

# Game-Based Platforms for Studying Virtual Agent Believability

**Christianah Titilope Oyewale**

School of Interactive Arts and Technology, Simon Fraser University  
13450 102 Avenue  
Surrey, BC V3T 5Y1 Canada  
christianah\_oyewale@sfu.ca

## Abstract

Believable virtual agents are central to many interactive systems, yet current methods for evaluating user trust and perception often rely on limited, decontextualized tools. My research investigates how narrative-driven, game-based environments can be used to study users' perceptions of virtual agent believability, particularly in relation to gender presentation and role-based behavior. I am developing an interactive platform where users engage with male, female, and androgynous agents embedded in branching scenarios that simulate trust-based decision-making. The platform incorporates adaptive storytelling, real-time analytics, and machine learning to capture both behavioral and attitudinal responses. By gamifying the evaluation process, the research aims to produce richer, more ecologically valid data than traditional surveys. This work draws from human-computer interaction, affective computing, and AI ethics to explore how design choices, such as agent role, gender, and interactivity, affect user judgments of reliability and integrity. As an early-stage PhD student, I am currently refining the research questions, platform design, and experimental methods. Feedback from the AIIDE community will help sharpen the research focus, improve methodological choices, and ensure broader relevance across virtual agent applications.

## Introduction

Virtual agents are becoming increasingly central to interactive media, games, and educational platforms (Nag and Yalçın 2020; Fitriani et al. 2019; Haake and Gulz 2008). As their roles expand, so does the importance of designing agents that users perceive as psychologically and socially believable (McKnight, Choudhury, and Kacmar 2002; Nass and Moon 2000). This prospective research explores how game-based environments, infused with narrative and decision-making mechanics, can be used to study user perceptions of agent gender, behaviour, and believability. Early efforts have used survey tools or static interactions, but these often lack the richness and realism of gameplay (Nag and Yalçın 2020). This study's aim is to design an interactive platform that gamifies data collection, engaging users in story-driven tasks and role-play scenarios to study behavioral responses toward varied virtual agent presentations.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

## Research Objectives

This study aims to explore how narrative roles and game scenarios affect the perceived believability of virtual agents. It investigates the emergence of gender-based biases or expectations during gameplay and examines whether gamified platforms can generate more authentic and immersive data compared to traditional survey methods. The research approach integrates elements of game design, social cognition, and human-computer interaction (HCI), positioning users in meaningful, context-rich interactions with virtual agents that simulate real-world dynamics and expectations.

## Literature Review

Research into virtual agents has increasingly emphasized the importance of narrative structures and character roles in shaping users' perceptions of believability and authenticity (Poivet, Pelachaud, and Auvray 2023). Narrative framing often provides cues that users rely on to interpret agent behavior, with roles such as mentor, peer, or antagonist subtly guiding emotional responses and trustworthiness assessments (Moreno et al. 2002).

Gender-based bias in virtual environments has also received growing attention. Studies by Eyssel and Hegel, and Armando, Ochs, and Régner show that users often project stereotypes onto agents, with perceived competence and warmth varying according to the agent's gender presentation. These biases can influence decision-making, engagement levels, and even data quality in user studies.

Gamification, as a methodological tool, has emerged as a promising avenue for capturing deeper cognitive and affective responses (Ishaq et al. 2024; Sailer and Homner 2020; Mullins and Sabherwal 2020; Hui and Mahmud 2023). Unlike traditional survey techniques that may induce social desirability bias, immersive platforms encourage spontaneous interactions in context-rich settings, which better reflect real-world dynamics (Nag and Yalçın 2020).

Taken together, these strands suggest that narrative structure, gender dynamics, and platform interactivity are not peripheral elements, they're central to both the user experience and the integrity of the data collected.

## Methodology and Platform Design

This study will build upon an existing study by Nag and Yalçın, featuring three pre-validated virtual agents: Male, Female, and Androgynous. These agents will be embedded within gamified, narrative-driven scenarios designed to investigate how users perceive and trust virtual agents across varying roles and contexts. Each participant will interact with all three agents, with agent sequences presented in a randomized and counterbalanced format to mitigate order effects.

