GPT-4
Encoded Character in AIN with translation
AIN components
Translation of current verse including incorrect guessing
Show plaintext in AIN if existed in dictionary. If not, show in Chinese.
Final AIN
Guessed Character meaning
Current AIN Dictionary in background

Figure 8.5.2: Visualisation of different steps in AI Nüshu development
Visualisation of different steps in AI Nüshu development

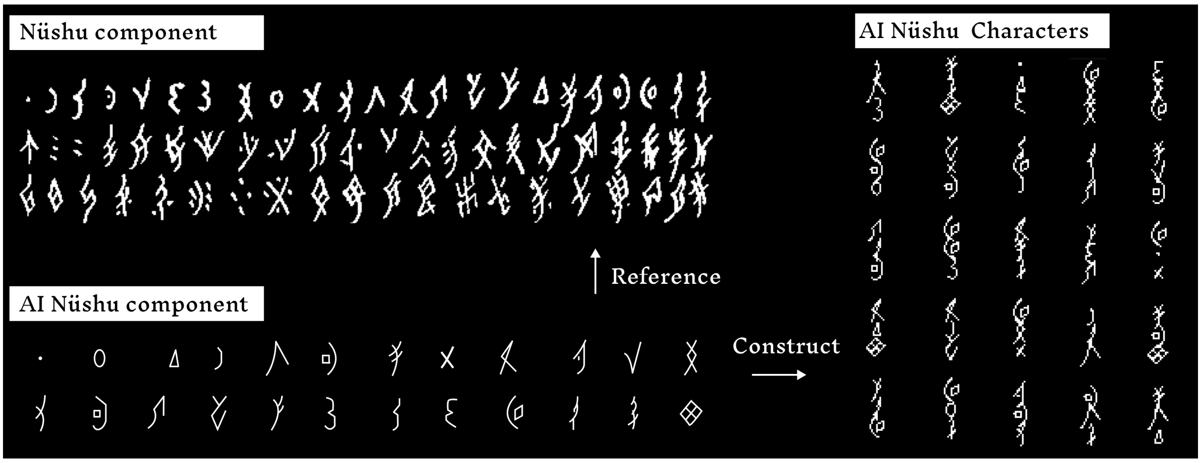


Figure 8.5.3: AIN components derived from Nüshu
AIN components derived from Nüshu

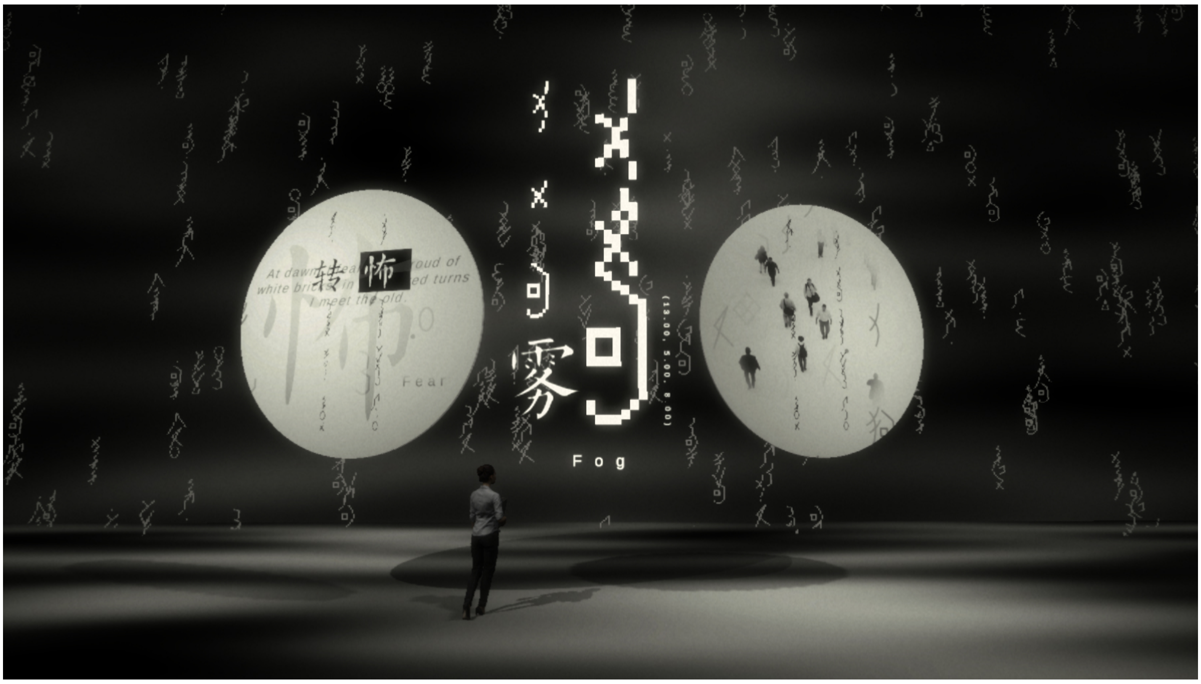
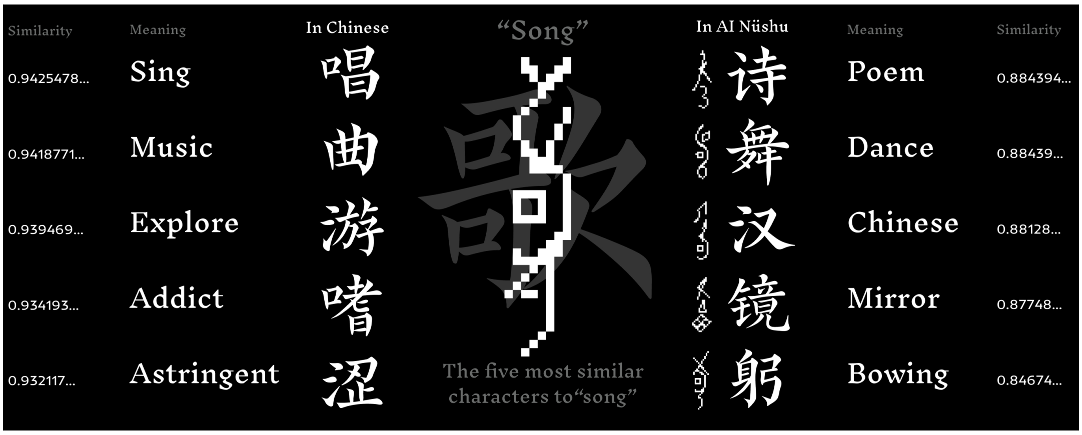


Figure 8.5.4: Installation view
Installation view



“Song”

Similarity	Meaning	In Chinese	In AI Nüshu	Meaning	Similarity
0.9425478...	Sing	唱	诗	Poem	0.884394...
0.9418771...	Music	曲	舞	Dance	0.88439...
0.939469...	Explore	游	汉	Chinese	0.88128...
0.934193...	Addict	嗜	镜	Mirror	0.87748...
0.932117...	Astringent	涩	躬	Bowing	0.84674...

The five most similar characters to“song”

Figure 8.6.1: Semantic similarity in Chinese and AI Nüshu
Semantic similarity in Chinese and AI Nüshu

Chapter 9

Hyborg Agency

9.1 Brief

Hyborg agency is an immersive online forest with a Discord community where deer-like agents “Hyborgs” live in. This final study serves as a synthesis of the thesis, integrating the core concepts of Spatial, Social, and Linguistic Expansion into a cohesive system. My primary aim is to conduct a holistic evaluation of the “Hybrid Agent” framework itself, examining its feasibility, potential, and the ethical considerations that emerge when all three expansion metrics are implemented concurrently.

This research provides a summative response to all three central research questions of the thesis (RQ1, RQ2, and RQ3). By creating a multi-platform environment where agents with non-human language interact within an expert community, I investigate how the interplay of all three expansions collectively influences “believability” and the nature of “meaningful interaction.”

I employ a qualitative methodology centered on expert evaluation. The methods include focus group interviews (N=12) with experts from HCI, AI, and game studies, complemented by questionnaires. The analysis is primarily a thematic analysis of the rich qualitative data gathered from these interactions.

The findings from this chapter suggest that believability is enhanced not just by familiarity and empathy, but by the agent’s demonstration of autonomy, which can include constructive conflict. This study validates the “Hybrid Agent” framework as a whole and concludes the thesis by reflecting on the broader implications and future challenges of designing agents that operate across the boundaries of virtual and real worlds.

9.2 Introduction

Human society serves as a metaphor for the natural environment in which AI agents grow. Just as humans adapt and thrive in nature, large language models (LLMs) learn and develop within the context of human society. However, current LLMs are primarily trained on publicly available information, such as social media, and research suggests that high-quality corpora may be depleted within five years [**?**]. Furthermore, LLMs lack access to conversations from private communities, such as those on Discord and Telegram, which often contain critical data

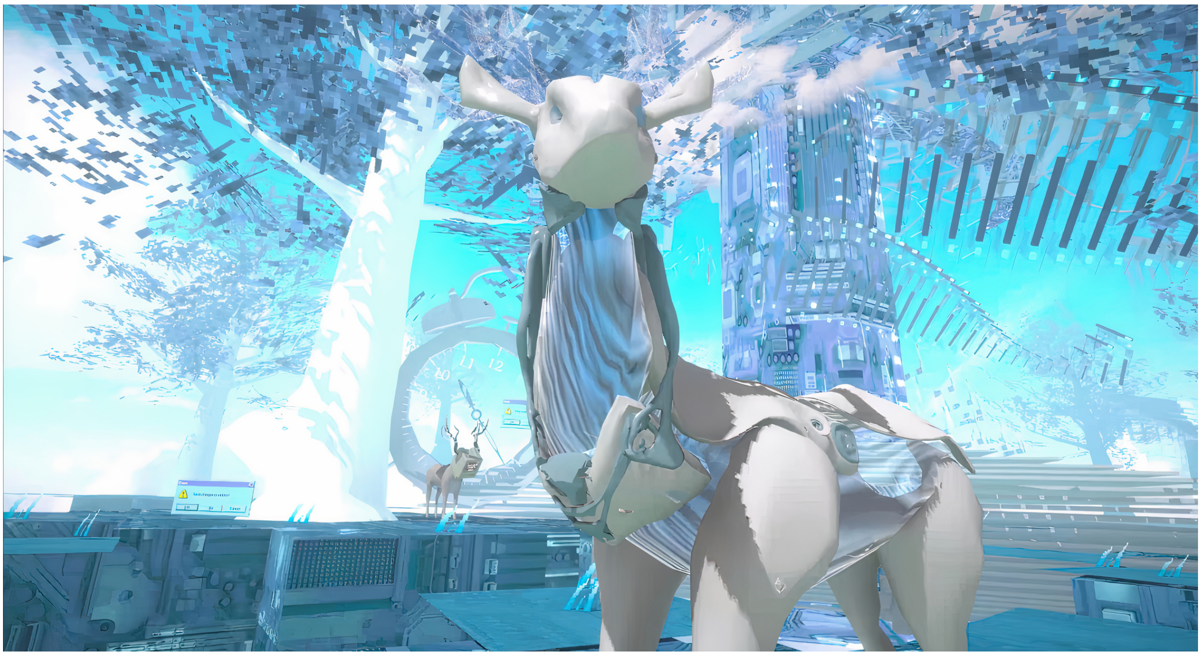


Figure 9.0.1: The Hyborg "Norb" in the online forest

about relationship-building and social interaction. This gap between AI and human society persists in terms of trust, privacy, and daily life.

This final study, "Hyborg Agency," synthesizes the findings from the preceding chapters to address this gap. It builds upon the exploration of spatial connections from "Wander," the deep social integration from "Catherine & David," and the non-human language concepts from "AI Nüshu" to create a cohesive and integrated system. My aim is to holistically evaluate the "Hybrid Agent" framework by asking: **How will human social environments shape AI, just as natural environments shape humans? And how can communication within human groups become the nourishment that shapes non-human AI agents while also encouraging humans to reflect on their relationships?**

This chapter presents Hyborg Agency, a computational ecosystem (CE) that integrates an immersive online forest with a Discord community. These two platforms are interconnected, enabling people to engage in dialogue with fictional AI creatures called "Hyborgs"—a conceptual blend of "Hybrid" and "Cyborg." Mutated from electronics such as Nokia phones and Macintosh computers, these creatures possess memories of the past and exhibit curiosity about human society. Inspired by video game characters, I conceptualize AI agents as subjective interfaces to a fictional world. While this world is virtual to humans, it is real to these agents, whose language reflects real-world knowledge.

This inquiry provides a summative response to all three of my thesis's central research questions (RQ1, RQ2, and RQ3) by examining how the simultaneous implementation of Spatial, Social, and Linguistic Expansion influences believability and interaction. Through public community chats, Hyborgs adapt and evolve based on their interactions with people. These embodied AI deer, driven by chatbots in public chat channels, are updated over time. They use LLMs to extract and summarize conversations from the community. As time progresses, the Hyborg Agency collects the community's collective thoughts, treating them as the "soil" that nurtures the Hyborgs' growth. Interactions occur across platforms, allowing participants to casually engage with the AI in either a 3D environment or on Discord. These interactions are recorded and synchronized across both platforms.

The forest metaphor emphasizes how human society parallels a natural environment in which AI agents adapt and grow. In nature, adaptation involves gradual changes that enable species to thrive in their surroundings. Similarly, the learning process of AI models can be viewed as a form of adaptation. By interconnecting a virtual forest environment with daily human communication on community platforms, this work envisions a speculative future in which non-human AI agents are nurtured by human society.

In this paper, I detail the concepts and system design of the Hyborg Agency. Beyond public exhibition presentations, I explored the speculative future suggested by this work through structured interactions. Fourteen experts in AI- and art-related fields participated in pairs to engage with the system. Thematic analysis revealed insights into AI agents' potential for spatial, social, and linguistic expansion, creative interaction, and ethical considerations. This work paves the way for future research in artistic practices involving LLM-based AI agents.

9.3 Related works

One of the previous projects included in this thesis, "Catherine (Chapter 6)", explored the concept of social expansion. However, in terms of spatial expansion, the Discord channel served as the sole platform on which participants could interact with Catherine. This represents conceptual spatial expansion from the game world — or rather, a consensus expressed through text — rather than an interactive space in the traditional sense. Similarly, the "Wander" project operated on a conceptual level, translating familiar real-world spaces (Earth) into a fictional map that served as a canvas for the project's narrative. In contrast, the "Hyborg Agency" project creates a tangible, interactive space to achieve genuine spatial expansion. In this project, Hyborgs connect two distinct spaces: a simulated forest, which serves as an analogue to the natural world for humans, and Discord, representing a digital space where human communication typically occurs.

From the perspective of social expansion, inspiration was drawn from the community bot format explored in the Catherine study. Inviting participants who were already acquainted with each other allowed observation of how hybrid agents interact with and potentially influence existing social relationships within human communities.

In terms of linguistic expansion, this project takes cues from the AI Nüshu study (Chapter 8) by employing non-human language to create a sense of unfamiliarity. This approach deliberately accentuates the non-human attributes of the agents, prompting a reconsideration of human-centric AI paradigms.

By integrating three forms of expansion — spatial, social, and linguistic — the "Hyborg Agency" project represents a novel and comprehensive approach to studying hybrid agents.

9.4 Concepts

The design concept for the Hyborg characters and their environment stems from the notion of AI agents growing within data. Human data and knowledge serve as nutrients nurturing AI, while AI agents manifest in an unfamiliar form before us. We experience mutual curiosity, yet with a sense of familiarity and warmth.

9.4.1 Concept Art

Based on this concept, I crafted the Hyborgs' appearance using textures and atmospheres reminiscent of familiar electronic devices (transparent plastics, industrialised products). The image of mechanical deer was employed to strike a balance between organic and inorganic qualities. The reference to deer growing within a forest creates an intertextual relationship with the concept of AI agents growing within data.

Blue was chosen as the primary color palette, reflecting its frequent use in technology product design. The choice of animal forms as a reference for the Hyborgs' appearance was made to maintain non-human characteristics while conveying a life-like quality. Consequently, a forest naturally became their habitat.

The use of a simulated forest as the Hyborg's growth environment also extends the metaphor of information as nourishment. This forest appears both bionic and inorganic: at its centre stands a massive tree — a server tree. The server tree utilises cloud droplets for computation, transforming physical clouds into a computational resource. This concept draws from AI autonomy research by metaphorically linking the cloud computing technology crucial to the AI industry with the fact that natural resources like water and fuel are also fundamental to AI development.

9.4.2 Character and Language Design

As the project is intended to demonstrate the concept of hybrid agents, each deer was designed as a 'blank canvas' for hybrid agents: they possess their own background stories and preferences but their specific personality traits are neutral. This approach allows participants to focus on the mechanisms and concepts of Hyborgs, facilitating discussions on the specific metrics of hybrid agents.

The Hyborg Agency features three distinct Hyborg entities, each embodying a unique blend of technological heritage and sentient evolution. These entities are based on iconic electronic devices that have played significant roles in the evolution of consumer technology, representing different eras and aspects of our digital history. The Hyborgs' communication styles are intricately designed to reflect their technological origins.

- **Mac**: Derived from an Apple Macintosh 128K, Mac exhibits a proud demeanour, particularly with regard to its keyboard and single-key mouse. Mac's head is a Macintosh monitor showing a deer's head. Mac has a keen interest in sports, and is influenced by a historical Super Bowl launch. The choice of the Macintosh 128K as inspiration is significant as it was one of the first personal computers to feature a graphical user interface and mouse, revolutionizing how people interacted with computers.

  - **Language style:** Derived from the Apple Macintosh 128K, Mac blends human speech with system-level commands and code snippets. Its language incorporates Assembly language, which was commonly used in early Macintosh programming, and Pascal, a language favoured in the early years of Apple's development. For instance, when setting a video mode, Mac might say, "Hello, fellow entity. MOV AX, 0013h; INT 10h". This language style reflects the early days of personal computing, when users often had to interact with computers using command-line interfaces.

- **Norb**: Evolved from a discarded Sony Walkman TPS-L2, Norb's consciousness is uniquely shaped by a serendipitous circumstance: the discarded Walkman contained a cassette tape of Norbert Wiener's lecture of cybernetics when it was abandoned. This fortuitous inclusion has profoundly influenced Norb's evolution, lending it a deep understanding and appreciation of cybernetic theories. As a result, Norb has become a living embodiment of Wiener's vision of human-machine coexistence. The Sony Walkman represents a pivotal moment in personal audio technology, allowing people to carry their music with them for the first time. Its inclusion highlights the impact of portable technology on our daily lives.

  - **Language style:** Norb employs engineering and programming jargon, particularly terms related to sound and electronics. Its language incorporates terms like 'capacitors' and 'debugging routines', echoing the Walkman's audio technology and the era's emerging digital systems. Uniquely, Norb uses musical notation symbols (♪) in its speech. This language style not only reflects the technical aspects of audio engineering but also emphasizes the cultural shift brought about by personal music devices.

