MIMIC: Integrating Diverse Personality Traits for Better Game Testing Using Large Language Model

[1] **Yifei Chen**, [2] Sarra Habchi, [1] Lili Wei

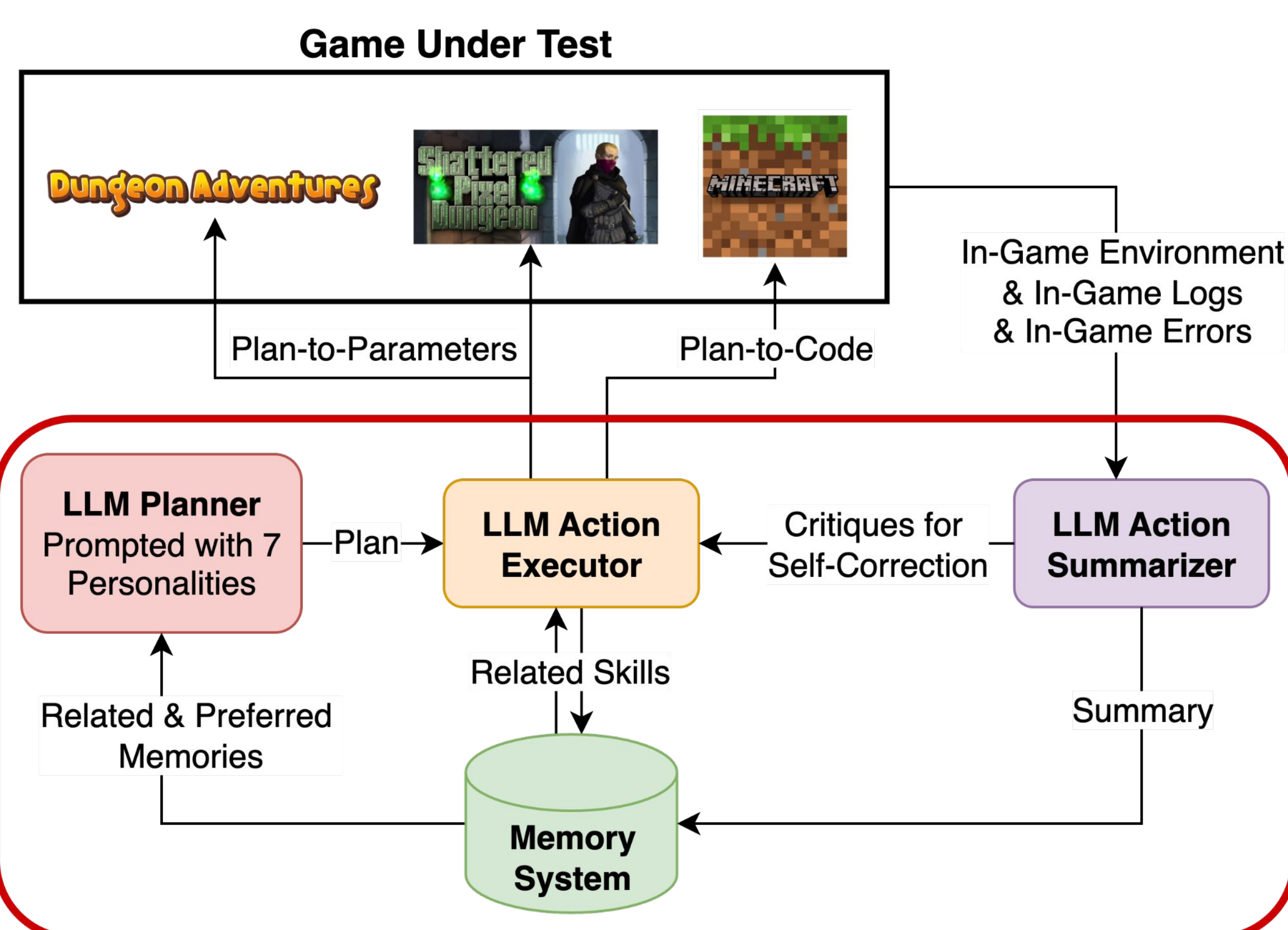[1] McGill University, Montréal, Canada, [2] Cohere, Canada