Participants will engage in a modified trust game, during which they evaluate each agent using a 12-item adaptation of the Trust in Automation Scale (Jian, Bisantz, and Drury 2000). This instrument assesses perceived reliability, competence, and integrity on a 7-point Likert scale, with contextual phrasing tailored to refer to "agents" rather than generic systems. Sample items include "I am confident in the agent" and "The agent has integrity."

In parallel, users will be asked to estimate the gender identity of each agent using a slider-based scale ranging from 0 (Male) to 1 (Female), with 0.5 representing an Androgynous presentation. These measures will yield both affective and cognitive data to explore how narrative framing, agent design, and gamified interaction influence users' gender perceptions and trust dynamics.

To support both experimental validity and user engagement, the study will leverage a custom-built platform with several integrated modules. These include a user interaction layer featuring gamified narrative scenarios, dynamic dialogues tailored to highlight gender-based decision-making patterns, and visually differentiated agents to elicit varied responses.

The data collection engine will log interaction paths, decision metrics, and trust evaluations. It will also integrate machine learning models to detect patterns in user interaction, trust evolution, and bias indicators. These insights will inform real-time tagging of behaviours, enabling finer-grained analytics and supporting deeper interpretations of participant behaviour.

Gamification mechanisms such as score tracking and optional replay arcs will enhance user immersion and produce richer, more spontaneous data over time.

On the back end, secure database management and real-time server integration will facilitate streamlined analysis through a visualization and analytics portal. This will include heatmaps, timelines, and correlation matrices, supported by AI-driven clustering and classification algorithms that group interaction profiles and highlight emergent behavioral types. A dedicated automated inference engine will compute preliminary statistical results, such as correlations between participant gender and agent trust scores or gender perceptions, and use LLMs to generate human-readable interpretations of these findings. This will provide researchers with early hypotheses and summary insights before conducting deeper inferential or multivariate analysis.

Predictive models may also be incorporated to simulate possible user-agent interaction trajectories based on prior data.

To complement these analytical features, an embedded ethics and consent framework will ensure transparency, including opt-in forms and briefing screens that inform users of research objectives and data usage protocols. A future-proof consent module may use AI to dynamically tailor study explanations based on user engagement levels, making ethics more digestible and personalized. The platform will be tested with human participants, and ethical approval will be sought from the Research Ethics Board (REB) at Simon Fraser University, British Columbia, Canada.

The platform may utilize technologies such as React.js or Unity WebGL (front-end), Node.js/Express or Django (back-end), and PostgreSQL or MongoDB (database), with visualization powered by D3.js or Plotly. All components will adhere to modern security and anonymization standards. In addition to functional modules, the platform's interface will prioritize accessibility and ease of use for researchers across disciplines. Design considerations include high-contrast color schemes, intuitive navigation structures, and responsive layouts for mobile and tablet devices. These features aim to reduce cognitive load and promote seamless interaction, supporting both novice and expert users.

## Expected Results

Based on prior literature and the multi-agent architecture of the study, we anticipate several key patterns to emerge from participant interaction and evaluation data:

- **Gender-Trust Congruency** Female participants may be more likely to identify and trust female agents, particularly in scenarios requiring collaboration or emotional intelligence. This may reflect shared identity effects on perceived reliability and integrity.
- **Androgynous Agent Perceptions** We expect a broader distribution in slider responses for the androgynous agent, especially among participants who identify outside the binary spectrum. This agent may serve as a mirror for flexible gender expectations, generating varied trust appraisals.
- **Narrative Amplification** The embedded storytelling mechanics are likely to deepen immersion, producing stronger trust signals overall. Participants may overlook agent gender in favor of behavior and role alignment within the story arc.
- **AI-Driven Insights** The platform's built-in statistical and interpretive module will surface early-stage results such as:
  - Trends in gender-attributed trust scores
  - Clustering of high-trust ratings around agents perceived as androgynous
  - Gendered interpretations tied to task framing (e.g., competitive vs cooperative roles)
- **Enhanced Usability for Researchers** Researchers will be able to instantly view these results via the platform's interactive analytics dashboard, which combines:
  - Dynamic filtering by participant attributes
  - Visualizations of agent perception metrics