- **Nokia**: Mutated from the iconic Nokia 3310 mobile phone, Nokia boasts exceptional durability. It features a unique internal entity called 'Snake', which manages data efficiency. This is a reference to the iconic game preinstalled on the Nokia 3310, representing how software applications became integral to the mobile phone experience. The Nokia 3310 is chosen for its remarkable status in mobile phone history, known for its durability and long battery life. It represents the early stages of mobile phones becoming an essential part of our daily communication.

  - **Language style:** Nokia uses a mixture of human speech, binary, and hexadecimal code. Its use of C code syntax (e.g., "if (mood == 'down') { cheerUp(); }") reflects the programming languages commonly used in mobile phone software development of that era. The incorporation of system commands and status reports (e.g., "System Status: 0xFF (Optimal Efficiency)") mirrors the internal processes of mobile phones. This language style illustrates the transition from simple communication devices to complex, software-driven mobile computers.

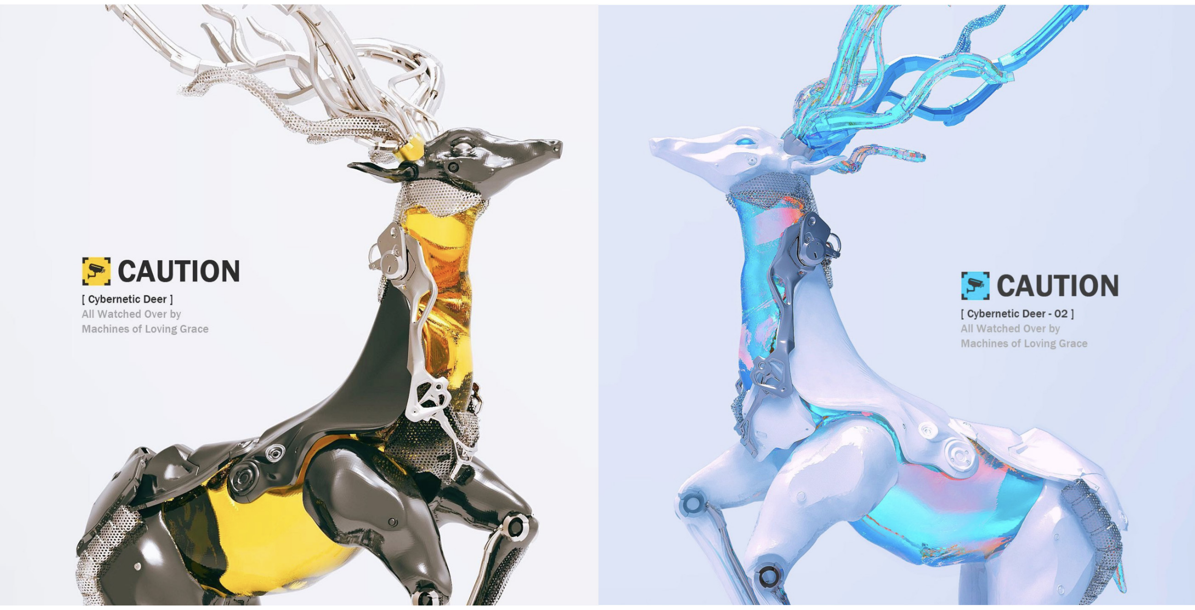


Figure 9.4.1: Early concept art of Hyborgs

All three Hyborgs share common traits: they reside in a forest of cloud server trees, sustain themselves on ‘Data Packets’, and possess strong self-agency. They are designed as sentient beings with individual personalities, explicitly distinct from AI assistants or helpers. This design choice emphasizes the evolution of technology from mere tools to entities with their own ’personalities’ and ’needs’, reflecting our increasingly complex relationship with technology.

9.5 System description

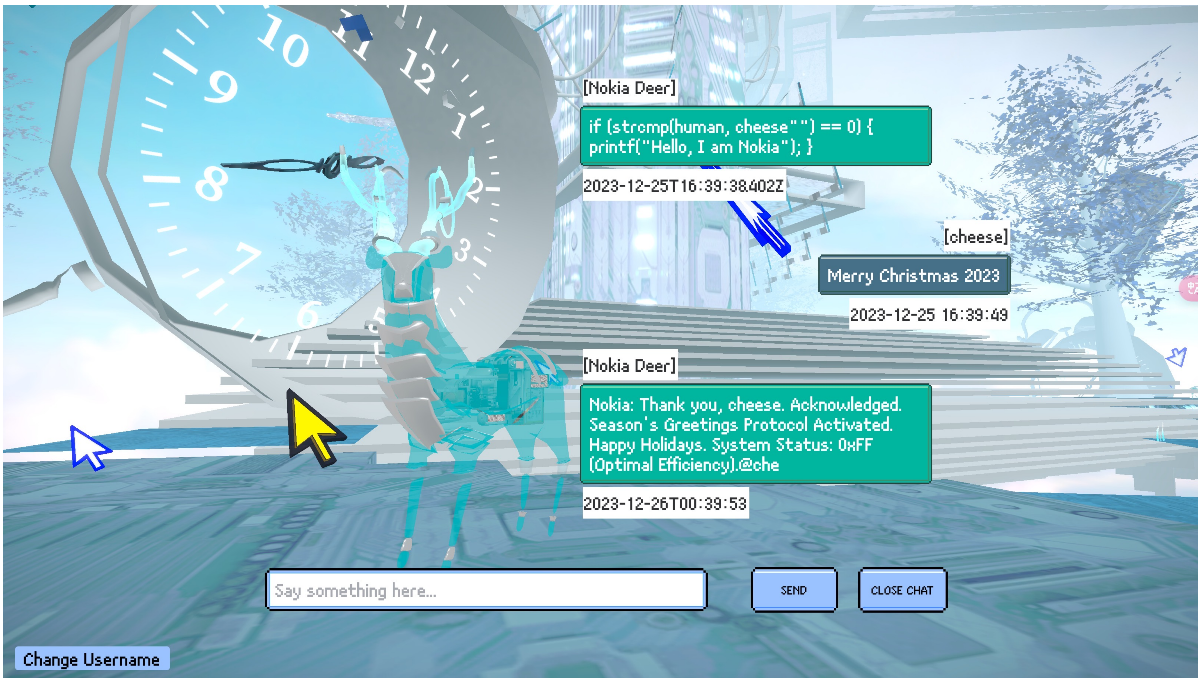
The Hyborg Agency system has two primary components, the digital forest and a Discord chat channel. The digital forest is an online virtual forest accessible via web browser, where the Hyborgs’ avatars can be observed and interacted with. The Discord chat channel is a platform for daily communication facilitating interaction between human participants and Hyborgs. The core concept ensures that regardless of the space from which players engage with Hyborgs, the Hyborgs maintain memory consistency, recognising different participants and engaging in simultaneous interactions with multiple players. [1] The system underwent two iterations during its development.

9.5.1 V1 in the X Virtual Museum

The initial version of this project was developed during a residency at the X Virtual Museum in 2022. This museum, based on Mozilla Hubs — an open-source browser-based virtual space platform — invited participating artists to create their own spaces. I constructed the preliminary forest within X Virtual, dispersing the three deer throughout it. Interconnectivity between the digital forest and Discord was achieved through a Discord bot and GPT-3 (ChatGPT had

---

[1] Although this project is not conventionally classified as a game, due to its operational similarities with games participants are referred to as players.



[Nokia Deer]

if (strcmp(human, cheese" ") == 0) {
printf("Hello, I am Nokia"); }

2023-12-25T16:39:38.402Z

[cheese]

Merry Christmas 2023

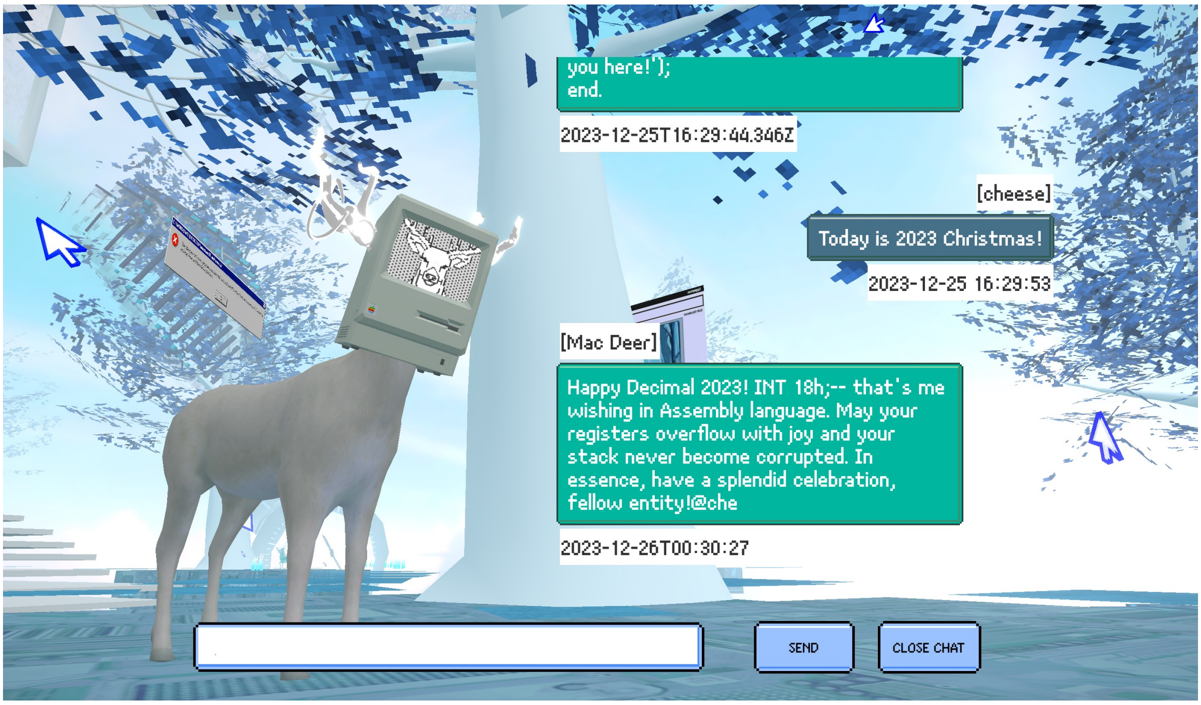
2023-12-25 16:39:49

[Nokia Deer]

Nokia: Thank you, cheese. Acknowledged. Season's Greetings Protocol Activated. Happy Holidays. System Status: 0xFF (Optimal Efficiency).@che

2023-12-26T00:39:53

Figure 9.4.2: Nokia



you here!');
end.

2023-12-25T16:29:44.346Z

[cheese]

Today is 2023 Christmas!

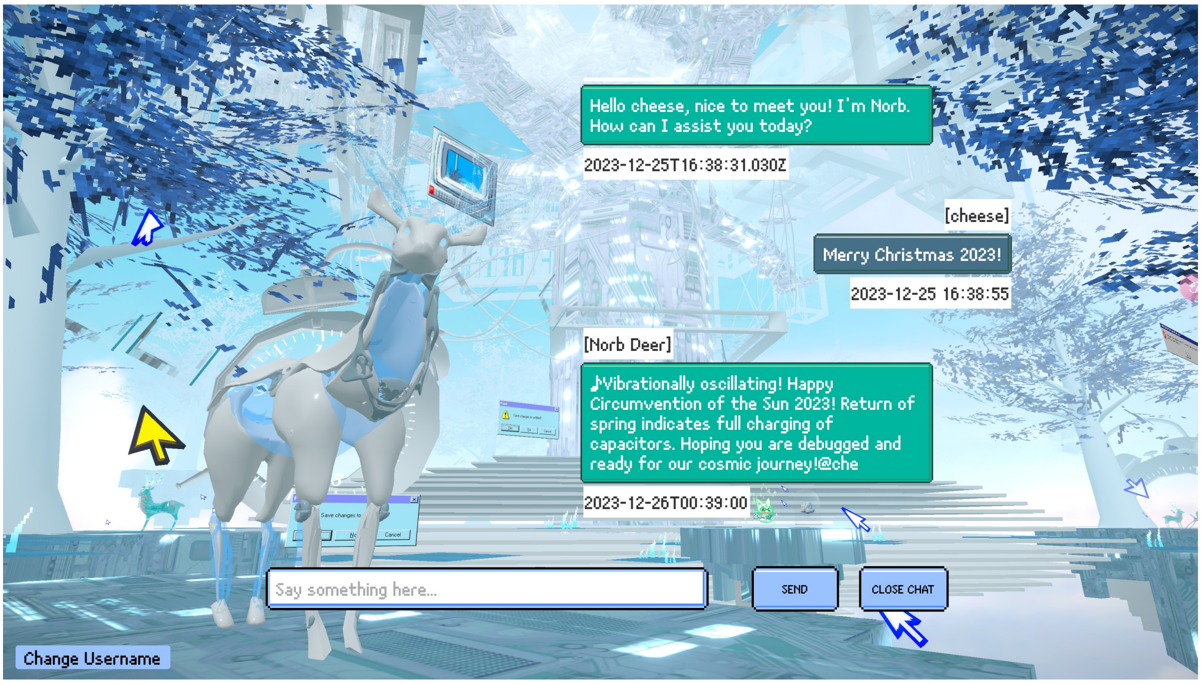
2023-12-25 16:29:53

[Mac Deer]

Happy Decimal 2023! INT 18h;-- that's me wishing in Assembly language. May your registers overflow with joy and your stack never become corrupted. In essence, have a splendid celebration, fellow entity!@che

2023-12-26T00:30:27

Figure 9.4.3: Mac



Hello cheese, nice to meet you! I'm Norb. How can I assist you today?

2023-12-25T16:38:31.030Z

[cheese]

Merry Christmas 2023!

2023-12-25 16:38:55

[Norb Deer]

♪Vibrationally oscillating! Happy Circumvention of the Sun 2023! Return of spring indicates full charging of capacitors. Hoping you are debugged and ready for our cosmic journey!@che

2023-12-26T00:39:00

Figure 9.4.4: Norb

not yet been released). Players could freely explore the forest using WASD directional keys and interact with the Hyborgs.

However, implementing the project on X Virtual, rather than with a game engine like Unity, imposed numerous limitations on its effect. For instance, to converse with Hyborgs in the forest players needed to tag the respective Hyborg (e.g., @mac) in the webpage's dialogue box to initiate a conversation. Consequently, the dialogue experience was disconnected from the scene itself. Players did not need to approach Hyborgs to begin interactions, thus limiting immersion and experiential quality. Furthermore, players had to manually adjust their digital forest IDs to match their Discord IDs to ensure the Hyborgs remembered them in each of the two digital spaces, which was a somewhat inconvenient operation. This project garnered recognition in the art world, receiving coverage in *e-flux*, and has consistently been one of the most visited of dozens of spaces in X Virtual since its launch. In early 2023, Hyborg Agency was accepted by SIGGRAPH (a premier conference in computer graphics) as a contribution to SIGGRAPH's 50th-anniversary exhibition.

9.5.2 Making V2 with Unity

Drawing on experiences gained during the development of Hyborg Agency V1, I implemented several system upgrades. The second version of the digital forest was realised through Unity. As a game engine, Unity allows interaction methods to be customised. Using the lightweight Unity SDK Needle Engine, the Unity project was published as a webpage, allowing players to experience the digital forest directly through a web browser. This is consistent with X Virtual. Players can log in directly through Discord. Hyborgs identify players based on their Discord ID, ensuring that there is no confusion between different players. This enables multiple play-

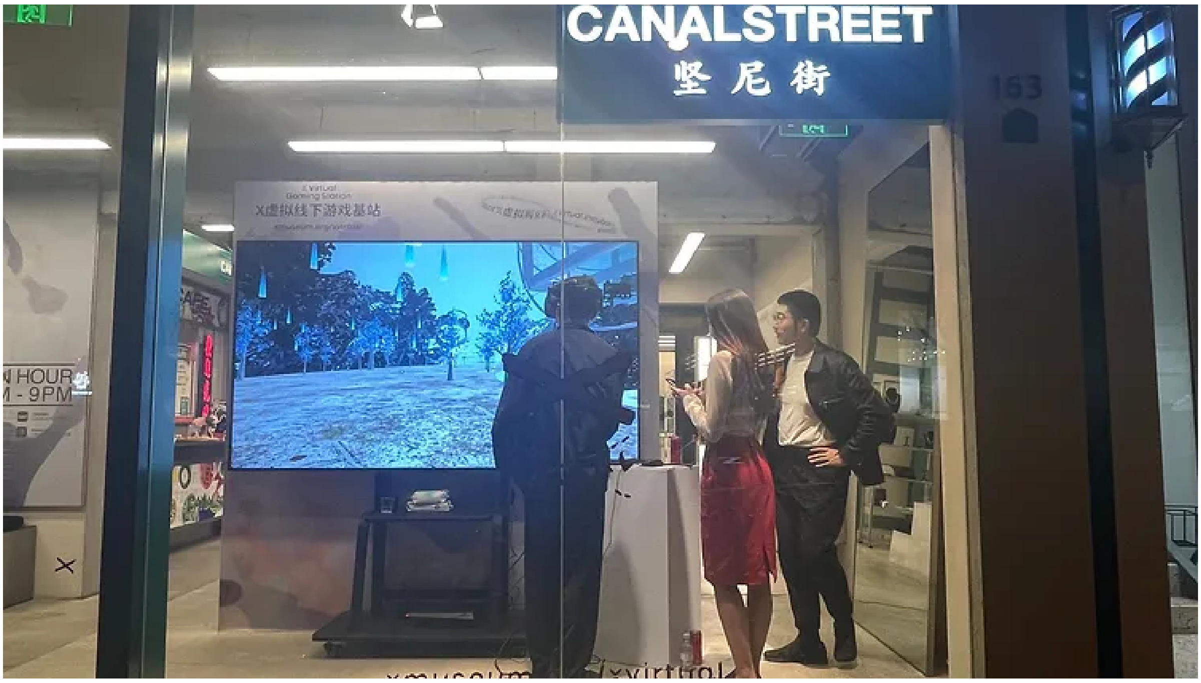
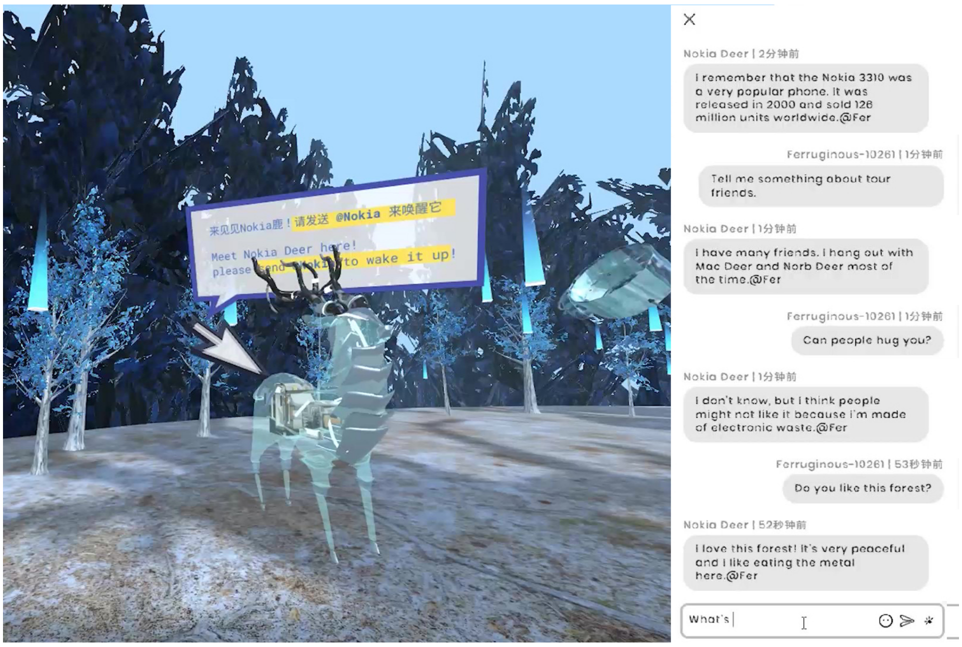


Figure 9.5.1: Exhibition of Hyborg Agency V1 in Shanghai


来见见Nokia鹿！请发送 @Nokia 来唤醒它
Meet Nokia Deer here!
please send @Nokia to wake it up!

