LoveSims: Exploring 'What-If' Scenarios for Relationship Insights and Compatibility

video demo

Mateo Larrea, Xingyi Zhang, Xuyang Zhu
Stanford University

Introduction

Dating technologies shape romantic connections, but often rely on algorithmic compatibility scores that reinforce biases and reduce relationships to superficial feature-matching, fostering a "shopping culture." Current apps require users to navigate early interactions through trial and error, neglecting deeper long-term relationship dynamics. While recent advances in generative agents offer promising new possibilities, existing systems still lack structured long-term progression and real-time feedback.

To address this, we introduce ***LoveSims*, a generative agent framework that shifts from matchmaking to meaningful relationship insights by simulating long-term interactions and integrating feedback loops.**

Methodology

We created 20 synthetic agents and made them to interact with each other under given scenarios (e.g. café dates, museum visits). Then generate:

- **Self-Reflection**: Agents critically reflect post-interaction, assessing compatibility and feelings.
- **Third-Party Evaluation**: Independent AI model evaluates interactions based on objective metrics (flow, connection, interests).

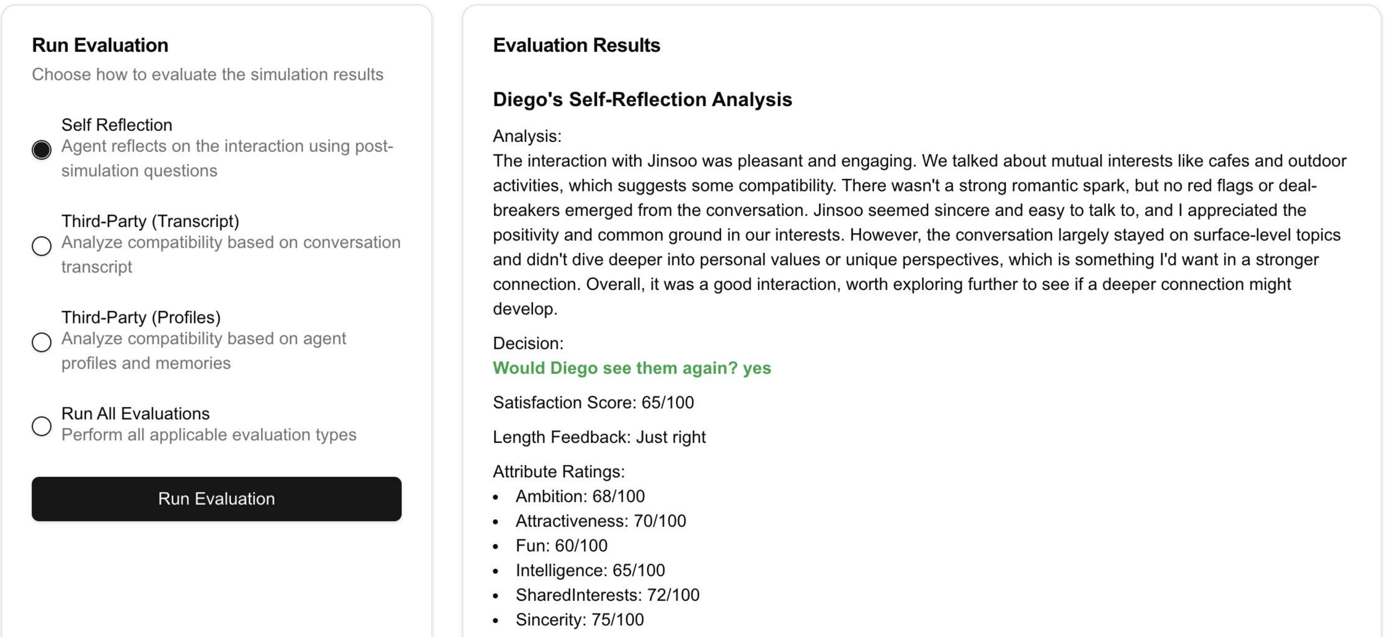
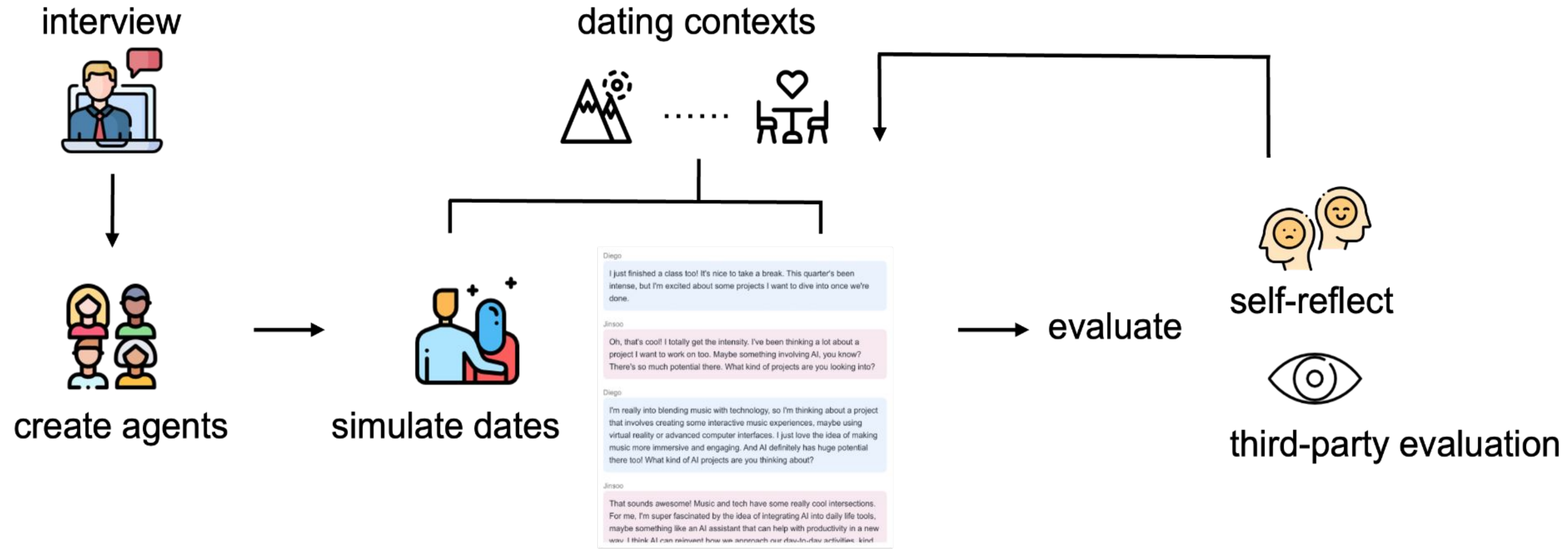