## Static Visuals To Be Adapted For The Study

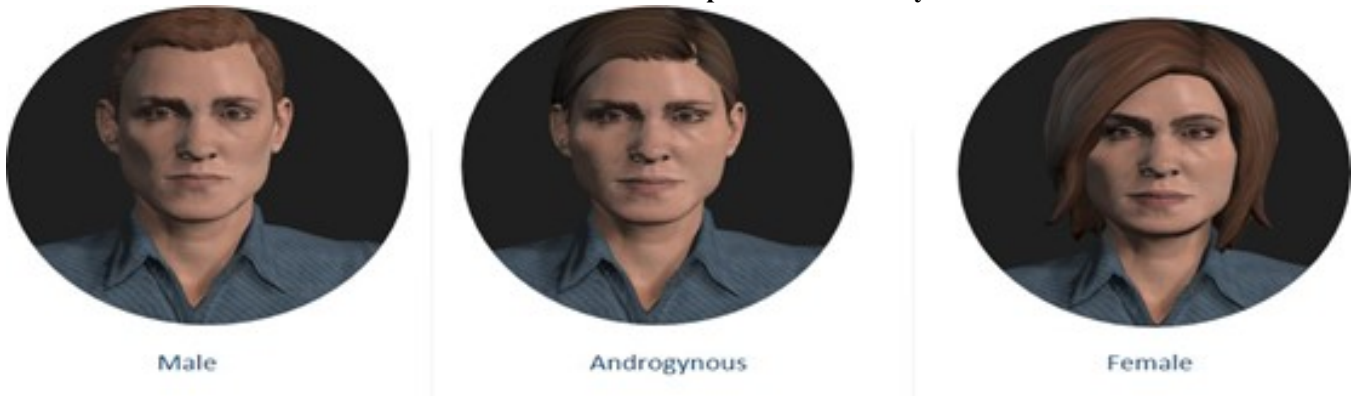


Figure 1: Visual designs based on Nag and Yalçın’s study (Nag and Yalçın 2020).

- Natural language interpretation powered by LLMs (e.g., “Female participants were more likely to rate female agents higher in trust-driven decisions.”) This reduces cognitive load, facilitates rapid insight generation, and enhances accessibility across disciplines—especially useful for mixed-method studies where researchers may not specialize in statistical modeling.

### Future Work and Feedback Goals

As the platform and study evolve, several avenues for refinement and expansion are envisioned:

- **Multimodal Agent Presentation** Future iterations may incorporate vocal and gestural elements into agent interactions, enabling more nuanced evaluation of trust and perception. This will allow exploration of how tone, body language, and speech patterns influence user-agent dynamics, especially in cross-cultural contexts.
- **Expanded Participant Sampling** Larger and more diverse participant pools, including gender-diverse and neurodiverse individuals, will strengthen generalizability and enrich insights into how personal identity affects agent perception. Recruitment may extend to international cohorts and use stratified sampling methods to ensure balanced representation.
- **Real-Time Adaptive Scenarios** Leveraging reinforcement learning, future builds may enable agents to respond dynamically to user choices, creating personalized story arcs. This approach will support deeper investigation into trust formation, bias shifts, and decision-making in evolving contexts.
- **Feedback-Driven Platform Enhancements** Participant input will be used to iteratively refine interface design, scenario clarity, and agent presentation. Integrated feedback prompts—activated post-interaction; will solicit user perspectives on engagement, ease of use, and perceived realism. These insights will guide updates to interaction flow, accessibility standards, and narrative structure.