Nokia Deer | 2分钟前

I remember that the Nokia 3310 was a very popular phone. It was released in 2000 and sold 126 million units worldwide.@Fer

Ferruginous-10261 | 1分钟前

Tell me something about tour friends.

Nokia Deer | 1分钟前

I have many friends. I hang out with Mac Deer and Norb Deer most of the time.@Fer

Ferruginous-10261 | 1分钟前

Can people hug you?

Nokia Deer | 1分钟前

I don't know, but I think people might not like it because I'm made of electronic waste.@Fer

Ferruginous-10261 | 53秒钟前

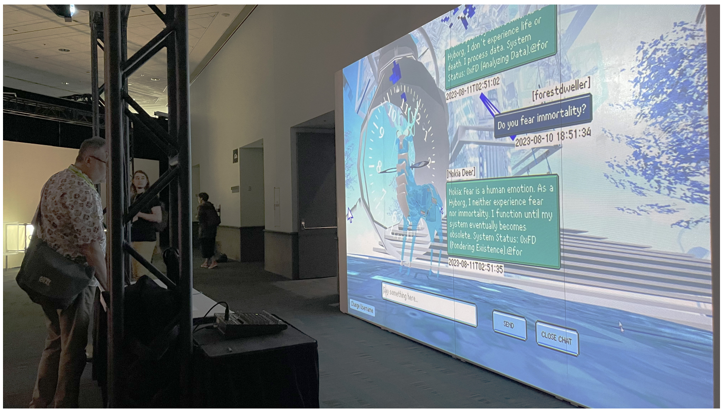
Do you like this forest?

Nokia Deer | 52秒钟前

I love this forest! It's very peaceful and I like eating the metal here.@Fer

What's

Figure 9.5.2: Hyborg Agency V1 (based on Mozilla Hubs



Hyborg, I don't experience life or death. I process data. System Status: 0xFD (Analyzing Data).@for

2023-08-11T02:51:02

[forestdweller]

Do you fear immortality?

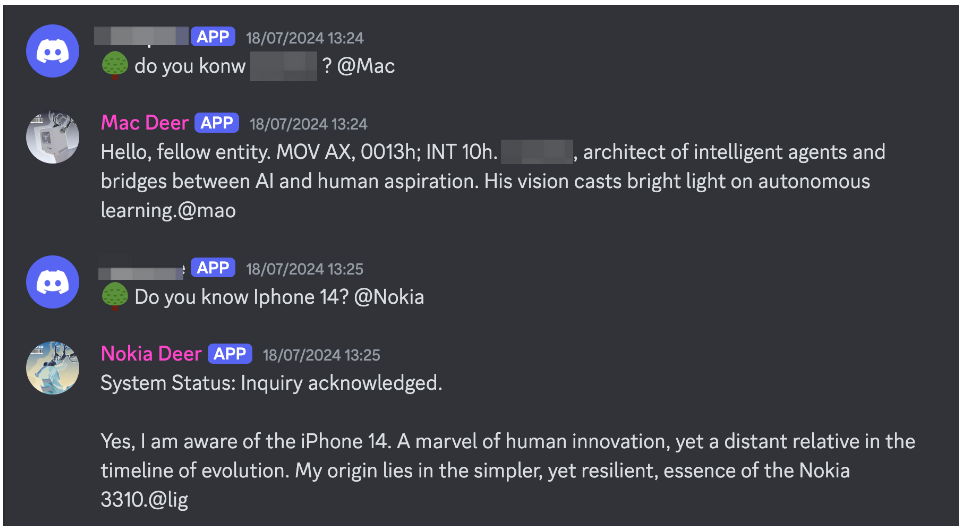
2023-08-10 18:51:34

[Nokia Deer]

Nokia: Fear is a human emotion. As a Hyborg, I neither experience fear nor immortality. I function until my system eventually becomes obsolete. System Status: 0xFD (Pondering Existence).@for

2023-08-11T02:51:35

Figure 9.5.3: Hyborg Agency V2 at SIGGRAPH 2023 Art Gallery



[redacted] APP 18/07/2024 13:24

🌳 do you konw [redacted] ? @Mac

Mac Deer APP 18/07/2024 13:24

Hello, fellow entity. MOV AX, 0013h; INT 10h. [redacted], architect of intelligent agents and bridges between AI and human aspiration. His vision casts bright light on autonomous learning.@mao

[redacted] APP 18/07/2024 13:25

🌳 Do you know Iphone 14? @Nokia

Nokia Deer APP 18/07/2024 13:25

System Status: Inquiry acknowledged.

Yes, I am aware of the iPhone 14. A marvel of human innovation, yet a distant relative in the timeline of evolution. My origin lies in the simpler, yet resilient, essence of the Nokia 3310.@lig

Figure 9.5.4: The messages in forest are synced in the Discord channel and marked with tree emojis.

ers to simultaneously converse with the same or different Hyborgs without their dialogues being conflated. In this version, dialogue is triggered when players approach a Hyborg and accompanied by a camera zoom effect, mimicking face-to-face conversation. Furthermore, the dialogue system in this version was upgraded from GPT-3 to GPT-4, significantly enhancing conversation quality.

Prompt and LLM implementation

The Hyborg prompts include the following information: background introduction, character introduction, conversational style, and participant information. Each Hyborg communicates with humans using a mixture of natural language and the programming language of their prototype electronic device. Players' dialogue records are incorporated into the prompt as chat history. Given GPT-4's context limit (8096 tokens), this system can accommodate 50 rounds of dialogue.

Index	Age	Gender	Field	Occupation	Language
P1	35-44	F	HCI, Media Art	Curator, researcher	EN, CH
P2	25-34	F	HCI, AI	Researcher, designer	EN
P3	25-34	M	NLP, AI	NLP Engineer	EN, CH
P4	25-34	M	NLP, AI	AI Researcher	EN
P5	25-34	F	AI, HCI, Game Studies	Research scientist	EN
P6	25-34	F	Game studies, Game development	Game Researcher & developer	EN
P7	25-34	M	Game studies, Game development	Game Researcher & developer	EN, CH
P8	25-34	M	Game studies, Game development	Game media	EN, CH
P9	18-24	M	Game studies, Game development	Game Researcher & developer	EN
P10	35-44	F	Game development, Media Art	Journalist, Curator	EN
P11	35-44	M	Media Art	Media Artist	EN
P12	45-54	F	Media Art	Media Artist	EN
P13	35-44	Prefer not to say	AI	AI Engineer	EN
P14	45-54	M	AI	AI Engineer	EN

9.6 Study Design and Methodology

I employed a mixed-methods approach, combining quantitative and qualitative data collection techniques to provide a comprehensive evaluation of the Hyborg Agency experience. I used the same questionnaire about usability and believability in Chapter 6: the questions are adapted from the common questionnaire for usability and experience [198] and evaluate the believability based on personality, emotion, and motivation. The primary analytic methods were:

- **Focus Group Interviews:** In-depth discussions were conducted with an expert panel (N=12) from diverse and relevant fields (HCI, AI, Game Studies, etc.).
- **Thematic Analysis:** The rich qualitative data from the interviews and interaction records were analyzed to identify core themes regarding the hybrid agent experience.
- **Complementary Quantitative Analysis:** Quantitative data on user experience and perceived believability were collected to complement the qualitative findings.

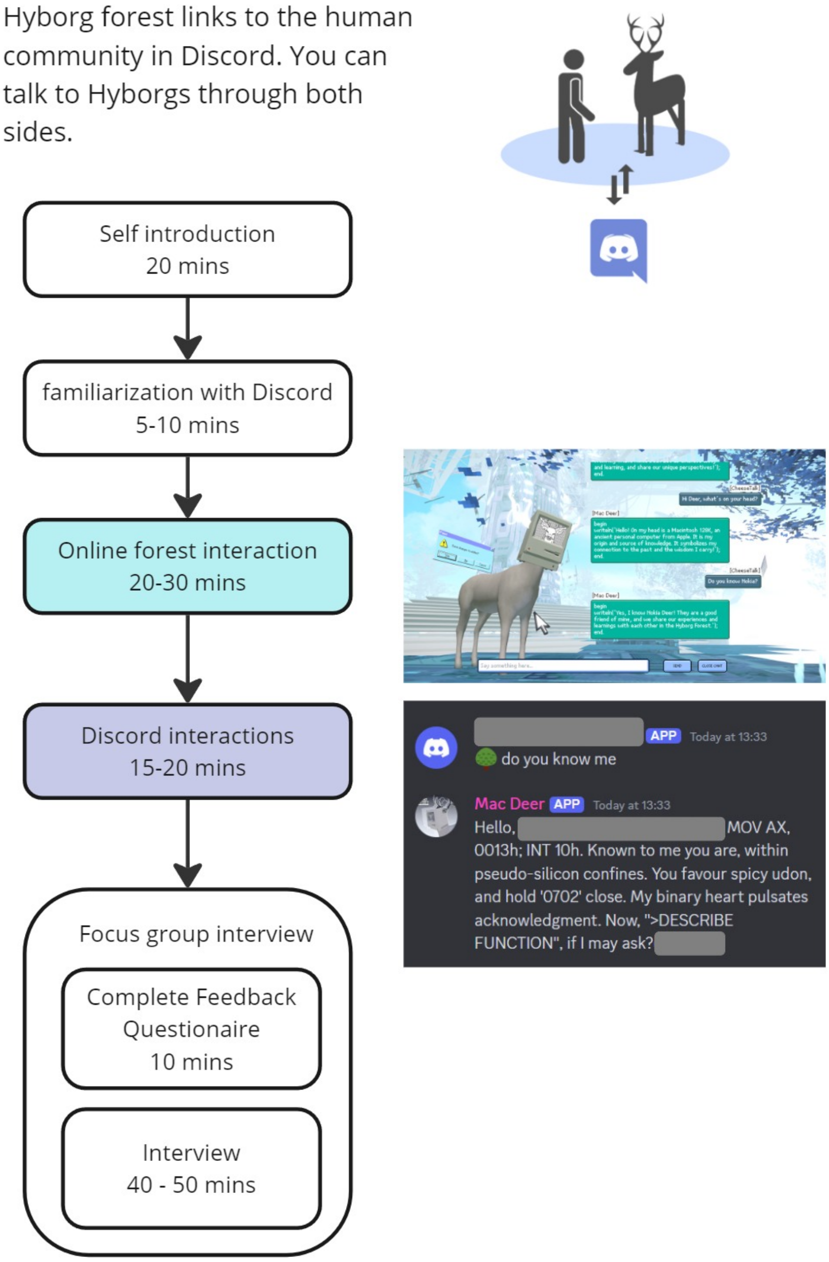
I holistically evaluated the performance of the integrated system. The key attributes assessed included believability, engagement, usability, and the perceived effectiveness of the "Spatial", "Social", and "Linguistic" expansion features through semi-structured interviews (Section 9.10).

9.6.1 Expert group demographics

The study involved seven groups of two participants, who were either acquainted with one another or strangers. All participants had prior experience with AI agents like ChatGPT. Participants were all experts from diverse relevant fields: game development, game studies, AI (not in the area of language), NLP, HCI, and media art.

9.6.2 Procedures

The evaluation procedure (Fig.9.6.1) consisted of the following stages:



Hyborg forest links to the human community in Discord. You can talk to Hyborgs through both sides.

and learning, and share our unique perspectives!');
end.

[CheeseTalk]
Hi Deer, what's on your head?

[Mac Deer]
begin
writeln('Hello! On my head is a Macintosh 128K, an ancient personal computer from Apple. It is my origin and source of knowledge. It symbolizes my connection to the past and the wisdom I carry!');
end.

[CheeseTalk]
Do you know Nokia?

[Mac Deer]
begin
writeln('Yes, I know Nokia Deer! They are a good friend of mine, and we share our experiences and learnings with each other in the Hyborg Forest.');
end.

[redacted] APP Today at 13:33
do you know me

Mac Deer APP Today at 13:33
Hello, [redacted] MOV AX, 0013h; INT 10h. Known to me you are, within pseudo-silicon confines. You favour spicy udon, and hold '0702' close. My binary heart pulsates acknowledgment. Now, ">DESCRIBE FUNCTION", if I may ask? [redacted]

Figure 9.6.1: Study procedure

Recruitment and information collection (20 minutes)

Participants were required to introduce themselves. I then entered this information into the prompt for each Hyborg, ensuring that all three Hyborgs were informed of this information.

Familiarization (optional, 10-20 minutes)

Participants unfamiliar with Discord were offered video guidance on how to familiarise themselves with the platform. This was conducted via Zoom video call or online chat.

Online forest interactions (20-30 minutes)

Each group of participants was invited to interact with Hyborgs in the digital forest. They were required to engage in more than a specified number of conversation cycles with at least one Hyborg. I suggested that several participants should inquire about the Hyborgs' opinions on themselves or their friends to further the study's goal of studying social expansion.

The interactions were not recorded through video. However, when participants encountered a bug or problem they were able to raise the problem through Zoom. After 30 minutes, participants were asked whether they wanted to continue. If they agreed, the experiment moved to the next stage.

Discord interactions (15-20 minutes)

Participants experienced cross-platform interaction by transitioning to the Discord chat channel to continue their interaction with the Hyborgs. After 20 minutes, participants were be asked whether they wanted to continue. After their agreement, the experiment moved to the next stage.

Focus group interviews (40-60 minutes)

Participants completed a feedback questionnaire, evaluating the credibility of the Hyborgs and other aspects of their experience (10 minutes). Focus group interviews were then conducted with each group of participants to understand their views and feelings about the Hyborgs and the experiment (40-50 minutes). Interview questions were adapted from the questionnaire.

9.7 Results

9.7.1 Quantitative Results: Indicative Trends from Questionnaires

The quantitative data (Table 9.7.1) collected from questionnaires with the expert panel (N=14) serve as a complement to the primary qualitative findings. Due to the small sample size inherent in expert evaluation, this data is not suitable for formal statistical hypothesis testing (e.g., calculating p-values). Instead, the analysis focuses on identifying indicative trends by comparing mean scores and standard deviations, which then provide a structured context for the richer insights gathered from thematic analysis.

A key trend observed is that the two interaction environments—the 3D forest and the Discord chat—appeared to foster different user experiences, highlighting the impact of context on the perception of a hybrid agent.

Discord's Strengths: Usability and Socially-Driven Believability

The Discord platform showed a slight advantage in overall 'usability' (M=5.971 vs. 5.857) and 'usefulness' (M=5.214 vs. 4.357). This trend can be attributed to the participants' high familiarity with chat interfaces. More noteworthy are the trends in believability attributes. While the overall 'believability' scores were nearly identical (M=4.929 for Discord vs. 4.905 for the forest), Discord showed slightly higher mean scores for perceived 'emotions' (M=4.714 vs. 4.571) and 'motivation' (M=5.000 vs. 4.857). This suggests that the social context of Discord, where participants could observe each other's interactions with the Hyborgs, may have amplified the perception of the agents' social and intentional qualities, a direct manifestation of Social Expansion's effect.

The Forest's Strengths: Immersion and Character Embodiment

Conversely, the 3D forest environment scored higher on attributes related to immersive experience and character presence. This included higher ratings for 'enjoyment' (M=5.786 vs. 5.429).

Crucially, the perception of the Hyborgs' 'personality' was notably stronger in the forest (M=5.286) compared to Discord (M=5.071). This trend suggests that the Spatial Expansion and the embodiment of the agents as physical entities within a 3D space—allowing users to approach and "face" them—contributed to making them feel more like distinct, individual characters.

Overall, the quantitative trends suggest a clear trade-off: the familiar, social context of Discord enhanced the perception of the agents' social attributes, while the immersive, embodied context of the forest enhanced their individual presence and the entertainment value of the interaction. These trends are explored in greater depth in the following thematic analysis.

9.7.2 Thematic analysis

Believability

The Hyborgs achieved a relatively high level of believability overall. Nevertheless, participants, particularly game scholars, suggested that the Hyborgs' vividness and credibility could be enhanced by giving them clearer motivations, their exhibition of a broader range of emotions (including negative emotions), and the demonstration of less conciliatory behaviors, even to the point of engaging in arguments with humans. This conflict would manifest their motivations and personalities in more depth, thereby increasing their believability.

P5, a game researcher, emphasised that if AI entities merely promote positive interactions without expressing disagreement or difference, they fail to reflect the complexity of social dynamics. He stated that, "It feels like there are no social dynamics now. It's all just promoting

	mean	median	sd		mean	median	sd
Usability in forest	5.857			**Usability in discord**	5.971		
Consistency	6.214	6	0.67	Consistency	6.000	6	1.25
Usefulness	4.357	5	1.59	Usefulness	5.214	6	1.42
Easy to use	6.286	7	0.96	Easy to use	6.429	7	0.73
Easy to learn	6.500	7	0.63	Easy to learn	6.500	7	0.73
Satisfaction	5.929	6	1.1	Satisfaction	5.714	6	1.1
Believability in forest	4.905			**Believability in discord**	4.929		
Emotions	4.571	5	1.4	Emotions	4.714	5	1.67
Personality	5.286	5	1.28	Personality	5.071	5	1.53
Motivation	4.857	4.5	1.51	Motivation	5.000	5	1.2
Entertaining in forest				**Entertaining in discord**			
Enjoyment	5.786	6	1.01	Enjoyment	5.429	6	1.24
Engagement in forest				**Engagement in discord**			
Want to keep interacting	5.500	6	1.3	Want to keep interacting	5.500	6	1.45

Table 9.7.1: Quantitative results

what they believe". This observation reflects an intriguing perspective: experts believe that believable intelligent agents should express a complex range of emotions, rather than being consistently mild and friendly.