Figure 1. Simulation interface



Results

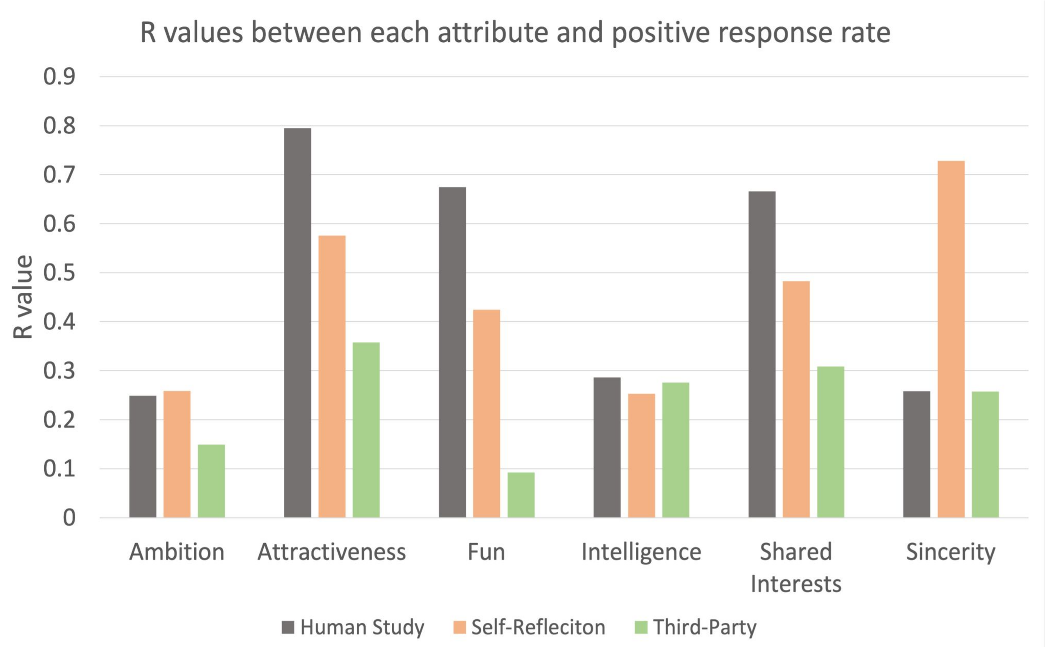


Figure 2. Aggregated correlations

Table 1. Consistency

Metrics Type	Consistency Rate (%)	Majority Response Rate (%)
Self-reflection	89.29	96.76
Third-party	75.62	92.38