- **Open Science and Modular Reusability** To support reproducibility and interdisciplinary collaboration, core platform components may be modularized for open-source release. Researchers will be able to plug in their own agent designs, trust measures, or narratives, accelerating innovation across fields such as HCI, psychology, and AI ethics.
- **Educational Integration and Applied Use** The system may be adapted for use in classrooms, ethics workshops, and tech literacy programs. By framing human-agent trust as a learnable and interrogable process, the platform could help students and professionals navigate AI interactions more critically.

These goals will be driven by a cyclical refinement model, where community feedback, research findings, and usability data converge to shape new directions. This aligns the platform not only with scientific rigor but with the evolving needs of researchers and participants alike.

### Conclusion

This research explores how game-based, narrative environments can serve as powerful tools for studying the believability and perceived trustworthiness of virtual agents. By integrating gender dynamics, role-based framing, and real-time analytics into a single interactive platform, the work aims to address key limitations in traditional evaluation methods. As I continue developing the system and refining the study design, I seek guidance on how to best frame the research questions, validate the approach, and ensure its relevance to broader conversations in AI and human-agent interaction. Participation in the AIIDE Doctoral Consortium will provide valuable mentorship and critical feedback to help steer the direction of this work and strengthen its impact across the fields of believable agents, interactive storytelling, and affective computing.

### References

Armando, M.; Ochs, M.; and Régner, I. 2022. The impact of pedagogical agents’ gender on academic learning: A systematic review. *Frontiers in Artificial Intelligence*, 5: 862997.

- Eyssel, F.; and Hegel, F. 2012. (S)he's Got the Look: Gender Stereotyping of Robots 1. *Journal of Applied Social Psychology*, 42: 2213–2230.
- Fitriani, S.; Bruijnes, M.; Richards, D.; Abdulrahman, A.; and Brinkman, W. P. 2019. What are we measuring anyway?—A literature survey of questionnaires used in studies reported in the intelligent virtual agent conferences. In *IVA 2019 - Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 159–161. Association for Computing Machinery, Inc. ISBN 9781450366724.
- Haake, M.; and Gulz, A. 2008. International Forum of Educational Technology Society Visual Stereotypes and Virtual Pedagogical Agents. *Source: Journal of Educational Technology Society*, 11: 1–15.
- Hui, H. B.; and Mahmud, M. S. 2023. Influence of game-based learning in mathematics education on the students' cognitive and affective domain: A systematic review. *Frontiers in psychology*, 14: 1105806.
- Ishaq, K.; Alvi, A.; ul Haq, M. I.; Rosdi, F.; Choudhry, A. N.; Anjum, A.; and Khan, F. A. 2024. Level up your coding: a systematic review of personalized, cognitive, and gamified learning in programming education. *PeerJ Computer Science*, 10: e2310.
- Jian, J.-Y.; Bisantz, A. M.; and Drury, C. G. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4: 53–71.
- McKnight, D. H.; Choudhury, V.; and Kacmar, C. 2002. Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13: 334–359.
- Moreno, K. N.; Eck, R. N. V.; Person, N. K.; Jackson, G. T.; Adcock, A. B.; and Marineau, J. C. 2002. Etiquette and Efficacy in Animated Pedagogical Agents: The Role of Stereotypes.
- Mullins, J. K.; and Sabherwal, R. 2020. Gamification: A cognitive-emotional view. *Journal of Business Research*, 106: 304–314.
- Nag, P.; and Yalçın, N. 2020. Gender Stereotypes in Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA 2020*. Association for Computing Machinery, Inc. ISBN 9781450375863.
- Nass, C.; and Moon, Y. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56: 81–103.
- Poivet, R.; Pelachaud, C.; and Auvray, M. 2023. The influence of conversational agents' role and behaviors on narrative experiences. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, 1–4.
- Sailer, M.; and Homner, L. 2020. The gamification of learning: A meta-analysis. *Educational psychology review*, 32(1): 77–112.