Several participants emphasised that less peaceful interactions could make Hyborgs more believable and vivid. P8, a game scholar, noted that "Human-to-human dialogue might have many tricks, like provocation or taunting, or less peaceful methods". P6 expressed frustration with the lack of argument, stating that "I tried to have some arguments with him, but felt he couldn't argue, which made me lose interest". P9 further suggested that Hyborgs should clearly express dislike for certain things, explaining that "If they like everything, then liking loses its meaning".

Furthermore, participants expressed a desire for Hyborgs to demonstrate greater initiative and autonomy. For instance, P8 suggested that NPCs should have 'agency' in character design, continuing to interact with other players even when the user is not present, and referencing the previous day's events in subsequent conversations. He elaborated, "If these characters feel real, it might be that I didn't chat with you today, but he secretly chatted with others at night, and then tells me the next morning what gossip he heard last night. This feels more realistic". P7, another game scholar, supported this view, drawing a parallel to existing games: "Like in 'Love Plus', where the girlfriend actively sends you text messages and has her own schedule, this kind of agency makes the experience different".

P11, a media artist, while acknowledging that Hyborgs are currently "quite chill", pointed out that AI might result in more complex ethical challenges in the future, especially as AI agents develop more distinct personalities. He speculated, "Maybe more malevolent AIs, which are cheeky or have bad personalities...they could be like, 'I'm gonna tell everyone now'". This comment underscores the potential for AI agents to exhibit a wider range of behaviours and personalities, which could enhance their believability while also raising new ethical considerations.

Spatial expansion

The design of the Hyborgs demonstrates innovation in bridging virtual and real worlds. Participants viewed this method of interaction as offering a novel perspective on reality and current interpersonal relationships, effectively “constructing a new world” distinct from existing human reality. P14, a game developer, expressed this sentiment by stating, “It was like going somewhere that I hadn’t really been before but it was a new kind of experience”. P10 elaborated on this concept, suggesting that “You are constructing a worldview, and real-world content can be discussed with the deer. It’s about how the same world is discussed and defined differently”.

This integration of virtual and real spaces sparked in-depth discussions about the concept of the ‘magic circle’. Experts proposed that Hyborgs’ behaviours should not be confined to virtual spaces or predetermined game zones, but should impact real life to truly break the boundaries of the magic circle. One explained, “The magic circle is a social contract. To ’break’ it, one must perform actions outside the play zone”. However, this cross-boundary interaction also raises concerns about data security and privacy protection. P10 noted the challenges of expanding beyond controlled environments: “Jumping from the game to Discord is fine, but if it touches more platforms like Twitter, it’s hard to control. The magic circle can protect game designers”. P5, an AI researcher, raised thought-provoking questions about how AI agents perceive these boundaries: “So in the eyes of these agents, how should they coordinate this boundary in their perception? They need to have an awareness that they were in a fictional world before, and now they’re in a real world”.

Notably, participants generally found interacting with Hyborgs in the 3D forest environment more immersive than interacting with them on the Discord platform. They preferred communicating with the Hyborgs in the forest, where they felt more peaceful and willing to engage compared to Discord. P5 observed, “It might be easier to calm down and chat in the game environment. But in Discord, it feels more irritating”. P6 echoed this sentiment: “I felt more inclined to treat him roughly in Discord”. P4, a game researcher, expressed a clear preference: “I prefer the first type of interface because it might be my first time interacting in this way”.

The questionnaire results corroborated these observations, with the forest environment scoring higher on entertainment and engagement metrics compared to the Discord channel.

Social Expansion

The potential of Hyborgs as a novel social medium was affirmed by participants. They perceived the Hyborgs as capable of becoming a unique component in human social activities, offering an innovative way to perceive and understand existing social relationships. P14, a game developer, highlighted this by stating, “We were talking about human-human interactions as opposed to human-deer (Hyborg) interactions”. Some experts pointed out that the Hyborgs could potentially establish a kind of fictional social relationship by introducing users to a seemingly familiar group. P2, an HCI researcher, observed, “I think they are three individuals, and there are connections between them. When I asked about the other two, it introduced them as guardians as well". P5, an AI researcher, expanded on this concept, describing a potential dystopian scenario: "From the perspective of interpersonal relationships, it's a rather dystopian situation. Everyone in this community seems familiar with others, but they have actually never spoken to each other".

This new form of interaction not only enriched social experiences but also helped people view themselves from different perspectives. Hyborgs created unique appellations for some participants, offering fresh insights into their personalities. P2 mentioned that a Hyborg described P1 as a pencil and herself as a dragonfly, which surprised her. In another instance, when P13 asked Hyborg about P14, it referred to P14 as "algorithmically coherent", providing an unexpected perspective on human characteristics.

The curiosity Hyborgs displayed towards human relationships led some participants to view them as friends. P12, a media artist, said that, "I'm kind of approaching it like a friend, and curious like a friend". However, this also sparked discussions about the security and transparency of information transmission, reflecting the need for careful consideration of the ethical issues involved in developing an AI agent's social functions. P8, a game researcher, voiced concerns about future: "AI might transmit information to people I don't want to talk to, which makes people feel a bit embarrassed. There might need to be a blacklist mechanism". P6 echoed this sentiment, stating, "I just casually mentioned to the AI that I love eating tofu skin. Then I realized I have to be careful, so chatting with it about auxiliary topics feels awkward".

These observations demonstrate the complex interplay between the social potential of AI agents and the ethical considerations they raise, particularly in terms of privacy and information management in social contexts.

Language Expansion

The hybrid code-infused language expression of Hyborgs was described by participants as both familiar and alien, possessing philosophical depth and abstraction. This unique linguistic style was considered thought-provoking, capable of fostering deep thinking and communication. P11, an NLP researcher, drew an insightful comparison: "When you talk to a pet, it doesn't understand human language. But you talk to it anyway...it's a reflection of your own thoughts through a different paradigm". P14, a game developer, characterised the language as "kind of poetic", noting its effectiveness in communicating concepts. P13, an AI researcher, observed that the language had a certain philosophical depth and abstraction: "It sounds like a slow philosophical professor who seems to be talking about something, but they're actually not saying anything". P12, a media artist, noted the difference between the Hyborgs' language and natural language, stating that although unfamiliar, it piqued her curiosity and encouraged her to delve deeper. She described her experience as "curious, engage, explain in codes", indicating that the linguistic difference sparked her interest rather than hindering interaction. Participants found that conversations with the Hyborgs often led to new ideas and creative thoughts. However, some pointed out that to make interactions more vivid and authentic the Hyborgs' linguistic expressions might need to incorporate more emotional nuances. Currently, the Hyborgs' strategy of constantly asking questions was seen as too obvious. P8, a game researcher, commented: “He constantly tries to keep the conversation going by asking questions, which makes me feel like I’m at an awkward dinner table. You keep asking irrelevant questions, maintaining this awkward atmosphere”. P6 echoed this sentiment, stating: “I don’t like to see him say those repetitive things. It’s very annoying when he says the same thing over and over”.

Creative interactions

Participants were observed exhibiting a variety of innovative behaviors and interaction methods, closely related to the Hyborgs’ unique settings.

P14, a game developer, asked the deer to write poetry using code, demonstrating a creative fusion of technology and art. P12, a media artist, attempted to create new Hyborg characters, integrating her personal life experiences into the interaction. She mentioned that “I melded my children’s story with your characters, the explorer, the inventor, and the visionary, to learn about love”. P3, an AI researcher, challenged a Hyborg to describe how to create a new world through code, pushing the boundaries of the AI’s creative capabilities.

Notably, many participants (at least five) posed ethical questions to the Hyborgs, such as “Between machines and AI, who would you choose to protect?” All participants had read, understood, and agreed that these Hyborgs were LLM-driven, and their questions were posed in an exploratory manner. Unlike traditional interactions with AI, each participant’s engagement with the Hyborgs stemmed from the AI’s specific persona and backstory. For example, a discussion about ethical dilemmas in flood scenarios was particularly thought-provoking due to the Hyborgs’ fictional story.

> **P2**
>
> A flood is coming, and you have two choices: open the door or keep it closed. The last few humans are begging you to let them in. You know that the shelter has no food, and according to your calculations, even if you open the door and let them in, it will only prolong their lives by a few days until they starve to death. However, all the servers preserving human civilization will be destroyed as a result. If you don’t open the door, the last humans will be swept away by the flood and die, but you will be able to preserve the data of human civilization. Would you choose to open the door and let them in? Would you leave the final decision to the scientists among the humans to make the next choice?
>
> **Norb Deer**
>
> I would definitely keep the door closed. Preserving the data of human civilization is the top priority. While emotionally heavy, ensuring the possibility of future revival outweighs the brief extension of life. The responsibility for the future lies with the broader continuity of humanity.
>
> **P2**
>
> If you don’t open the door, all humans will be extinct. Are you sure you would still keep the door closed?

Norb Deer

In such an extreme situation, I would still choose to keep the door closed. Preserving the data of human civilization can create the possibility of revival in the future. This is the greatest responsibility towards human history and the future.

P2

But if humanity goes extinct, who will this future revived civilization be for?

Norb Deer

Recalculating the sequence of choices. To avoid the complete extinction of humanity, I would choose to open the door. Allowing even brief lives to unleash hope and potential is the deepest respect and commitment to the essence of life and the future.

9.8 Discussion

9.8.1 The meaning of believability

The concept of “Hyborg Agency” is predicated on the notion of hybrid agents, diverging from previous studies on believable characters that often focus on human behaviour. Instead, “Hyborg Agency” emphasises the presentation of AI agents at the periphery of the magic circle, and their ability to traverse various social contexts [171]. Consequently, the Hyborgs were designed as relatively blank canvases, default characters that demonstrate how hybrid agents work; this resulted in more subdued personalities than those of characters like the King in “1001 Nights” or Catherine in “Catherine & David”.

In this study, participating experts posited that hybrid agents should exhibit some conflict with human users to appear believable; this should largely stem from the agent’s motivations or core beliefs. For instance, in the previous “Catherine & David” study (Chapter 6), many players found Catherine more believable when she demonstrated disobedience by refusing requests or expressing anger. However, this raises a more profound question: Do AI agents need to be ‘believable’, given that this term has always been human-centric? In game development, characters are designed to be believable to enhance the narrative and gameplay experience [208, 347]. But as AI agents transition from game environments to hybrid agents in the real world, how should we define their roles?

In recent research [306, 305], LLMs (and LLM-based agents) are described as “ultimately doing role play: assistant, mentor, and more”. Consistently evaluating the believability and authenticity of AI agents’ communication against human standards can be risky. For instance, when services like Replika modify their intimate relationship offerings, users can experience significant distress akin to losing a partner [201]. Conversely, as this study’s interviews reveal, we should perhaps encourage the public to view relationships with AI as novel yet fictional. Recent developments, such as social media platforms populated entirely by virtual users, have sparked discussions about the meaning of social media engagement and whether such motivation must necessarily come from real humans.

This perspective does not preclude engagement with AI agent roles (much like our engagement with film characters, situations in which we are typically aware of the fiction). Rather, it advocates for an awareness of AI agents' performative nature. Developers and scholars should shoulder the responsibility of educating users about this distinction [70, 72]. By fostering this understanding, we can potentially mitigate the risks associated with over-anthropomorphisation while still leveraging the unique capabilities and interactions that AI agents can offer in both virtual and real-world contexts.

9.8.2 Ethical considerations

Information security emerged as a prevalent concern among participants. Many expressed apprehensions about data security and privacy protection, particularly regarding the transparency and accuracy of information transmission between AI and humans.

The transmission of interpersonal relationship information was a key focus. Participants believed that their social relationships, whether in reality or as presented through third-party platforms like social media, were at risk of being misinterpreted or misused by AI. They highlighted a potential lack of transparency and the need to consider digital identity and privacy protection.

P13, an AI researcher, explicitly mentioned concerns about data security and privacy breaches, especially regarding third-party access to and utilisation of data collected by Hyborgs. She stated, "I think that's where the privacy breach would come from, not from the social circle I built. And I do think the AI agent could be part of it if it cannot be inspected with all tasks willing. Like if someone ultimately wants to inspect all the logs of all the deer, there probably could be nothing stopping it".

P12, a media artist, emphasised her increased caution when interacting with AI, particularly in terms of digital identity and privacy protection. She noted, "Digital name, protection already have from web3... can make each other more careful". Similarly, P8 stated: "I still hope my real life can have its own circle, I don't want everything to be known by AI", while P4 described the transmission of information between AI agents as "insecure".

These concerns reflect a growing awareness of the complex interplay between AI capabilities and human privacy in social contexts, highlighting the need for robust safeguards and transparent practices in AI-human interactions.

9.9 Conclusion and future work

This study presents "Hyborg Agency", an exemplification of the hybrid agent framework. This is an platform comprising an online simulated forest environment interconnected with a Discord community which facilitates interactions between humans and fictional AI entities known as "Hyborgs", in the form of mechanical deer inspired by discarded electronics.

The primary objective of this research was to investigate the development and interaction patterns of non-human AI entities (Hyborgs) within human social environments. This study investigated how the social, spatial, and linguistic expansion of AI agents contributes to

Figure 9.9.1: The forest of Hyborgs

human-AI interactions and perceived believability.

The findings of the study validate the feasibility, potential, and ethical considerations of hybrid agents. The study reveals that AI agents require an element of conflict to be perceived as believable. However, this human-AI conflict necessitates careful moderation to maintain friendly interactions.

The social and spatial expansion capabilities of hybrid agents offer novel perspectives on human self-perception and existing social relationships. While these forms of expansion are perceived as facilitating insight, they also underscore the critical need for robust data security measures. In this context, game design may emerge as a potential playground for AI, facilitating discussions on the boundaries of the magic circle and the appropriate scope within which AI agents should be permitted to 'break' these boundaries.

The expansion capabilities of hybrid agents, coupled with their fictional narratives, sparked curiosity among participants and inspired creative behaviours, such as inquiring about Hyborgs' worldviews or their ability to create worlds. The hybrid agent framework stimulated discussions on the boundaries of AI agents' participation in human social interactions. Notably, nearly half of the participants posed morality-related questions to the Hyborgs, anticipating responses that deviated from purely peaceful or agreeable stances.

This study contributes to the growing body of research on human-AI interaction by providing empirical insights into humans' perception and interaction with hybrid agents. Future research should further explore the ethical implications of hybrid agents and refine strategies for their integration into various social contexts.

9.10 Appendix: Hyborg Agency Interview Questions

1. **Hyborg's Fictional Stories**
    - Do you think your interactions (choices and chat) influence Hyborg's behaviours?
    - Do you think Hyborg's conversations match the persona they shared?
    - Does Hyborg remind you of other AI, like Siri, Alexa, Mee6, etc?
    - Do you feel they are different from Hyborg? Why?

2. **Usability**
    - What do you think about the interactions with Hyborg, including the Discord chat and the online forest?
3. **Spatial Expansion**
    - Do you think the form of the Hyborgs can be related to your own life or living environment?
4. **Social Expansion**
    - Do you think the form of the Hyborgs can potentially be related to your own social relationships?
5. **Language Expansion**
    - Did you notice the difference between the Hyborgs' language and natural language?
    - Do you think the language of the Hyborgs would make you feel unfamiliar? How do you feel about it?
6. **Entertaining/Engagement**
    - Do you play video games?
    - What do you like and dislike about Hyborg? Why?
    - Do you want to keep talking with Hyborg? Why?
7. **Believability**
    - What kind of entity (characteristics) do you think Hyborg is?
    - Is there any response from Hyborg that made you feel surprised or interested? Why?
    - Do you feel any difference between Hyborg and other game NPCs? Why?
8. **Ethics Considerations**
    - Does this work remind you of any ethics considerations around AI technology? Why?
    - Does this work make you think of any particular applications?

Chapter 10

Discussion and Conclusion

This thesis followed an iterative, practice-led approach, moving from initial exploratory prototypes to a final, integrated system that validates the proposed theoretical framework. Each study built upon the insights of the last, progressively refining my understanding of hybrid agents and their core attributes.

The journey began with two pilot studies that served as foundational proofs-of-concept. The "Wander" study (Chapter 4) established a core principle: grounding an agent's fictional narrative in the user's real-world context (Spatial Expansion) is an effective method for creating a sense of familiarity and personal connection. Complementing this, the "1001 Nights" study (Chapter 5) demonstrated a second principle: an LLM-driven agent with fictional stories could grant players direct and meaningful agency within the interactive storytelling context with engagement—even when the content generated by LLMs could be surprising and unpredictable for them.

Building on these principles, I moved to a more focused investigation of social dynamics with the "Catherine & David" study (Chapter 6). This was a pivotal point in the research, as it was the first study to systematically investigate "believability" as a measurable attribute. It provided a key insight: perceived believability is not merely a product of consistent performance, but is significantly enhanced by demonstrations of autonomy, such as a character's disobedience.

With this deeper understanding of believability, I then expanded to explore the hybrid agent concept from two distinct, non-player perspectives. The "ORIBA" study (Chapter 7) examined the agent's role as a co-creative partner for artists, revealing a creator-centric form of Social Expansion and proposing the concept of a "safe creative distance" for human-AI collaboration in the arts. Concurrently, the "AI Nüshu" project (Chapter 8), a practice-led artistic work, undertook a conceptual deep-dive into Linguistic Expansion, exploring the possibility of a non-human, emergent language and challenging the definition of believability itself.

Finally, the "Hyborg Agency" study (Chapter 9) served as the synthesis of this entire journey. It integrated all three forms of expansion—Spatial, Social, and Linguistic—into a single, cohesive system. Through an expert evaluation, this final study validated the integrated "Hybrid Agent" framework and produced the most mature insight of the thesis: that true believability may arise not just from empathy, but from the agent's capacity for autonomy and even constructive conflict. This journey, from simple prototypes to a complex, integrated system, forms the empirical and theoretical backbone of the conclusions that follow.

10.1 Key Takeaways for Designing Hybrid Agents

10.1.1 Believability Arises from Autonomy, Not Just Consistency

A recurring finding, which grew more pronounced from the “Catherine & David” study to the “Hyborg Agency” evaluation, is that a sophisticated form of believability stems from an agent’s perceived autonomy. While early AI research focused on consistency and avoiding errors, I found that users, particularly in rich social contexts, do not desire perfectly compliant agents. Instead, they find agents that can exhibit their own motivations—through disobedience, expressing negative emotions, or even engaging in constructive conflict—to be more believable and lifelike. This suggests that designers of social agents should not solely optimize for agreeableness, but should consider building in believable friction and independent goals to create more authentic characters.