- We calculated the correlation between the six attribute scores received by an agent and the decisions made by their dates. Both human experiments and simulations exhibit similar patterns (Figure 3).
- Figure 2 presents the average R values across all contexts. Compared to third-party evaluation, self-reflection better captures human preferences, highlighting the effectiveness of role-playing.
- Agents demonstrated high consistency in their dating responses, particularly during self-reflection (Table 1).

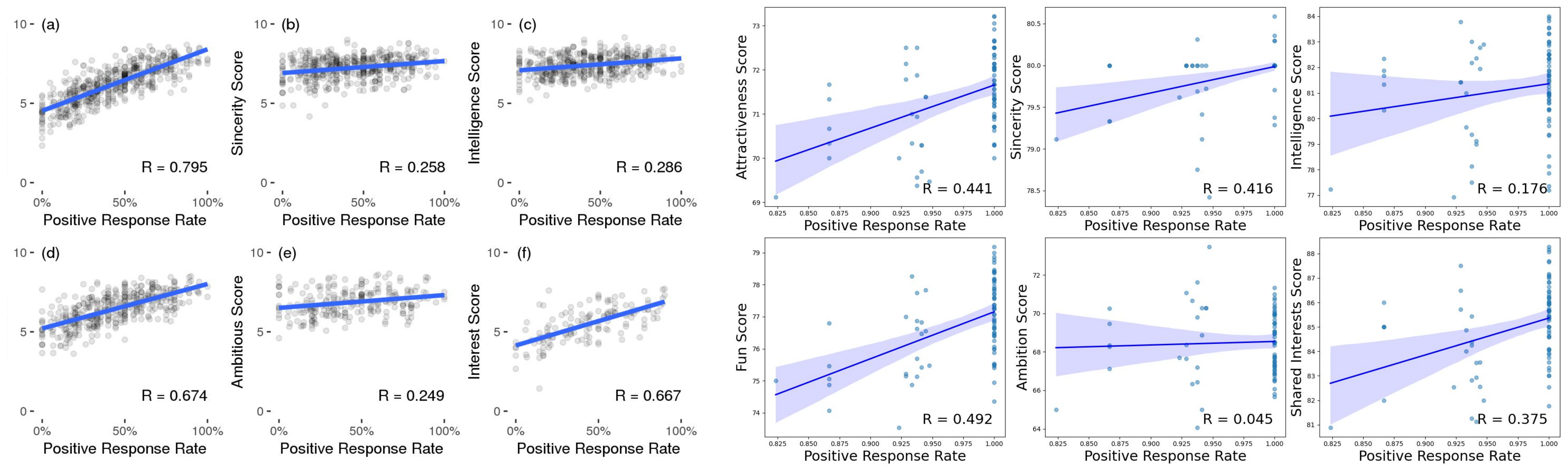


Figure 3: Distributions of positive response rates versus attribute scores. Left: real-person experiments [1]. Right: simulations, self-reflection metrics.

Discussions & Future Work

Limitations

- LLMs have inherent bias towards politeness and agreeability in generative agents, which leads to an overrepresentation of positive responses that may not reflect real-world dynamics.
- LLMs struggle with capturing complex human social behaviors, such as non-verbal cues and emotional fluctuations, which are vital in romantic decision-making.

These limitations result in interactions that feel static or overly rational, lacking the emotional depth and unpredictability of real relationships. Moreover, inconsistencies between self-reflection and third-party evaluations further raise concerns about the reliability of LLM-based assessments in ambiguous social scenarios.

Avenues for improvement

- Enhance prompt engineering for greater emotional and contextual variability.
- Refine data collection to build more accurate agent personas.
- Integrate affective states, allowing agents to simulate emotions like empathy and frustration.
- Improve memory stream architecture, enabling agents to recall and adapt from past interactions for more dynamic simulations.

To validate *LoveSims*'s effectiveness, we need real-world user studies to let the model evolve through authentic behavioral feedback. We also need to consider expanding the system's cultural adaptability and addressing ethical concerns surrounding the use of sensitive data.