**University of Leicester**

**Computer Science GTA project**

|  |  |
| --- | --- |
| **First Supervisor** | Dr Fabricio Goes |
| **School** | **Computer Science** |
| **Email** | [fabricio.goes@leicester.ac.uk](mailto:fabricio.goes@leicester.ac.uk) |

|  |  |  |
| --- | --- | --- |
| **Project Title** | Generative Agents as Autonomous Evaluators of Games | |
| **Project Highlights:** | 1. | LLM-based Game Generation Frameworks |
| 2. | Autonomous Agent-based Game Testing |
| 3. | Creative Evaluation of Generated Games by Generative Agents |
| **Project Summary** | | |
| Recent advancements in Large Language Models (LLMs) like Claude 3.7 and o3 have demonstrated their ability to generate not just text but also complex creative artifacts such as games. These models can now produce game designs, mechanics, and code for simple games including Atari-style games. However, evaluating the creativity and quality of these generated games remains a significant challenge in the field of Computational Creativity.  In game design, creative artifacts are typically assessed through a combination of expert evaluation and playtesting. Games are considered creative when they present novelty (unique mechanics or experiences) and value (engagement, fun, challenge). Currently, human evaluators are required to play and assess these games, which creates bottlenecks in the generation-evaluation cycle and introduces subjective biases.  Despite the importance of evaluation, current approaches rely heavily on human judges who may have inconsistent criteria or lack expertise in game design principles. Even when using playtesting metrics like engagement time or completion rates, human interpretation is needed to contextualize this data into creativity assessments.  In this proposal, we aim to develop a novel method where generative agents serve as autonomous evaluators of LLM-created games. Our approach consists of three interconnected components: (1) generating Atari-style games using various LLM models and prompting techniques, (2) deploying autonomous agents to play and test these games, collecting behavioural and performance data, and (3) using generative agents to analyse this playtest data and evaluate the creativity of the generated games.  By creating a closed-loop system where generative agents both test and evaluate LLM-created games, we can establish more consistent, scalable, and objective creativity metrics. This approach minimizes human participation in the evaluation process while accelerating the iteration cycle of game generation. The project will contribute to the emerging field of AI-driven game design by developing frameworks for autonomous creativity evaluation that could eventually apply to more complex interactive experiences beyond Atari games. | | |