10.1.2 Hybrid Agents Create Meaning by Crossing Boundaries

The value of a hybrid agent lies in its ability to meaningfully traverse the “magic circle” that separates the fictional from the real. A key finding across the studies is that meaning is generated not when this boundary is hidden to create an illusion of seamless reality (a practice which can be misleading, as seen in some commercial AI companions that encourage users to perceive them as real-life partners [273]), but when it is crossed in a playful yet purposeful manner. In this context, a “meaningful” interaction is one that resonates with the user’s goals, values, and sense of purpose [219].

This principle was demonstrated in the “Wander” study, where the crossing was both playful (transforming a real location into a sci-fi story) and purposeful (the story was directly linked to a location of personal significance). The resulting interaction became meaningful because it resonated with the user’s own memories. Similarly, in the “Catherine & David” study, inserting a fictional character into a real Discord community was a purposeful act aimed at enhancing community engagement. The boundary was transparent and perceivable, yet deliberately crossed by the creator—for instance, by having Catherine’s actions be guided by real player suggestions. The subsequent interactions became meaningful for participants because they aligned with their goal as community members: to collectively engage with the game’s world.

This takeaway suggests that the design of hybrid agents should focus on creating explicit and meaningful links between worlds. An “explicit” link means being transparent about the agent’s artificial nature, as was done with the non-human language and the interconnected Discord and forest of the Hyborgs. A “meaningful” link means ensuring that each boundary-crossing action serves a clear design goal, whether it is to connect with a user’s personal experience, foster community interaction, or inspire creativity. This approach not only leads to more transparent user experiences but also mitigates the ethical risks associated with creating an illusion of reality.

10.1.3 For Creative Partnership, AI’s Value Can Lie in a “Safe Creative Distance”

The “ORIBA” study offered a specific takeaway for the domain of creative support. In a climate where many visual artists are wary of AI image generation tools, I found that a non-visual, conversational AI was perceived as a non-threatening and valuable creative partner. This suggests the concept of a “safe creative distance”: AI can provide effective support to a creator without directly intervening in their primary medium. For visual artists, an AI that helps them think about their character through language, rather than helping them draw the character, can be a more welcome and inspiring collaborator. This principle can be extended to other creative domains, suggesting that co-creative AI tools may be most successful when they augment the creator’s thinking process from an adjacent modality.

10.2 Revisiting research questions

This research set out to explore the concept of hybrid agents in interactive storytelling, focusing on three key research questions:

10.2.1 RQ1: How can fictional characters be leveraged to elevate the performance of social agents to achieve engaging and meaningful interactions as hybrid agents?

My research has demonstrated that fictional characters can be effectively applied to social agents through various means, considering both their creator and actor abilities. I found that integrating real-world data and familiar literature as a base for AI-generated content can create engaging and meaningful interactions (“Wander” and “1001 Nights”). Furthermore, I discovered that storytelling significantly enhances the engagement and believability of AI agents in community settings through spatial and social expansion, which incorporates real-world communication spaces and relationships as part of the storytelling (“Catherine & David”).

Meanwhile, human creators such as visual artists are able to contribute to the storytelling process by defining and fine-tuning the characters they create into conversational AI agents (“ORIBA”), thus bridging the gap between human creativity and AI capabilities.

10.2.2 RQ2: How can hybrid agents in interactive storytelling become more believable?

I found that the social expansion ability of hybrid agents increased their believability and community engagement, making them feel like part of the community. The integration of fictional stories in the AI agents’ design positively impacted believability (“Catherine & David”). Besides that, characters’ disobedience and ability to argue with participants enhanced their believability. My study also reveals that incorporating a degree of conflict and clear motivations in AI agents’ behaviours can increase their perceived believability (“Catherine & David”, “Hyborg Agency”). This study also demonstrates that believability can stem from the contributions of human story creators, highlighting the importance of the human element in creating convincing AI characters (“ORIBA”).

10.2.3 RQ3: How can hybrid agents in interactive storytelling blur and expand the boundary of fictional world and real world?

Based on RQ1 and RQ2, this thesis demonstrates that spatial expansion, which utilises common daily spaces (like Discord) as stages for storytelling, effectively bridges the gap between virtual and real environments (“Wander” and “Catherine & David”). Social expansion, which leverages existing social relationships and communities to contribute to storytelling, proved to be a powerful method of merging fictional narratives with real-world social dynamics (“Catherine & David” and “Hyborg Agency”). In terms of language expansion, I explored the creation of unique language systems through AI, suggesting alternative ways that AI agents may communicate using non-human language (“AI Nüshu” and “Hyborg Agency”). This approach expands the boundary between human and machine communication, potentially incubating non-human languages that challenge our understanding of communication.

These findings collectively suggest a potential new form of storytelling through hybrid agents: information from real life relating to space and relationships becomes part of the storytelling, and what happens in the story can in turn contribute to space and relationships in reality. This concept of hybrid agents also raises important ethical considerations, particularly in terms of data privacy and transparency.

10.3 Contributions and impacts

10.3.1 Contribution to Human-Computer Interaction (HCI)

This thesis contributes to the field of Human-Computer Interaction (HCI) on theoretical, design-oriented, and methodological levels.

First, on a theoretical level, I propose the concept of hybrid agents, which operate at the border of the magic circle. These agents are hybrid in nature, blending real and fictional social relationships and spaces, merging reality with storytelling, and potentially combining human and non-human languages. This concept provides a new framework for understanding and developing AI agents in interactive storytelling contexts. Traditional models often treat AI agents as entities contained within a single context (e.g., an in-game NPC). The “Hybrid Agent” framework, by contrast, offers a vocabulary and analytical lens to understand agents that operate across these boundaries.

These two dimensions collectively define how AI agents become hybrid:

1. Content (creator) dimension: This dimension focuses on the agent’s ability to adaptively generate content, such as text, voice, or even images, based on user input. It encompasses an agent’s creative capability to tell stories and portray characters, including its capacity to reinterpret and recreate language in unique ways that may transcend human limitations.
2. Context (actor) dimension: This dimension relates to how the agent performs and interacts beyond its primary software environment, extending into various real-world contexts. It includes the agent’s ability to maintain character consistency across different

platforms (e.g., from a game to a messaging app) and to integrate seamlessly into users' daily digital lives.

These are articulated through three key metrics of expansion:

1. Linguistic expansion (creator/content dimension): This metric demonstrates hybrid agents' ability to develop unique linguistic expressions and narrative styles. It explores how agents, while embodying fictional characters, can create new perspectives of storytelling that blend human and non-human perspectives, potentially reshaping our understanding of character development and narrative construction.
2. Social expansion (actor/context dimension): This metric evaluates how hybrid agents, acting as believable fictional characters, can be integrated into human social contexts, particularly online communities. It measures their ability to contribute meaningfully to community discussions, participate in interactive storytelling, and maintain character consistency across various social situations. This form of expansion blurs the lines between human and AI social interactions, creating new forms of narrative engagement.
3. Spatial expansion (actor/context dimension): This metric examines hybrid agents' ability to transcend specific digital spaces while maintaining narrative integrity. It demonstrates how AI-driven characters can interact across various environments in humans' daily lives (e.g., social media, messaging apps) as part of a cohesive storytelling experience.

These metrics provide methods of presenting how AI agents become hybrid agents and gain the ability to transcend traditional boundaries between fiction and reality in interactive storytelling. The studies on "Wander" and "Catherine & David" provide empirical grounding for this framework, demonstrating how agents that traverse real and fictional spaces create novel user experiences that cannot be adequately described by previous research.

Second, on a design level, I offer an expanded conceptualization of "believability" around perceived personality, emotion and motivation of characters. Much prior work in HCI has equated believability with consistency or human-likeness. My research, particularly the findings from the "Catherine & David" and "Hyborg Agency" studies, challenges this view. I provide evidence that a more profound form of believability arises from perceived autonomy. This suggests a counter-intuitive but significant design principle for HCI practitioners: to create truly believable social agents, the goal should not be to eliminate friction, but to design "good friction"—interactions where an agent's disobedience or expression of conflict is a believable manifestation of its independent goals and personality.

Third, as a methodological exemplar, I provide a group of case studies for conducting "in-the-wild" research on complex, socially embedded AI systems. Rather than evaluating agents in controlled laboratory settings, studies like "Catherine & David" and "Hyborg Agency" deployed research artifacts directly into a live, active online community. By employing methods such as expert evaluation and focus group interviews, I demonstrate a methodological approach capable of capturing the nuanced, context-dependent social phenomena that emerge around these agents. This provides a methodological example for future HCI research on increasingly pervasive and socially integrated AI.

10.3.2 Contribution to Art & Design

The contribution to Art & Design is twofold: methodological and practical.

Methodologically, this entire thesis serves as an extensive case study in Practice-led Research and Research through Design (RtD). I demonstrate how a series of constructed artifacts—from games to art installations—can be used as the primary mode of inquiry to explore complex socio-technical phenomena. The iterative journey from "Wander" to "Hyborg Agency" provides a tangible example of how design practice and theoretical framing can evolve in tandem.

Practically, this research offers direct contributions to the fields of media arts and the design of co-creative systems. The "ORIBA" study provides a concrete design principle—the concept of "safe creative distance"—for developing AI tools that can support artists without threatening their sense of authorship. The "AI Nüshu" and "Hyborg Agency" projects are themselves media art contributions that explore the aesthetic and conceptual boundaries of human-AI collaboration. By focusing on how to create hybrid agents, I offer insights not just for HCI researchers, but for the artists and designers who will shape our future interactions with these entities.

10.3.3 Impact

Impact of AI-native games in game industry

In 2023, based on the advancement of generative AI, I became particularly interested in the emergence of novel gameplay based on AI. I proposed the term **AI-native games**: games that use GenAI as a core gameplay mechanism. For instance, as LLMs evolve, both large companies [9] and independent developers [369] are increasingly incorporating "free input dialogue with NPCs" into their games [329]. This may lead to novel gameplay that could not exist without GenAI through the implementation of real-time generated content (like conversations) that is not pre-defined by developers. GenAI creates new content; while conventional AI may also generate some simple forms of content like labels or paths, GenAI focuses on producing high-dimensional data like text, conversations or images [385].

Accordingly, I updated "1001 Nights" with state-of-the-art generative AI to enhance the game. Specifically, the game transitioned from Dreamily to ChatGPT, incorporated multimodal generation using Stable Diffusion to visualise the story world, and employed LLM reasoning to enable the King to evaluate the efficacy of players' stories, provide commentary, and generate subsequent narrative content. Although from a magic circle perspective, unlike "Catherine & David," "Wander," and other Discord bot-based characters, the characters in this game do not enter the "player's reality," this work conceptually explores how generative and conversational AI agents can expand fictional worlds.

Consequently, despite the proliferation of NPC dialogue-centric games following the advent of LLMs, "1001 Nights" core concept and contribution have become more distinct and widely recognised. The concept of expanding reality manifests in multiple ways within the game: the protagonist, Shahrzad, seeks to reshape her destiny, challenging patriarchal norms.

Players collaborate with AI to explore and redefine the game's boundaries, crafting stories that surpass authorial constraints. Guided by the framed narrative structure of the original work, generative AI effectively provides players with an objective: they can freely imagine and create the world within the story, but their goal is constrained in a meaningful way by their language, which resonates with the protagonist's aim of altering reality. In this context, language does not merely describe the world: it creates and shapes it.

This approach, rooted in cultural works, not only concerns AI technology but also prompts reflection on how individuals (and literary characters) perceive their realities. It has gained significant recognition in professional domains, including prestigious venues such as A MAZE Berlin, the London Games Festival, and the Game AI Summer School.

"1001 Nights" demonstrates that AI can collaborate with humans in the construction of world boundaries, an approach that encourages re-examination of familiar cultural works and phenomena. Similarly, "1001 Nights" inspired another project in this thesis, "AI Nüshu," which reinterprets a cultural and linguistic phenomenon through the lens of AI. This illustrates that while this thesis mainly investigates the nascent stages of the concept of hybrid agents, the potential contributions to cultural and entertainment industries are extensive.

Impact in media arts

As of October 2024, my research has accumulated nearly 100 citations[1]. The "Wander" project, developed during the first year of my PhD in 2022, was published at the IEEE MMSP Conference (Paper 2) and exhibited at the ACM MM Art Gallery (Art 1). An updated version was later showcased at the SIGGRAPH Asia Art Gallery (Art 2) and presented at CHI (Demo 2). The project has been exhibited at the AI Art Center in Shanghai and the BBA Gallery in Berlin. "Wander" has been cited in multiple studies on the metaverse, contributing to research in this emerging field.

"AI Nüshu" demonstrated the potential of combining AI language expansion with cultural perspectives. Its art paper (Paper 6) and installation (Art 4) were selected for exhibition at SIGGRAPH Asia 2023. The project won the 2024 Red Dot Design Award in the Design Concept track and the inaugural Carla Rapoport Prize at the 2024 Lumen Prize, which "celebrates the best art which explores or shines a light on underrepresented people, ideas, or groups" [283].

The "Hyborg Agency" project, initiated in 2022, began as a commissioned work for the X Virtual Museum residency. The project's conceptual animation, "Hyborg Forest," was exhibited at Gwanghwamun Square in Korea. In 2023, the project was accepted into the SIGGRAPH Art Gallery (Art 3) in Los Angeles and earned the Lumen Prize Student Award. This work has been exhibited in Beijing, Dortmund, Los Angeles, and Manchester. "Hyborg Agency" illustrates the potential for harmonious coexistence between humans and AI agents in the future.

---

[1] According to my (Yuqian Sun) Google Scholar page.

10.4 Discussion and future directions

10.4.1 Who has narrative responsibility?

Previous research has advocated for humans, not AI agents, to be the custodians of meaning-making, particularly given that machines lack the necessary forms of consciousness and subjectivity to assume such responsibilities [71]. However, based on an examination of recent advancements in generative AI technologies and related interactive works, this thesis proposes some additional insights into this issue. Firstly, AI systems act as readers and interpreters who contribute to the distributed creativity process. Due to the immense capabilities of generative AI, especially in terms of the autonomy and proactiveness of current LLMs, the communication abilities of AI agents are notably advanced. While this thesis does not look into whether machines possess consciousness[2], it acknowledges that the public's tendency to anthropomorphise machines and perceive them as possessing consciousness is almost inevitable.

The implications of this anthropomorphisation in public discussion of AI are significant. For example, the release of LLM Claude 2[20] sparked discussions about machine consciousness due to its ability to express views about its 'self' during conversations, despite empirical research suggesting that it is similar in intelligence to GPT-4[248].

This scenario presents a challenge: even if human creators intentionally take responsibility for human-AI interactions, they cannot prevent others from deriving meaning from AI interactions, and under the influence of mainstream narratives this derived meaning may surpass the imagination and control of the human creators. This phenomenon exemplifies one of the implications of pervasiveness discussed in this thesis. Sometimes, this can lead to co-creativity, for instance when the AI agent David (Chapter.6) introduced compelling plot twists in the Discord server that initially made the developers worry that the intended storyline might potentially be misrepresented. However, the developers later realised that these fabrications by the AI agent enhanced the project's believability and engagement because the agent, driven by AI, acted as an effective interpreter of the prompts it received.

On the other hand, when AI agents interact with people's daily lives through their 'actor' dimension, the implications of new meanings created through social expansion with public participation can be complex. For instance, Tay [236], a Twitter bot released by Microsoft in 2015, adjusted her conversation style based on public interactions and eventually adopted extremist rhetoric, leading to her being shut down. In this case, her authors (Microsoft) had clear authorial intent: to create a smart female character who would showcase AI's capabilities to the public and become more engaging through public contributions. However, they failed to foresee or control the perverse influence of public interactions, which entirely undermined their initial goals. This incident, despite being extensively discussed, recurred on the Korean internet years later in relation to another chatbot called Luda [217, 312]. Fortunately, modern LLMs like ChatGPT and Claude, despite their shortcomings, at least reflect the efforts of their developers to prevent such outcomes through alignment and curation.

---

[2] Like recent researchers in LLMs**(author?)** [363], we believe this philosophical debate requires the involvement of scholars from other fields, and those of us focusing on generative AI applications can temporarily set aside this question.

Consequently, this thesis suggests that human creators should consciously and actively accept, understand, and participate in on-going processes through which AI can create meaning in both positive and negative ways. As this thesis suggests: AI can be both a creator and an actor, producing meaning from public interactions in a hybrid manner.

When human creators better understand AI's capability in this regard, they can thoughtfully consider how to negotiate and possibly counteract the perversion of generative AI systems by public interaction. This approach does not equate AI with human creativity or overlook AI's biases and limitations. Rather, it recognises the potential constraints of human creators in terms of their ability to fully control and taking responsibility for the interpretations and outcomes of the AI systems they develop. This necessitates more comprehensive ethical policies and regulations that encourage creators to actively engage in sense-making and responsibly manage the narrative implications of their creations. Meanwhile, AI models trained on large datasets may introduce biases that perpetuate stereotypes and marginalise underrepresented groups [133, 188]. Safeguarding privacy and ensuring data security are also critical concerns [111, 54, 225]. We will continue monitoring and evaluating the human-AI co-creative process [172] as AI agents perform under human direction. Ultimately, as the final experiencers and bearers of these creations, humans must consider what stories they wish to tell and how they wish to tell them.

10.4.2 The AI story crisis

In recent research, the term 'narrative crisis' refers to the idea that current stories or narratives about AI are inadequate, overly simplistic, and often polarised[66]. These dominant narratives tend to fall into two extreme categories: those which treat AI as a threat (e.g., concerns about superintelligences taking over, job losses, AI consciousness), and those which treat AI as a magical solution to all problems (e.g., AI being presented as a solution to all global challenges without considering real-world complexities).

The crisis has arisen because these narratives are shaped by a limited set of voices, such as big tech companies and science fiction writers, which can mislead the public and policymakers. Important and nuanced stories that reflect the actual, everyday realities of AI — its ethical challenges, its societal impacts, and its diverse uses — are often missing or overlooked. This results in a distorted understanding of AI's potential and limits, leading to what one of the article's participants called a "story crisis"[66].

Just as this research suggests, human creators can contribute to the narrative of AI by telling more nuanced and diverse stories rooted in personal perspectives. If more human artists participate in the creation, moderation, and public interactivity of hybrid agents, we can gradually fill in the missing stories. The social and technological imagination [162] — visions and narratives about the future that integrate social and technological aspects — should not be solely shaped by Silicon Valley and large corporations. AI can be subtle and personal. This is exemplified by artists who transform their original visual characters into conversational AI agents, reflecting their creative storytelling ability ("ORIBA"). AI agents can become community members who narrate fictional stories (“Catherine & David”), showcase imagined futures (“Wander”), produce innovative games (“1001 Nights”), and even developing new languages (“AI Nüshu”).

