香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

数据科学与分析学域
DATA SCIENCE AND ANALYTICS THRUST
信息枢纽
INFORMATION HUB

DSA Research Experiences for Undergraduates

Research Project

Section1: Faculty Information

Full Name	LUO, Yuyu	Tel	13717902271
Thrust/Hub	DSA/INFH	Office	E2-615
Email	yuyuluo@hkust-gz.edu.cn		

Section2: Research Project Proposal

Project Title	Exploring Reasoning Models as Textual World Models
Project Description (max 800 words)	The rise of large language models (LLMs) has transformed artificial intelligence, enabling the creation of LLM Agents—systems designed to not only process and generate text but also interact with real-world environments through reasoning and decision-making. Unlike traditional language models focused on dialogue or text completion, LLM Agents are increasingly tasked with executing actions, such as running terminal commands, solving mathematical problems, or simulating interactions with user interfaces (UIs). This project explores the potential of LLM Agents to serve as Textual World Models (TWMs)—frameworks that build internal representations of the world based on textual input and use them to predict outcomes or execute tasks. Specifically, it investigates whether these agents can accurately reason about practical scenarios, such as interpreting command-line outputs, verifying mathematical computations, or anticipating the results of UI interactions on a smartphone. Tailored as an accessible undergraduate research project, this study combines theoretical exploration with hands-on experimentation to assess the capabilities and limitations of LLM Agents in simulating real-world understanding through text. The core question is: Can LLM Agents function as TWMs by reasoning about practical scenarios described in text, such as terminal operations, math problems, or UI navigation, and predict their outcomes reliably? For instance, if given a prompt like "Execute ls -l in a terminal," can the agent infer that this is a Linux command listing files (despite ls not applying to all systems) and describe the expected output? Or, if asked, "Subtract 15 from 42," can it compute the correct result and explain its reasoning? Similarly, for a scenario like "Tap the 'Settings' icon on a phone UI," can it predict subsequent options (e.g., "Wi-Fi" or "Sound") based on common phone layouts? These examples test the agent's ability to model environments textually and reason causally, bridging abstract knowledge with actionable outcomes.

	The project unfolds in three phases. First, the student will review literature on LLM Agents (e.g., models like GPT or specialized frameworks like AutoGPT) and the concept of TWMs, focusing on how these systems encode practical knowledge versus theoretical reasoning. This will clarify the gap between language understanding and task execution. Second, the student will select an accessible LLM Agent (e.g., via Hugging Face or an API like OpenAI's) and design textual scenarios to test its TWM potential. These will include: (1) terminal tasks (e.g., "Run cd dir1 then pwd—what's the result?"), (2) math problems (e.g., "If a store discounts a $50 item by 20%, what's the final price?"), and (3) UI interactions (e.g., "Swipe left on a phone lock screen—what happens next?"). The student will input these prompts, record the agent's responses, and evaluate them for accuracy, coherence, and contextual awareness. Third, the student will synthesize findings into a report, assessing whether the agent behaves as a TWM and identifying errors—like misinterpreting commands, miscalculating, or failing to generalize UI patterns. This project is ideal for undergraduates because it uses relatable, real-world scenarios and requires only basic tools (e.g., Python, a text editor, and internet access), not advanced programming or hardware. The focus is on qualitative and observational analysis—e.g., does the agent "understand" a terminal's file system or a phone's navigation?—making it manageable within a short timeframe (e.g., a semester). Students will learn research skills like hypothesis testing, data collection, and critical evaluation, while engaging with a timely AI topic. They can also personalize the project by choosing scenarios relevant to their interests, such as coding, math, or mobile app design. Expected outcomes include insights into how LLM Agents process practical knowledge and whether they can simulate world-like reasoning in text-based tasks. Limitations may surface, such as reliance on training data (e.g., confusing Windows and Linux commands) or struggles with dynamic UI contexts. These findings could spark further questions, like enhancing agents with real-time feedback or multimodal inputs, but the project's simplicity ensures it remains an approachable introduction to AI research. By exploring LLM Agents in actionable contexts, this study offers a fresh perspective on their potential as TWMs while equipping students with foundational scientific experience.
Proposed Research Duration	Start Date: _____2025__ / ____03____ / ___20_____ End Date: _____2025_ / _____08___ / ___30_____
Student/Researcher Duties	● Conduct a literature review on reasoning models and world models.

	● Select and set up a reasoning model for experimentation using available computational tools. ● Design and test 5-10 textual scenarios to evaluate the model' s reasoning and world-representation capabilities. ● Document model inputs, outputs, and observations systematically. ● Analyze results to assess the model' s strengths and limitations as a Textual World Model. ● Write a concise report (±2000 words) summarizing findings and reflections. ● Present the project orally (e.g., 10-minute presentation) to peers or a supervisor. ● Write and submit a research paper to the top-tier conference.
Technical Skills Required	☑ Python ☑ Machine Learning ☐ Big Data ☐ R ☑ Deep Learning ☐ SQL ☐ C/C++ ☐ Other: _____
Preferred Student/Researcher Background	The ideal student for this project is self-motivated and eager to explore the intersection of AI and knowledge representation. Basic familiarity with programming (e.g., Python) is helpful but not required, as the project can be adapted to the student's skill level with proper guidance. A curiosity about how machines "think" and a willingness to learn through trial and error are essential. No advanced coursework in AI or mathematics is necessary—enthusiasm and initiative will drive success
Maximum Number of Students/Researchers	☐ 1 ☑ 2

Section3: Pre-Application Research Exposure Meeting

Faculty members are encouraged to schedule a Research Exposure Meeting to introduce students to their projects.

Preferred Date	
Preferred Time	
Meeting Mode	☐ In-Person ☐ Online
Venue (if in-person)	
Meeting Link (if online)	