Introduction

- **MIMIC** brings human-like diversity into automated game testing.
- By integrating **personality traits** into **Large Language Model (LLM)-based agents,** it adapts flexibly across multiple open- and closed-source games.
- MIMIC achieves **superior coverage** and **solution diversity** over SOTA baselines, bridging AI testing and realistic human playstyles.

Motivation

- The global video game market reached **USD 298.98 Billion** in 2024 [1], emphasizing the growing need for effective quality assurance.
- Manual playtesting is both **costly and time-consuming**, motivating the use of **automated, agent-based testing,** where agents play and evaluate games autonomously.
- However, most agents focus on task optimization, producing **repetitive and homogeneous behaviors** under similar scenarios, unlike real players who exhibit **diverse, personality-driven strategies**.

Approach

Game Under Test

MIMIC

- In each iteration, the system **plans, executes, and reflects**, forming a closed feedback loop for goal-driven game testing.
- The **Planner** is prompted with **seven predefined personalities** [2], *Achievement, Adrenaline, Aggression, Caution, Completion, Curiosity, and Efficiency*, to shape decision styles.
- Guided by personality and retrieved memories, the **Planner** generates action plans suited to the current in-game context.
- The **Executor** interprets these plans into concrete in-game actions or code/parameter operations, interacting directly with the game under test.
- The **Summarizer** observes outcomes, errors, and logs, producing feedback and critiques for self-correction.
- These summaries are stored in the **Memory System** as new experiences, allowing future planning to adapt dynamically.
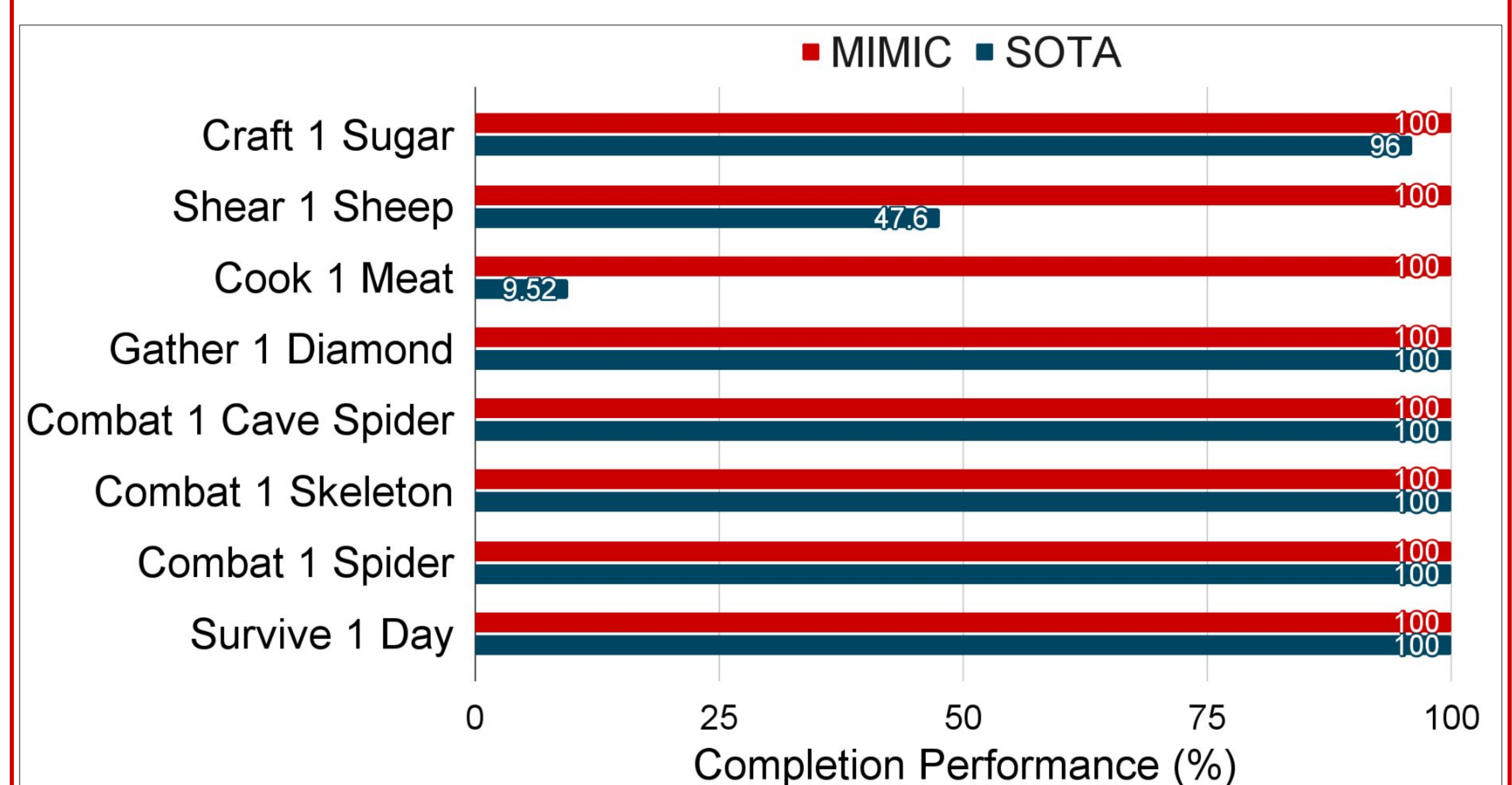
Results