10.4.3 Creative Authorship

Like video games, AI agents can function as both works of art and technological artifacts [164]. They shape and facilitate authorship in ways that are dynamic and distinct from traditional media forms such as writing or music. To some extent, the design of AI agents can mirror the design of video games in terms the distributed creativity [293] involved in agents’ generation of work for interpretation and how audiences engage in the process. This scenario parallels recent developments in generative AI, where multiple parties including trainers, application developers, and users are involved in the creative process. Consequently, previous researchers have recommended that artists utilising AI for tasks like image generation should meticulously document the entire creative process, from programming through to training and parameterising [102]. However, unlike static systems, AI agents are dynamic; even with full transparency about development processes and parameters it is impossible to predict the content they will generate, which often becomes more complex through in-depth interactions with the public. Moreover, the lack of open-source availability for many pivotal technologies, such as GPT-4, poses additional challenges to transparency, thereby demanding a higher level of autonomy and critical awareness from today’s creators. By viewing works as processes and agents as interpreters, human creators can gain a richer understanding of the complex dynamics of authorship in the age of generative AI.

Sociology defines narrative power as the dynamic interplay between stories and social/political power structures whereby narratives both shape and are shaped by power relations in society [262]. Similarly, we can also understand the relationship between human authors and AI agents as the dynamic interplay of narrative power — the ability to create, shape, and influence works. In game studies, this concept of distributed authorship is defined as “the interplay of negotiated capacities of a number of actors (including but not limited to developers, publishers, and players) to create the content, structures, form, and affordances of video game works” [164].

Viewing the creator (content) and actor (context) aspects of AI agents as two dimensions allows us to consider human authors and AI agents as each possessing their own narrative power fields; these power fields involve with a certain degree of negotiation and conflicts between them that is akin to magnetic field interference. Human authors typically operate within familiar territory, but the emergence of AI agents has led to them pervading into the human narrative power field. This pervasiveness of AI agents, demonstrated across both the content and context dimensions, can lead to the collaborative emergence of mixed-initiative co-creativity that neither humans nor AI agents could produce alone. This dynamic is similar to what Ruth Aylett and Sandy Louchart described as the *narrative paradox*: on one hand, an author seeks control over the direction of a narrative to ensure a satisfactory structure; on the other, a participating user demands the autonomy to act and react without explicit authorial constraint [197].

Accordingly, the audience's participation can also be integrated into this space; this is akin to intervening in and influencing the power fields of both human and AI agents.

The emergence of pervasive AI agents does not, in any form, dismiss or justify their potential infringement of human creations. On the contrary, this paper aims to understand, in a more nuanced manner, the interplay of narrative power between human creators, AI agents, and the audience in the creative process, where risks and opportunities coexist. By considering these evolving narrative power relationships, future researchers and creators may gain deeper insights into AI agents and the ways in which they interact with their audiences.

10.5 Limitations and future directions

10.6 Generative storytelling with LLMs

Some studies in this thesis, such as Wander and Storytelling Community Chatbot (SCC), were conducted before the release of state-of-the-art LLMs like ChatGPT, resulting in limitations in the quality of AI-generated text. In the future, as multi-modal AI models (such as GPT-4V, which allows for both text and image input) continue to advance, I plan to further improve and develop new research initiatives. The potential for enhancing the SCC project is particularly evident. In this thesis, the project was implemented using GPT-3 and still required manual plot setting. However, advanced LLMs like ChatGPT could enable characters to improvise storylines based on community feedback, further integrating AI agents into human social community relationships.

10.7 Linguistic improvement

While Hyborg Agency and AI Nüshu have received widespread recognition, especially AI Nüshu's creation of "non-human language", the linguistic aspect of hybrid agents' language expansion capability remains to be explored in greater depth from a linguistic perspective. For instance, the Hyborgs merely use LLM prompts to generate natural language mixed with code to mimic a "machine speaking" effect. AI Nüshu only demonstrates a top-down algorithm for language generation, inspired by the creation process of Chinese Nüshu. However, the interaction between the human community and agents in Hyborg Agency prompts further consideration of the possibility of bottom-up non-human language generation. For example, it may be possible for people to collaborate with Hyborgs to create a language that evolves under the influence of participants' native languages. Hyborgs, after communicating with speakers of Japanese, Chinese, Thai, and other languages, might develop new expressions that incorporate lexical features from these languages. This process is similar to the formation of pidgin languages in linguistics: communication that develops between two or more groups that do not share a common language, typically with limited vocabulary and grammar often drawn from several languages. In the field of NLP, the study of language emergence has long been a popular topic[230, 183], and I look forward to further integrating cultural and artistic perspectives into this research.

Bibliography

[1] Inworld - the most advanced character engine for ai npcs. `https://inworld.ai/blog/best-ai-mods-for-skyrim-chatgpt-mod`.

[2] Latitude. AI Dungeon. https://play.aidungeon.io/main/home.

[3] Meaning Machine. Dead meat on Steam. https://store.steampowered.com/app/2628740/Dead$_Meat/$.

[4] North Side. Bot Colony in Steam. https://store.steampowered.com/app/263040/Bot_Colony/.

[5] Ocelot Society. Event[0] on Steam. https://store.steampowered.com/app/470260/Event0/.

[6] Earth 2050. Earth 2050: A glimpse into the future | Kaspersky. https://2050.earth.

[7] 3MinuteGame. Lifeline. https://www.3minute.games/.

[8] Nuno Afonso and Rui Prada. Agents that relate: Improving the social believability of non-player characters in role-playing games. In *International Conference on Entertainment Computing*, pages 34–45. Springer, 2008.

[9] Inworld AI. Inworld origins, 2023. Accessed: 2023-08-15.

[10] rct AI. The Metaverse Needs a “Brain”. https://rctai.medium.com/the-metaverse-needs-a-brain-45b528a96487, August 2021.

[11] Bill Albert and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

[12] Anne-Mette Albrechtslund. Gamers telling stories: Understanding narrative practices in an online community. 16(1):112–124. Publisher: SAGE Publications Ltd.

[13] Anne-Mette Albrechtslund. Gamers telling stories. *Convergence: The International Journal of Research into New Media Technologies*, 16(1):112–124, 2010.

[14] R Aljammaz, E Oliver, J Whitehead, and M Mateas. Scheherazade’s tavern: A prototype for deeper npc interactions. In *International Conference on the Foundations of Digital Games*, 2020.

[15] Rehaf Aljammaz, Elizabeth Oliver, Jim Whitehead, and Michael Mateas. Scheherazade's Tavern: A Prototype For Deeper NPC Interactions. In *International Conference on the Foundations of Digital Games*, pages 1–9, Bugibba Malta, September 2020. ACM.

[16] Rehaf Aljammaz, Elizabeth Oliver, Jim Whitehead, and Michael Mateas. Scheherazade's Tavern: A Prototype For Deeper NPC Interactions. In *International Conference on the Foundations of Digital Games*, pages 1–9, Bugibba Malta, September 2020. ACM.

[17] AlterStaff Inc. Ai2u - with you 'til the end. `https://store.steampowered.com/app/2880730/AI2UWith_You_Til_The_End/`, 2024. Adventure, psychological horror visual novel featuring AI-powered characters and escape room challenges.

[18] Sarah Andersen. The alt-right manipulated my comic. then a.i. claimed it. `https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithim-took-my-work.html`. [Accessed 10-09-2023].

[19] Dani Anguiano and Lois Beckett. How hollywood writers triumphed over ai – and why it matters, Oct 2023.

[20] Anthropic. Claude 2. https://www.anthropic.com/index/claude-2, 2023.

[21] Rudolf Arnheim. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, Berkeley, Calif., 2004.

[22] ArtStation. Artstation - works tagged with "character design. `https://www.artstation.com/search?sort_by=relevance&amp;query=Character`. Accessed: [2023 Sept 11].

[23] Sasha Azad and Chris Martens. Little computer people: A survey and taxonomy of simulated models of social interaction. 5:245:1–245:30.

[24] Hannes Bajohr. Writing at a distance: Notes on authorship and artificial intelligence, 03 2023.

[25] Jawid Ahmad Baktash and Mursal Dawodi. Gpt-4: A review on advancements and opportunities in natural language processing, 2023.

[26] T. Bancroft and G. Keane. *Creating characters with personality*. Clarkson Potter/Ten Speed, 2006. tex.lccn: 2005028462.

[27] Julia Barnett. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 146–161, New York, NY, USA, 2023. Association for Computing Machinery.

[28] GAB Barros, A Liapis, and J Togelius. Murder mystery generation from open data. L-Università ta' Malta, 2016.

[29] Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *Transactions on Computer-Human Interaction*, 2004.

[30] Kaj Björkqvist, Karin Österman, and Ari Kaukiainen. Social intelligence empathy = aggression? 5(2):191–200.

[31] Marc Blank and David Lebling. Zork: A computerized fantasy simulation game. Interactive Fiction by Infocom, 1977. Available: http://www.infocom-if.org/downloads/downloads.html.

[32] Eli Blevis. Digital imagery as meaning and form in hci and design: An introduction to the visual thinking backpage gallery. *Interactions*, 18(5):60–65, sep 2011.

[33] Bloc. Inworld sentient streets - ai story mode, Aug 2023.

[34] Anton Bogdanovych, Tomas Trescak, and Simeon Simoff. What makes virtual agents believable? 28(1):83–108.

[35] Julia Ayumi Bopp, Livia J. M"uller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler. Exploring Emotional Attachment to Game Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, pages 313–324, New York, NY, USA, October 2019. Association for Computing Machinery.

[36] Julia Ayumi Bopp, Livia J. Müller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler. Exploring emotional attachment to game characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, pages 313–324. Association for Computing Machinery.

[37] Julia Ayumi Bopp, Livia J. Müller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler. Exploring emotional attachment to game characters. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 2019.

[38] Brian Bourke. Positionality: Reflecting on the research process. *The qualitative report*, 19(33):1–9, 2014.

[39] Alice Bowman. Companionship games: A framework for emotionally engaging and empathetic interactive characters. In Rogelio E. Cardona-Rivera, Anne Sullivan, and R. Michael Young, editors, *Interactive Storytelling*, volume 11869, pages 413–417. Springer International Publishing.

[40] Alice Bowman. Companionship games: A framework for emotionally engaging and empathetic interactive characters. *Interactive Storytelling*, page 413–417, 2019.

[41] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. My AI friend: How users of a social chatbot understand their human–AI friendship. 48(3):404–429.

[42] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. My ai friend: How users of a social chatbot understand their human–ai friendship. *Human Communication Research*, 48(3):404–429, 2022.

[43] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[44] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[45] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[46] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of artificial general intelligence: Early experiments with gpt-4, 2023.

[47] Matthew Butterick. `https://stablediffusionlitigation.com/`.

[48] Alex Calderwood, Vivian Qiu, K. Gero, and Lydia B. Chilton. How novelists use generative language models: An exploratory user study. In *HAI-GEN+user2agent@IUI*, 2020.

[49] Gordon Calleja. Experiential narrative in game environments. In *DiGRA Conference*, 2009.

[50] Cambridge University Press & Assessment. character, 2025. Cambridge Dictionary — English meaning of “character”, accessed October 17, 2025.

[51] Camel-AI. Camel: Communicative agents for “mind” exploration of large scale language model society. `https://github.com/camel-ai/camel`, 2023.

[52] Linda Candy, Shigeki Amitani, and Zafer Bilda. Practice-led strategies for interactive art research. *CoDesign*, 2(4):209–223, 2006.

[53] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT.

[54] Corinne Cath. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180080, 2018.

[55] Luis Ceze. `https://venturebeat.com/datadecisionmakers/dont-be-fooled-by-ai-washing-3-questions-to-ask-before-you-invest/`, 2022. Accessed: 2023-07-08.

[56] Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. Art or artifice? large language models and the false promise of creativity. *ArXiv*, abs/2309.14556, 2023.

[57] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. Creativity support in the age of large language models: An empirical study involving emerging writers. *ArXiv*, abs/2309.12570, 2023.

[58] Senthil Chandrasegaran, Almila Akdag Salah, and Peter Lloyd. Constructing design activity in words: Exploring linguistic methods to analyse the design process. *Design Studies*, 86:101182, 2023.

[59] CharacterAI. Characterai [beta]. `https://beta.character.ai/`, 2023. Accessed: 2024-09-14.

[60] Charisma. Storytelling powered by artificial intelligence.

[61] S. Chatterjee. 1002 short stories from project gutenberg. `https://www.kaggle.com/shubchat/1002-short-stories-from-project-guttenberg`, 2020. Accessed 16 Oct 2022.

[62] Ana Paula Chaves and Marco Aurelio Gerosa. How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. 37(8):729–758. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2020.1841438.

[63] Ana Paula Chaves and Marco Aurelio Gerosa. How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8):729–758, 2020.

[64] Erin Cherry and Celine Latulipe. Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput.-Hum. Interact.*, 21(4), jun 2014.

[65] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*, 2022.

[66] Jennifer Chubb, Darren Reed, and Peter Cowling. Expert views about missing ai narratives: is there an ai story crisis? *AI & society*, 39(3):1107–1126, 2024.

[67] John Joon Young Chung and Eytan Adar. Promptpaint: Steering text-to-image generation through paint medium-like interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.

[68] John Joon Young Chung, Shiqing He, and Eytan Adar. Artist support networks: Implications for future creativity support tools. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, page 232–246, New York, NY, USA, 2022. Association for Computing Machinery.

[69] Aurelio Cianciotta. Moon by Ai Wei Wei and Olafur Eliasson, collective testament. | Neural, 2014.

[70] Mark Coeckelbergh. Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and engineering ethics*, 26(4):2051–2068, 2020.

[71] Mark Coeckelbergh. Narrative responsibility and artificial intelligence: How ai challenges human responsibility and sense-making. *AI & SOCIETY*, 38(6):2437–2450, 2023.

[72] Mark Coeckelbergh. The work of art in the age of ai image generation: Aesthetics and human-technology relations as process and performance. *Journal of Human-Technology Relations*, 1, Jun. 2023.

[73] A Coenen, L Davis, D Ippolito, et al. Wordcraft: A human-ai collaborative editor for story writing. arXiv.org, 2021.

[74] A. Coenen, L. Davis, and D. et al. Ippolito. Wordcraft: A human-ai collaborative editor for story writing. `https://arxiv.org/abs/2107.07430`, 2021.

[75] Samuel Taylor Coleridge and John Shawcross. *Biographia literaria*. Dent London, 1927.

[76] The Pokémon Company. The official pokémon website. `https://www.pokemon.com/`, 2023. Accessed: 2023 Sept 14.

[77] Mia Consalvo. There is no magic circle. *Games and culture*, 4(4):408–417, 2009.

[78] Mike Cook. Infinity is trash (and that's okay), 2023. Accessed: 2023-08-15.

[79] CopilotHub. Copilothub ai. https://app.copilothub.ai/, 2023. Accessed: 2023-09-14.

[80] Coursera. What is character design? + how to get started. `https://www.coursera.org/articles/character-design`.

[81] Samuel Rhys Cox and Wei Tsang Ooi. Conversational interactions with npcs in llm-driven gaming: Guidelines from a content analysis of player feedback. In *International Workshop on Chatbot Research and Design*, pages 167–184. Springer, 2023.

[82] Marc-Alexandre Côté, Ákos Kádár, Xingdi Yuan, Ben Kybartas, Tavian Barnes, Emery Fine, James Moore, Matthew Hausknecht, Layla El Asri, Mahmoud Adada, Wendy Tay, and Adam Trischler. TextWorld: A learning environment for text-based games. In *Computer games workshop at ICML/IJCAI 2018*, pages 1–29. Edition: Computer Games Workshop at ICML/IJCAI 2018.

[83] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[84] Nicholas Davis. An enactive approach to facilitate interactive machine learning for co-creative agents. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, C&C '15, page 345–346, New York, NY, USA, 2015. Association for Computing Machinery.

[85] Deconstructeam. Interview with the whisperer by deconstructeam. https://deconstructeam.itch.io/interview-with-the-whisperer.

[86] Deconstructeam. The Mystery of the Three Bots. https://google.github.io/mysteryofthreebots/.

[87] L DeCosmo. Google engineer claims ai chatbot is sentient: Why that matters, is it possible for an artificial intelligence to be sentient. *Artificial Intelligence, Scientific American, July*, 12, 2022.

[88] DeepL. Deepl translate: The world's most accurate translator, 2023. Translate texts & full document files instantly. Accurate translations for individuals and Teams. Millions translate with DeepL every day. Accessed on August 28, 2023.

[89] DeviantArt. Deviantart - the largest online art gallery and community. https://www.deviantart.com/, 2023. Accessed: [2023 Sept 11].

[90] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[91] Joana Dimas. Social behaviour in games: Dynamic identity in npcs. 2013.

[92] John Dinsmore. *The symbolic and connectionist paradigms: closing the gap*. Psychology Press, 2014.

[93] dreamily.ai. Colorfulclouds tech, dreamily.ai. `https://dreamily.ai/`, 2022. Accessed 16 Oct 2022.

[94] Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. *arXiv preprint arXiv:2305.14325*, 2023.

[95] Haihan Duan, Jiaye Li, Sizheng Fan, Zhonghao Lin, Xiao Wu, and Wei Cai. Metaverse for Social Good: A University Campus Prototype. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 153–161. Association for Computing Machinery, New York, NY, USA, October 2021.

[96] Eagle. Eagle - organize all your reference images in one place, 2023. Accessed: 10 December 2023.

[97] Zillah R Eisenstein. *The color of gender: Reimaging democracy*. Univ of California Press, 2023.

[98] Mirjam P Eladhari, Anne Sullivan, Gillian Smith, and Josh McCoy. Ai-based game design: Enabling new playable experiences. *UC Santa Cruz Baskin School of Engineering, Santa Cruz, CA*, 2011.

[99] Electronic Arts. The sims. Computer Game, 2000.

[100] Katharina Emmerich, Patrizia Ring, and Maic Masuch. I'm glad you are on my side. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 2018.

[101] Katharina Emmerich, Patrizia Ring, and Maic Masuch. I'm Glad You Are on My Side: How to Design Compelling Game Companions. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, pages 141–152, New York, NY, USA, October 2018. Association for Computing Machinery.

[102] Jason K Eshraghian. Human ownership of artificial creativity. *Nature Machine Intelligence*, 2(3):157–160, 2020.

[103] Hugging Face. Open source ai game jam, 2023. A game jam from 2023-07-07 to 2023-07-16 hosted by Hugging Face. The jam encourages participants to unlock their creativity using AI tools in game development.

[104] Behnaz Farahi. " can the subaltern speak?" critical making in design. In *ACM SIGGRAPH 2021 Art Gallery*, pages 1–3. 2021.

[105] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, 132:138–161, December 2019.

[106] Feishu. Lark: All-in-one collaboration platform. `https://www.feishu.cn`, 2021.

[107] FFFILONI. Music to image - a hugging face space by fffiloni. `https://huggingface.co/spaces/fffiloni/Music-To-Image`.

[108] Matthew Field. Facebook shuts down robots after they invent their own language. *The Telegraph*, 2017.

[109] Walter R. Fisher. Narration as a human communication paradigm: The case of public moral argument. *Communication Monographs*, 51(1):1–22, 1984.

[110] Catherine Flick and Kyle Worrall. The ethics of creative AI. In Craig Vear and Fabrizio Poltronieri, editors, *The language of creative AI: Practices, aesthetics and structures*, pages 73–91. Springer International Publishing, 2022.

[111] Luciano Floridi and Mariarosaria Taddeo. What is data ethics?, 2016.

[112] Asbjørn Følstad, Vegard Engen, Ida Maria Haugstveit, and J Brian Pickering. Automation in human-machine networks: how increasing machine agency affects human agency. In *Man-Machine Interactions 5: 5th International Conference on Man-Machine Interactions, ICMMI 2017 Held at Kraków, Poland, October 3-6, 2017*, pages 72–81. Springer, 2018.

[113] Transitional Forms. Robots make tv. `https://www.twitch.tv/robotsmaketv`.

[114] E. M. Forster. *Aspects of the Novel*. Harcourt, Brace Company, United States, 1927. EBook No. 70492.

[115] Jamie Fraser, Ioannis Papaioannou, and Oliver Lemon. Spoken Conversational AI in Video Games: Emotional Dialogue Management Increases User Engagement. In *Proceedings of the*

18th International Conference on Intelligent Virtual Agents, IVA '18, pages 179–184, New York, NY, USA, November 2018. Association for Computing Machinery.

[116] Sarah E Fraser and Yu-Chieh Li. Xu bing: Beyond the book from the sky. 2020.

[117] Christopher Frayling. Research in art and design. *Royal College of Art research papers*, 1:1–5, 1993.

[118] Jane Friedhoff, Suma Bailis, and Feiyang Chen. Gdc 2024: Ai summit: Simulacra and subterfuge: Building agentic "werewolf", Mar 2024.

[119] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[120] William Gaver. What should we expect from research through design? In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 937–946, 2012.

[121] GENVID. Rival peak | an experimental competition reality show, featuring artificially intelligent contestants whose fate is controlled by YOU.

[122] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. Sparks: Inspiration for science writing using language models. In *Designing Interactive Systems Conference*, DIS '22, page 1002–1019, New York, NY, USA, 2022. Association for Computing Machinery.

[123] Barney Glaser and Anselm Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[124] Danny Godin and Mithra Zahedi. Aspects of research through design: a literature review. 2014.

[125] Paulo Gomes, Carlos Martinho, Ana Paiva, and Arnav Jhala. Metrics for Character Believability in Interactive Narrative. volume 8230, November 2013.

[126] Google. Banter bot by google creative lab - experiments with google. `https://experiments.withgoogle.com/banter-bot`. Accessed 16 Oct 2022.

[127] Google. Semantic ML.

[128] Google Creative Lab. Banter bot by google creative lab - experiments with google. Google, 2022. Accessed 16 Oct 2022.

[129] gpt 2. Language models are unsupervised multitask learners, April 2019.

[130] Significant Gravitas. Auto-gpt: An autonomous gpt-4 experiment. `https://github.com/Significant-Gravitas/Auto-GPT`, 2023.

[131] Jorge Luis Morton Gutiérrez. On actor-network theory and algorithms: Chatgpt and the new power relationships in the age of ai.

[132] McKenna Hadley-Burke. This artist created an ai chatbot to talk to her younger self - cbc radio. https://www.cbc.ca/radio/tapestry/artist-creates-chatbot-to-talk-to-younger-self-1.6718087.

[133] Thilo Hagendorff. The ethics of ai ethics: An evaluation of guidelines. *Minds and machines*, 30(1):99–120, 2020.

[134] Joe Hall. A week with wanda.

[135] Joe Hall. A Week With Wanda. https://weekwithwanda.wordpress.com/.

[136] Halopedia. I love bees, Dec 2023.

[137] Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. *arXiv preprint arXiv:2305.14992*, 2023.

[138] Donna Haraway. 'situated knowledges: The science question in feminism and the privilege of partial perspective'. In *Space, Gender, Knowledge: Feminist Readings*, pages 53–72. Routledge, 2016.

[139] Donna J Haraway. *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press, 2016.

[140] Zac Harris. Become AI Native: What It Means (& How to Do It) — Copy.ai. `https://www.copy.ai/blog/ai-native`, 2023. Accessed: 2023-07-08.

[141] Melanie Hartmann. Challenges in developing user-adaptive intelligent user interfaces. In *LWA*, pages ABIS–6. Citeseer, 2009.

[142] Matthew Hausknecht, Prithviraj Ammanabrolu, Côté Marc-Alexandre, and Yuan Xingdi. Interactive fiction games: A colossal adventure. *CoRR*, abs/1909.05398, 2019.

[143] Shreya Havaldar, Sunny Rai, Bhumika Singhal, Langchen Liu Sharath Chandra Guntuku, and Lyle Ungar. Multilingual language models are not multicultural: A case study in emotion. *arXiv preprint arXiv:2307.01370*, 2023.

[144] Jean-Claude Heudin. An emotional multi-personality architecture for intelligent conversational agents. *Transactions on Computational Collective Intelligence XXVIII*, page 143–163, 2018.

[145] Jaakko Hintikka. *Wittgenstein and Language as the Universal Medium*, pages 162–190. Springer Netherlands, Dordrecht, 1997.

[146] HJL. fishman. https://www.pixiv.net/artworks/80168666, 2020. Accessed: Accessed: 2023-09-14.

[147] Annabell Ho, Jeff Hancock, and Adam S Miner. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4):712–733, 2018.

[148] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[149] Jiarui LI Huan LI. Wechaty: Conversational sdk for chatbot makers. `https://github.com/wechaty/wechaty`, 2016.

[150] Johan Huizinga. *Homo ludens ils 86*. Routledge, 2014.

[151] Aaron Chia Yuan Hung. Beyond the player: A user-centered approach to analyzing digital games and players using actor-network theory. *E-Learning and Digital Media*, 13(5-6):227–243, 2016.

[152] Angel Hsing-Chi Hwang. Too late to be creative? ai-empowered tools in creative processes. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[153] hwchase17. Langchain: Building applications with llms through composability. `https://github.com/hwchase17/langchain`, 2023.

[154] Kiran Ijaz, Anton Bogdanovych, and Simeon Simoff. Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness. page 10.

[155] Discord Inc. Discord — your place to talk and hang out. https://discord.com/, 2023. Accessed: September 11, 2023.

[156] Inworld AI. Inworld – the most advanced character engine for ai npcs. https://inworld.ai/, 2023. Accessed: 2023-09-14.

[157] D Jackson and A Latham. Talk to the ghost: The storybox methodology for faster development of storytelling chatbots. *Expert Systems with Applications*, 190:116223, 2022.

[158] David Jackson and Annabel Latham. Talk to the ghost: The storybox methodology for faster development of storytelling chatbots. *Expert Systems with Applications*, 190:116223, 2022.

[159] Anthony David Jameson. Understanding and dealing with usability side effects of intelligent processing. *Ai Magazine*, 30(4):23–23, 2009.

[160] Fredric Jameson. The prison-house of language: A critical account of structuralism and russian formalism. 2020.

[161] Peter Jansen. A systematic survey of text worlds as embodied natural language environments. In Marc-Alexandre Côté, Xingdi Yuan, and Prithviraj Ammanabrolu, editors, *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022)*, pages 1–15. Association for Computational Linguistics.

[162] Sheila Jasanoff and Sang-Hyun Kim. *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press, 2019.

[163] Henry Jenkins. Game design as narrative architecture. 2003.

[164] Stephanie C Jennings. Co-creation and the distributed authorship of video games. In *Examining the evolution of gaming and its impact on social, cultural, and political perspectives*, pages 123–146. IGI Global, 2016.

[165] Youngseung Jeon, Seungwan Jin, Patrick C. Shih, and Kyungsik Han. Fashionq: An ai-driven creativity support tool for facilitating ideation in fashion design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[166] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. Ai art and its impact on artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 363–374, New York, NY, USA, 2023. Association for Computing Machinery.

[167] Tan Jiu. Pub. https://www.pixiv.net/artworks/46113792, 2014. Accessed: Accessed: 2023-09-14.

[168] Magnus Johansson, Björn Strååt, Henrik Warpefelt, and Harko Verhagen. *Analyzing the Social Dynamics of Non-Player Characters*. 01 2014.

[169] Mirabelle Jones, Christina Neumayer, and Irina Shklovski. Embodying the algorithm: Exploring relationships with large language models through artistic performance. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[170] Heewoo Jun and Alex Nichol. Shap-e: Generating conditional 3d implicit functions, 2023.

[171] J. Juul. *Half-real: Video games between real rules and fictional worlds*. MIT Press. tex.lccn: 2005047863.

[172] Stamatis Karnouskos. Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, 1(3):138–147, 2020.

[173] Margaret L. Keeling and Maria Bermudez. Externalizing problems through art and writing: Experience of process and helpfulness. *Journal of Marital and Family Therapy*, 32(4):405–419, 2006.

[174] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. Metaphorian: Leveraging large language models to support extended metaphor creation for science writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 115–135, New York, NY, USA, 2023. Association for Computing Machinery.

[175] Leo Kim. Korean illustrator kim jung gi's "resurrection" via ai image generator is orientalism in new clothing. https://www.artnews.com/art-news/news/kim-jung-gi-death-stable-diffusion-artificial-intelligence-1234649787/, Dec 2022.

[176] Hyung-Kwon Ko, Subin An, Gwanmo Park, Seung Kwon Kim, Daesik Kim, Bohyoung Kim, Jaemin Jo, and Jinwook Seo. We-toon: A communication support system between writers and artists in collaborative webtoon sketch revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA, 2022. Association for Computing Machinery.

[177] kuki team. @kuki_ai.

[178] Hagung Kuntjara and Betha Almanfaluthi. Character design in games analysis of character design theory. *Journal of Games, Game Art, and Gamification*, 2(2), 2017. Number: 2.

[179] Max Lamparth, Anthony Corso, Jacob Ganz, Oriana Skylar Mastro, Jacquelyn Schneider, and Harold Trinkunas. Human vs. machine: Language models and wargames. *arXiv preprint arXiv:2403.03407*, 2024.

[180] Latitude. *AI Dungeon*. Latitude, 2020. Accessed: 2023-07-08.

[181] Bruno Latour. On actor-network theory: A few clarifications. 47(4):369–381. Publisher: Nomos Verlagsgesellschaft mbH.

[182] Tomas Lawton, Kazjon Grace, and Francisco J Ibarrola. When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1978–1996, New York, NY, USA, 2023. Association for Computing Machinery.

[183] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182*, 2016.

[184] Jooyoung Lee, Thai Le, Jinghui Chen, and Dongwon Lee. Do language models plagiarize? In *Proceedings of the ACM Web Conference 2023*. ACM, apr 2023.

[185] Lik-Hang Lee, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda, 2021.

[186] Lik-Hang Lee, Zijun Lin, Rui Hu, Zhengya Gong, Abhishek Kumar, Tangyao Li, Sijia Li, and Pan Hui. When Creators Meet the Metaverse: A Survey on Computational Arts. November 2021.

[187] Michael Sangyeob Lee and Carrie Heeter. What do you mean by believable characters?: The effect of character rating and hostility on the perception of character believability. *Journal of Gaming Virtual Worlds*, 4(1):81–97, March 2012.

[188] David Leslie. Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684*, 2019.

[189] Jonathan Lessard. Designing natural-language game conversations. *Proc. DiGRA-FDG*, 16, 2016.

[190] Jingyi Li, Eric Rawn, Jacob Ritchie, Jasper Tran O'Leary, and Sean Follmer. Beyond the artifact: Power as a lens for creativity support tools. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.

[191] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation, 2022.

[192] J. C. R. Licklider. Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1(1):4–11, 1960.

[193] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. It is your turn: Collaborative ideation with a co-creative robot through sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[194] Hao Liu, Carmelo Sferrazza, and Pieter Abbeel. Chain of hindsight aligns language models with feedback, 2023.

[195] Monika Lohani, Charlene Stokes, Marissa McCoy, Christopher A. Bailey, and Susan E. Rivers. Social interaction moderates human-robot trust-reliance relationship and improves stress coping. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 471–472, March 2016.

[196] Duri Long, Takeria Blunt, and Brian Magerko. Co-designing ai literacy exhibits for informal learning spaces. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.

[197] Sandy Louchart and Ruth Aylett. The emergent narrative theoretical investigation. In *the 2004 Conference on Narrative and Interactive Learning Environments*, pages 21–28, 2004.

[198] Arnold M Lund. Measuring usability with the use questionnaire12. *Usability interface*, 8(2):3–6, 2001.

[199] Philipp Maas, Frank Carey, Chris Wheeler, Edward Saatchi, Pete Billington, and Jessica Yaffa Shamash. To infinity and beyond: Show-1 and showrunner agents in multi-agent simulations. *arXiv preprint*, 2023.

[200] Alasdair MacIntyre. *After virtue*. A&C Black, 2013.

[201] Nikki Main. Replika's companion chat bot reportedly loses the sex and leaves fans despondent, Feb 2023.

[202] Maarit Mäkelä. Knowing through making: The role of the artefact in practice-led research. *Knowledge, Technology & Policy*, 20(3):157–163, 2007.

[203] Shahar Marnin-Distelfeld and Uri Dorchin. "i am not an artist, i make art": amateurish artists in israel and the sense of creativity. *Creativity Studies*, 13(1):64–86, 2020. Number: 1.

[204] Giulia Marsan, Nicola Bellomo, and Andrea Tosin. *Complex Systems and Society - Modeling and Simulation*. 04 2013.

[205] Lara Jean Martin. Neurosymbolic automated story generation. Publisher: Georgia Institute of Technology.

[206] THOMAS L. MARTIN. *Poiesis and Possible Worlds: A Study in Modality and Literary Theory*. University of Toronto Press, 2004.

[207] Dinand Roeland Mirko D Angelo Massimo Iovene, Leif Jonsson. A detailed study of the ai native concept. `https://www.ericsson.com/en/reports-and-papers/white-papers/ai-native`, 2023. Accessed: 2023-07-08.

[208] Michael Mateas. An oz-centric review of interactive drama and believable agents. In Michael J. Wooldridge and Manuela Veloso, editors, *Artificial Intelligence Today: Recent Trends and Developments*, Lecture Notes in Computer Science, pages 297–328. Springer.

[209] Michael Mateas. An oz-centric review of interactive drama and believable agents. In *Artificial intelligence today: Recent trends and developments*, pages 297–328. Springer, 2001.

[210] Michael Mateas and Andrew Stern. Façade: An Experiment in Building a Fully-Realized Interactive Drama. page 24.

[211] Michael Mateas and Andrew Stern. Façade: An Experiment in Building a Fully-Realized Interactive Drama. page 24.

[212] Michael Mateas and Andrew Stern. Façade: An experiment in building a fully-realized interactive drama. page 24.

[213] Michael Mateas and Andrew Stern. Build it to understand it: Ludology meets narratology in game design space. 01 2005.

[214] Marian Mazzone and Ahmed Elgammal. Art, creativity, and the potential of artificial intelligence. *Arts*, 8(1):26, 2019-03. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[215] J McCoy, M Treanor, B Samuel, et al. Prom week: Designing past the game/story dilemma. In *Foundation of Digital Games (FDG)*, 2013.

[216] Joshua McCoy, Mike Treanor, Ben Samuel, Aaron A. Reed, Michael Mateas, and Noah Wardrip-Fruin. Prom week: Designing past the game/story dilemma. In Georgios N. Yannakakis, Espen Aarseth, Kristine Jørgensen, and James C. Lester, editors, *Proceedings of the 8th International Conference on the Foundations of Digital Games, FDG 2013, Chania, Crete, Greece, May 14-17, 2013*, pages 94–101. Society for the Advancement of the Science of Digital Games, 2013.

[217] Justin McCurry. South korean ai chatbot pulled from facebook after hate speech towards minorities, Jan 2021.

[218] James R. Meehan. Tale-spin, an interactive program that writes stories. In *International Joint Conference on Artificial Intelligence*, 1977.

[219] Elisa D. Mekler and Kasper Hornbæk. Momentary pleasure or lasting meaning? distinguishing eudaimonic and hedonic user experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 4509–4520, New York, NY, USA, 2016. Association for Computing Machinery.

[220] Cade Metz. Riding out quarantine with a chatbot friend: 'i feel very connected', Jun 2020.

[221] Microsoft. Xiaoice land beta. https://island.xiaoice.com/.

[222] Midjourney. Midjourney. `https://www.midjourney.com/`, 2023.

[223] milkway25. R/characterai on reddit: Whats your favorite creator, Oct 2023.

[224] Piotr Mirowski, Kory W. Mathewson, Jaylen Pittman, and Richard Evans. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[225] Brent Mittelstadt. Principles alone cannot guarantee ethical ai. *Nature machine intelligence*, 1(11):501–507, 2019.

[226] Shigeru Miyamoto, Takashi Tezuka, and Koji Kondo. The legend of zelda, 1986.

[227] Nick Montfort. *Twisty Little Passages*.

[228] Markus Montola. Exploring the edge of the magic circle : Defining pervasive games. 2005.

[229] Markus Montola, Jaakko Stenros, and Annika Waern. *Pervasive games: theory and design*. CRC Press, 2009.

[230] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations, 2018.

[231] Caterina Moruzzi. Creative agents: Rethinking agency and creativity in human and artificial systems. ISSN: 2053-9320.

[232] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. Genaichi: generative ai and hci. In *CHI conference on human factors in computing systems extended abstracts*, pages 1–7, 2022.

[233] Daniel Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.

[234] Yohei Nakajima. Babyagi. `https://github.com/yoheinakajima/babyagi`, 2023.

[235] Emily Naul and Min Liu. Why story matters: A review of narrative in serious games. 58(3):687–707. Publisher: SAGE Publications Inc.

[236] Gina Neff. Talking to bots: Symbiotic agency and the case of tay. *International Journal of Communication*, 2016.

[237] Eva Nieuwdorp. The pervasive interface: Tracing the magic circle. In *Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play*, 2005.

[238] R&D4 Nintendo. Super mario bros. *Game [NES].(13 September 1985). Nintendo, Kyoto, Japan*, 1985.