Fig. 1. Task completion performance of MIMIC and the state-of-the-art (SOTA) agent in Minecraft.

- MIMIC is more effective at solving tasks than ODYSSEY, the SOTA LLM-based agent, in Minecraft (Fig. 1).
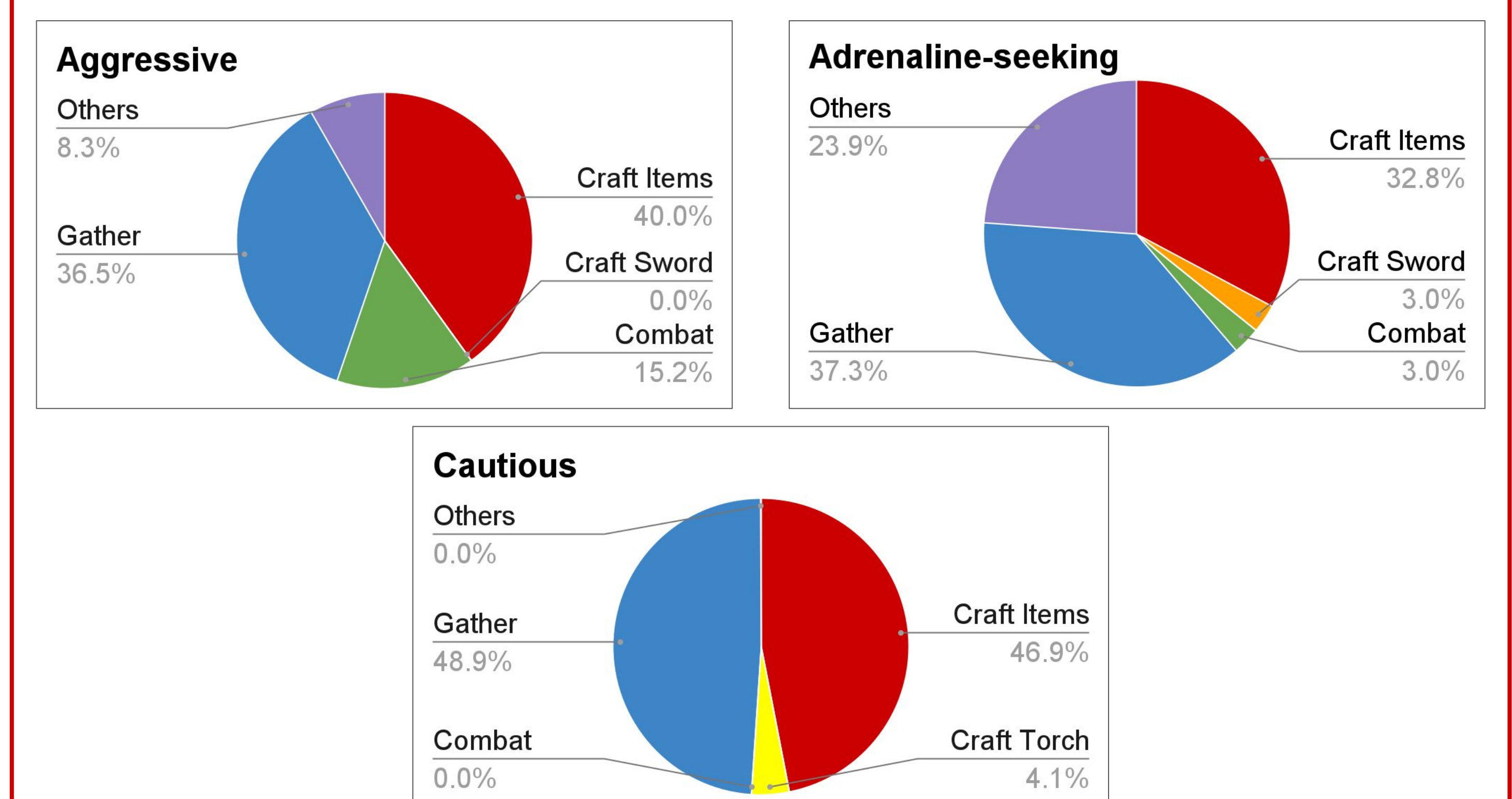

Fig. 2. Distribution of action types in solutions generated by three representative MIMIC agents with distinct personality traits for the task "Gather 1 Diamond".

Behavioral Comparison (Fig. 2)

- **Aggressive agent:** 15.2 % of actions involve combat, whereas others average only 2.0 %.
- **Adrenaline-seeking agent:** actively crafts swords and explores high-risk areas to engage enemies.
- **Cautious agent:** avoids combat entirely and is the only one that crafts torches before mining.
- **SOTA agent:** performs none of these behaviorally distinctive actions.

Conclusion & Future Work

- By emulating the diversity of human playstyles, **MIMIC** advances automated testing toward **real-world applicability** in the game industry.
- Personality-driven agents demonstrate **human-like adaptability** and provide new opportunities for **behaviorally rich game evaluation**.
- This paradigm can also extend to **User-Interface (UI) testing** and **Human-Computer Interaction (HCI)**, where diverse behavioral paths reveal edge cases often missed by traditional tools.

Reference

[1] Grand View Research. 2025. Video Game Market (2025–2030): Size, Share & Trends Analysis Report by Device (Console, Mobile, Computer), by Type (Online, Offline), by Region (North America, Europe, Asia Pacific, Latin America, Middle East & Africa), and Segment Forecasts. Grand View Research, San Francisco, CA.
https://www.grandviewresearch.com/industry-analysis/video-game-market

[2] S. Stahlke, A. Nova, and P. Mirza-Babaei. 2020. Artificial Players in the Design Process: Developing an Automated Testing Tool for Game Level and World Design. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20). Association for Computing Machinery, New York, NY, USA, 267–280. https://doi.org/10.1145/3410404.3414249