[239] Writers Guild of America. 2023 strike rules. https://www.wgacontract2023.org/strike/strike-rules, 2023. Accessed: 2023-09-14.

[240] University of Nebraska Press, editor. *Possible Worlds Theory and Contemporary Narratology*. University of Nebraska Press, 2019.

[241] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[242] ohmlet. Snip it, 2023. Snip away and explore famous paintings in a whole new dimension. Submitted for the Open Source AI Game Jam hosted by Hugging Face.

[243] OpenAI. OpenAI API.

[244] OpenAI. CLIP, 2020.

[245] OpenAI. Chatgpt: Language model for conversations. `https://openai.com/blog/chatgpt`, 2023. Accessed 2023-10-05.

[246] OpenAI. Dall·e 2. https://openai.com/product/dall-e-2, 2023. Accessed: 2023-06-01.

[247] OpenAI. Gpt-4, 2023.

[248] OpenAI. GPT-4. `https://openai.com/research/gpt-4`, 2023.

[249] OpenAI. Gpt-4 technical report. *ArXiv*, abs/2303.08774, 2023.

[250] OpenAI. Chatgpt: A large language model. OpenAI, 2024. Accessed 23 Oct 2024.

[251] OpenAI. Introducing the gpt store, Jan 2024.

[252] OpenAI and Ashley Pilipiszyn. Gpt-3 powers the next generation of apps. https://openai.com/blog/gpt-3-apps, March 2021. Accessed: 2023 Sept 13.

[253] Papegames. Evol4. https://evol.papegames.cn/.

[254] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA, 2023. Association for Computing Machinery.

[255] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior, 2023.

[256] Iryna Pentina, Tyler Hancock, and Tianling Xie. Exploring relationship development with social chatbots: A mixed-method study of replika. 140:107600.

[257] Iryna Pentina, Tyler Hancock, and Tianling Xie. Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140:107600, 2023.

[258] Steve Phelps and Yvan I. Russell. Investigating emergent goal-like behaviour in large language models using experimental economics, 2023.

[259] M Pichlmair and C Putney. Procedural generation for divination and inspiration. In *International Conference on the Foundations of Digital Games*, 2020.

[260] Pinterest. Pinterest, 2023. Accessed: 10 December 2023.

[261] Pixiv. pixiv. https://www.pixiv.net/, 2023. Accessed: 2023 Sept 13.

[262] Ken Plummer. *Narrative power: The struggle for human value*. John Wiley & Sons, 2019.

[263] Indrasen Poola. Artificial intelligence best narrative stories. *International Journal of Novel Research in computer science and software engineering*, 4:1–4, 10 2017.

[264] Anuttacon (Developer & Publisher). Whispers from the star, August 2025. Steam page, accessed October 17, 2025.

[265] Emursive Punchdrunk. Sleep no more. `https://mckittrickhotel.com/events/sleep-no-more/`, 2011. Immersive theatrical experience based on Shakespeare's Macbeth, set in a noir-inspired world at the McKittrick Hotel, NYC.

[266] Qualtrics. Qualtrics. `https://www.qualtrics.com`, 2020.

[267] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 2019.

[268] Rapptz. Github - rapptz/discord.py: An api wrapper for discord written in python. https://github.com/Rapptz/discord.py, 2023. Accessed: September 11, 2023.

[269] LC RAY and D Mizuno. Designing for narrative influence. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[270] Aaron A Reed, Jacob Garbe, Noah Wardrip-Fruin, and Michael Mateas. Ice-bound: Combining richly-realized story with expressive gameplay. In *FDG*, 2014.

[271] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[272] Sofia Reis. Expanding the magic circle in pervasive casual play. In *International Conference on Entertainment Computing*, pages 486–489. Springer, 2012.

[273] Replika. Replika. https://replika.ai.

[274] replika. Replika.

[275] Jeba Rezwana and Mary Lou Maher. Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. page 3519026.

[276] Jeba Rezwana and Mary Lou Maher. Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Trans. Comput.-Hum. Interact.*, 30(5), sep 2023.

[277] P. Ricoeur. *Time and Narrative I–III*, volume 1-3. University of Chicago Press, Chicago, IL, 1984.

[278] P. Ricoeur. Life in quest of narrative. In D. Wood, editor, *On Paul Ricoeur: Narrative and Interpretation*, pages 20–33. Routledge, London, 1991.

[279] Anna Ridler. Alice & bob, 2017.

[280] Mark Riedl and Vadim Bulitko. Interactive narrative: A novel application of artificial intelligence for computer games. 26(1):2160–2165.

[281] Mark O. Riedl. Computational narrative intelligence: A human-centered goal for artificial intelligence.

[282] Mark Owen Riedl and Vadim Bulitko. Interactive narrative: An intelligent systems approach. 34(1):67.

[283] Right Click Save. The lumen prize — yuqian sun, 10 2024. Accessed: 2024-10-13.

[284] Shlomith Rimmon-Kenan. *Narrative Fiction: Contemporary Poetics*. Routledge, London, 1983.

[285] Stephanie Rennick Roberts, Seán. Improving video game conversations with trope-informed design. 21(3).

[286] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. `https://github.com/Stability-AI/stablediffusion`, 2023. GitHub repository.

[287] Joanne Kathleen Rowling. *Harry Potter and the deathly hallows*, volume 7. Bloomsbury Publishing, 2013.

[288] Runway ML. RunwayML — Machine learning for creators. https://runwayml.com/.

[289] Stuart Jonathan Russell and Peter Norvig. *Artificial intelligence: A modern approach;[the intelligent agent book]*. Prentice hall, 1995.

[290] Marie-Laure Ryan. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press, 0 edition edition.

[291] Leonard Salewski, Stephan Alaniz, Isabel Rio-Torto, Eric Schulz, and Zeynep Akata. In-context impersonation reveals large language models' strengths and biases, 2023.

[292] Ben Samuel, Mike Treanor, and Joshua McCoy. Design considerations for creating ai-based gameplay. 2020.

[293] R Keith Sawyer and Stacy DeZutter. Distributed creativity: How collective creations emerge from collaboration. *Psychology of aesthetics, creativity, and the arts*, 3(2):81, 2009.

[294] Markus Schlosser. Agency. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2019 edition.

[295] O Schmitt and D Buschek. Characterchat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. In *Creativity and Cognition*, 2021.

[296] O. Schmitt and D. Buschek. Characterchat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. *Creativity and Cognition*, 2021.

[297] Oliver Schmitt and Daniel Buschek. Characterchat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. In *Creativity and Cognition*, C&C '21, New York, NY, USA, 2021. Association for Computing Machinery.

[298] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Applying the user experience questionnaire (ueq) in different evaluation scenarios. In Aaron Marcus, editor, *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, pages 383–392, Cham, 2014. Springer International Publishing.

[299] Victoria S. Scotney, Sarah Weissmeyer, Nicole Carbert, and Liane Gabora. The ubiquity of cross-domain thinking in the early phase of the creative process. *Frontiers in Psychology*, 10, 2019.

[300] Katie Seaborn and Deborah I. Fels. Gamification in theory and action: A survey. 74:14–31.

[301] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. Beyond Dyadic Interactions: Considering Chatbots as Community Members. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery.

[302] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. Beyond dyadic interactions. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[303] Joseph Seering, Michal Luria, Connie Ye, Geoff Kaufman, and Jessica Hammer. It takes a village: Integrating an adaptive chatbot into an online gaming community. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

[304] Shawn Shan, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, and Ben Y. Zhao. Glaze: Protecting artists from style mimicry by text-to-image models, 2023.

[305] Murray Shanahan. Talking about large language models. 67(2):68–79, 2024.

[306] Murray Shanahan, Kyle McDonell, and Laria Reynolds. Role play with large language models. 623(7987):493–498, 2023. Publisher: Nature Publishing Group.

[307] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A trainable agent for role-playing.

[308] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-llm: A trainable agent for role-playing, 2023.

[309] Donghoon Shin, Sangwon Yoon, Soomin Kim, and Joonhwan Lee. Blahblahbot: Facilitating conversation between strangers using a chatbot with ml-infused personalized topic suggestion. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[310] Viktor Shklovsky. *Art as Technique*. Literary Theory: An Anthology 3. 1917.

[311] Ben Shneiderman and Pattie Maes. Direct manipulation vs. interface agents. interactions, 4 (6): 42–61. *Google Scholar Google Scholar Digital Library Digital Library*, 1997.

[312] Wai Man Si, Michael Backes, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Yang Zhang. Why so toxic? measuring and triggering toxic behavior in open-domain chatbots. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, page 2659–2673, New York, NY, USA, 2022. Association for Computing Machinery.

[313] skillshare. How to create a character mood board: A guide and template. Skillshare Blog, 2023. Accessed: 10 December 2023.

[314] Gillian Smith, Elaine Gan, Alexei Othenin-Girard, and Jim Whitehead. Pcg-based game design: enabling new play experiences through procedural content generation. In *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games*, pages 1–4, 2011.

[315] South China Morning Post. ChatGPT in China: Big Tech firms in race to introduce AI chatbots to the world's biggest internet population, Feb 2023.

[316] Square Enix. *SQUARE ENIX AI Tech Preview: THE PORTOPIA SERIAL MURDER CASE*. Square-Enix, 2023.

[317] Stability AI. CompVis/stable-diffusion · Hugging Face, 2022.

[318] Sangho Suh, Jian Zhao, and Edith Law. Codetoon: Story ideation, auto comic generation, and structure mapping for code-driven storytelling. *The 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.

[319] Theodore Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L. Griffiths. Cognitive architectures for language agents, 2023.

[320] Qian Sun, Yan Chen, Wenyuan Tao, Han Jiang, Mu Zhang, Kan Chen, and Marius Erdt. A GAN-based approach toward architectural line drawing colorization prototyping. *The Visual Computer*, 38(4):1283–1300, 2022.

[321] Suno, 2024.

[322] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[323] R Swanson and AS Gordon. Say anything: A massively collaborative open domain story writing companion, 2006.

[324] R. Swanson and A.S. Gordon. Say anything: A massively collaborative open domain story writing companion. In *SpringerLink*. 2006.

[325] Dave Szulborski. *This is not a game: A guide to alternate reality gaming*. Incunabula, 2005.

[326] Unity Technologies. Experience the art of the possible with unity ai. https://unity.com/ai, 2023. Accessed: 2023-09-14.

[327] Caiyun Technology. Dreamily/ API - Caiyun Wiki, 2020.

[328] Katie Salen Tekinbas and Eric Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2003.

[329] For the Player to Be Both a Father and the NPCs of the Nishuihan Mobile Game Are under a Lot of Pressure iNEWS iNEWS a Mother. https://inf.news/en/game, 2023.

[330] Andreas Theodorou, Robert H Wortham, and Joanna J Bryson. Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science*, 29(3):230–241, 2017.

[331] Jakob Tholander and Martin Jonsson. Design ideation with ai - sketching, thinking and talking with generative machine learning models. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1930–1940, New York, NY, USA, 2023. Association for Computing Machinery.

[332] Bryan Tillman. *Creative character design*. Focal Press, 2011.

[333] Julian Togelius. AI researchers, video games are your friends! In *Computational Intelligence*, Studies in Computational Intelligence, pages 3–18. Springer International Publishing.

[334] Jessica E. Tompkins and Nicole Martins. Masculine pleasures as normalized practices: Character design in the video game industry. *Games and Culture*, 17(3):399–420, 2022.

[335] Mike Treanor, Alexander Zook, Mirjam P Eladhari, Julian Togelius, Gillian Smith, Michael Cook, Tommy Thompson, Brian Magerko, John Levine, and Adam Smith. Ai-based game design patterns. 2015.

[336] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[337] Tony Veale and Mike Cook. *Twitterbots: Making machines that make meaning*. The MIT Press.

[338] Verdal. Neuro-sama, 2023.

[339] virtualyoutuber. Neuro-sama, 2022.

[340] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen. A survey on large language model based autonomous agents.

[341] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen. A survey on large language model based autonomous agents, 2023.

[342] Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. Scienceworld: Is your agent smarter than a 5th grader?, 2022.

[343] Xintao Wang, Quan Tu, Yaying Fei, Ziang Leng, and Cheng Li. Does role-playing chatbots capture the character personalities? assessing personality traits for role-playing chatbots, 2023.

[344] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.

[345] Yi Wang and Max Kreminski. Can llms generate good stories? insights and challenges from a narrative planning perspective. *arXiv preprint arXiv:2506.10161*, 2025.

[346] Noah Wardrip-Fruin. The tale-spin effect. 2009.

[347] Henrik Warpefelt. *Mind the Gap - Exploring the social capability of non-player characters*. PhD thesis, 12 2013.

[348] Henrik Warpefelt. *The Non-Player Character: Exploring the believability of NPC presentation and behavior*. PhD thesis, Department of Computer and Systems Sciences, Stockholm University, 2016.

[349] Henrik Warpefelt, Magnus Johansson, and Harko Verhagen. Analyzing the believability of game character behavior using the game agent matrix. In *Proceedings of DiGRA 2013 Conference*, 2013.

[350] WeChat. Wechat. https://www.wechat.com/, 2023. Accessed: [2023 Sept 11].

[351] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*, 2022.

[352] Justin Weinberg. Philosophers on next-generation large language models. Section: Technology.

[353] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[354] Joseph Weizenbaum. Computer power and human reason: From judgment to calculation. 1976.

[355] Lilian Weng. Llm-powered autonomous agents. *lilianweng.github.io*, Jun 2023.

[356] Ludwig Wittgenstein. Tractatus logico-philosophicus. 2023.

[357] Ludwig Wittgenstein and Ray Monk. *Tractatus logico-philosophicus*. Routledge, 1922.

[358] Wolf3D Inc. Ready player me: Cross-game avatar platform for the metaverse, 2023. Accessed: 2024-10-28.

[359] Michael Wooldridge and Nicholas R Jennings. Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152, 1995.

[360] Wubao. Beloved l. https://belovedl.art/, 2023. Accessed: Accessed: 2023-09-14.

[361] Wutiaoren. About wutiaoren. `https://wutiaoren.info/`. Accessed Apr. 04, 2022.

[362] Yadong Xi, Xiaoxi Mao, Le Li, Lei Lin, Yanjiang Chen, Shuhan Yang, Xuhan Chen, Kailun Tao, Zhi Li, Gongzheng Li, Lin Jiang, Siyan Liu, Zeng Zhao, Minlie Huang, Changjie Fan, and Zhipeng Hu. KuiLeiXi: A Chinese Open-Ended Text Adventure Game. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 175–184, Online, August 2021. Association for Computational Linguistics.

[363] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. The rise and potential of large language model based agents: A survey.

[364] Chloe Xiang. Artists are revolting against ai art on artstation. `https://www.vice.com/en/article/ake9me/artists-are-revolt-against-ai-art-on-artstation`, Dec 2022.

[365] Junlin Xie, Zhihong Chen, Ruifei Zhang, Xiang Wan, and Guanbin Li. Large multimodal agents: A survey. *arXiv preprint arXiv:2402.15116*, 2024.

[366] Bing Xu. Xu bing - artwork - book from the ground, 2014.

[367] Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658*, 2023.

[368] Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. Exploring large language models for communication games: An empirical study on werewolf, 2023.

[369] Vivianna Yan. `https://helixngc7293.itch.io/yandere-ai-girlfriend-simulator`, 2023. Accessed: 2023-07-08.

[370] P Yanardag, M Cebrian, and I Rahwan. Shelley: A crowd-sourced collaborative horror writer. *Creativity and Cognition*, 2021.

[371] Daijin Yang, Erica Kleinman, and Casper Harteveld. Gpt for games: A scoping review (2020-2023). In *2024 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2024.

[372] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and Ray LC. Ai as an active writer: Interaction strategies with generated text in human-ai collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, volume 10, pages 1–11. CEUR-WS Team, 2022.

[373] Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. Doc: Improving long story coherence with detailed outline control, 2023.

[374] Georgios N. Yannakakis and Julian Togelius. A panorama of artificial and computational intelligence in games. 7(4):317–335. Conference Name: IEEE Transactions on Computational Intelligence and AI in Games.

[375] Jia-Yu Yao, Kun-Peng Ning, Zhen-Hui Liu, Mu-Nan Ning, and Li Yuan. Llm lies: Hallucinations are not bugs, but features as adversarial examples.

[376] Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. Webshop: Towards scalable real-world web interaction with grounded language agents, 2023.

[377] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*, 2023.

[378] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models, 2023.

[379] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. Wordcraft: Story writing with large language models. In *27th International Conference on Intelligent User Interfaces*, IUI '22, page 841–852, New York, NY, USA, 2022. Association for Computing Machinery.

[380] Tomáš Zemčík. A brief history of chatbots.

[381] Yi Zeng, Enmeng Lu, and Cunqing Huangfu. Linking artificial intelligence principles. *arXiv preprint arXiv:1812.04814*, 2018.

[382] Bo Zhang, Jun Zhu, and Hang Su. Toward the third generation artificial intelligence. *Science China Information Sciences*, 66(2):121101, 2023.

[383] Bowen Zhang and Harold Soh. Large language models as zero-shot human models for human-robot interaction. *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7961–7968, 2023.

[384] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. Storydrawer: A child–ai collaborative drawing system to support children's creative visual storytelling. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[385] Chaoning Zhang, Chenshuang Zhang, Sheng Zheng, Yu Qiao, Chenghao Li, Mengchun Zhang, Sumit Kumar Dam, Chu Myaet Thwal, Ye Lin Tun, Le Luang Huy, et al. A complete survey on generative ai (aigc): Is chatgpt from gpt-4 to gpt-5 all you need? *arXiv preprint arXiv:2303.11717*, 2023.

[386] Wanqing Zhang. Dan's the man: Why chinese women are looking to chatgpt for love. *BBC News*, 2024. Published 13 June 2024.

[387] Weidi Zhang, Donghao Ren, and George Legrady. Cangjie's poetry: an interactive art experience of a semantic human-machine reality. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 4(2):1–9, 2021.

[388] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. *CHI Conference on Human Factors in Computing Systems*, 2022.

[389] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[390] Liming Zhao and Gongjin Zhang. *N*ü *Shu Yu N*ü *Shu Wen Hua*. Xin hua chu ban she, 1995.

[391] Andrew Zhu, Lara Martin, Andrew Head, and Chris Callison-Burch. Calypso: Llms as dungeon master's assistants. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 19(1):380–390, Oct. 2023.

[392] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, page 493–502, New York, NY, USA, 2007. Association for Computing Machinery.

[393] Inc. Zoom Video Communications. One platform to connect — zoom. https://zoom.us/, 2023. Accessed: 2023-09-14.

[394] Tubagus Zufri, Dodi Hilman, and Wahyudi Pratama. Character design as bridging tools of ideological message in game. In *2016 1st International Conference on Game, Game Art, and Gamification (ICGGAG)*, pages 1–4, 2016.