Action Model-Based Natural Language Control System for Digital Twins

Younghwan Jeong
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
cjstntjd@keti.re.kr

Won Gi Choi
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
cwk1412@keti.re.kr

Taemin Hwang
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
taemin.hwang@keti.re.kr

Jin-Young Lee
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
jylee@keti.re.kr

Sang-Shin Lee *
Autonomous Intelligent System Research Center
Korea Electronics Technology Institute
Gyeonggi-do, Republic of Korea
sslee@keti.re.kr

***Abstract*— Digital twins, which enable seamless interaction between the physical and virtual worlds for simulation and control, have garnered significant attention across various domains. However, operating and managing digital twins beyond mere monitoring typically requires substantial expertise and effort, presenting a high barrier to entry for non-experts. In this study, we propose an Action Model-Based Digital twin control system that allows non-specialists to intuitively operate digital twins. By leveraging a large action model, the system interprets user intent from natural language commands and dynamically controls key functions of the digital twin without the need for domain-specific knowledge.**

Keywords—Digital twin, Large Action Model, Action planning

I. Introduction

Digital twins precisely replicate physical objects or systems in virtual environments, enabling real-time monitoring, simulation, prediction, and control [1]. However, most digital twin systems are designed with domain-specific, rigid control interfaces based on statically defined functions, making them closed and inflexible. This structure poses significant challenges for non-expert users, who find it difficult to intuitively operate the system, integrate diverse functionalities, or execute complex task scenarios. Consequently, such systems lack the flexibility to accommodate diverse user intents and fail to adapt effectively to dynamic situations.

Recently, increasing attention has been given to Large Action Models (LAMs) in the field of artificial intelligence. LAMs are designed to combine traditional language modeling with action reasoning and execution capabilities. These models dynamically interpret natural language commands, plan appropriate action sequences, and execute them, thereby enabling more robust and context-aware interaction [2][3].

In this study, we extend the capabilities of LAMs into the domain of digital twin control. We propose an intelligent dynamic control system that enables non-expert users to intuitively operate digital twins through natural language. The proposed system accurately infers user intent from natural language input and dynamically maps it to core digital twin functionalities. This approach significantly enhances the flexibility and adaptability of digital twin systems, improving overall usability and accessibility.

To evaluate the system, we constructed a dataset comprising natural language commands that reflect a diverse range of linguistic expressions and semantic variations related to digital twin operations. These include realistic imperfections commonly found in practical user input, such as typographical errors, pronominal references, and omitted information. Experimental results demonstrate that the proposed LAM-based system successfully handles a wider range of queries compared to conventional keyword-matching methods and shows superior performance in processing complex command structures.

In summary, the contributions of this paper are as follows:

- We propose a LAM-based digital twin system that enables dynamic control via natural language interaction.
- We design an intent inference and execution architecture capable of handling unstructured and informal language expressions.

II. Proposed Architecture

This section describes the overall architecture and operational principles of each component in the proposed Action Model-Based Natural Language Control System for Digital Twins**.** The system is designed to infer user intent from natural language commands issued by non-expert users and to flexibly execute various tasks by interfacing with the core functionalities of the digital twin. The system architecture consists of three main modules: 1) Natural Language Command Processing Module, 2) LAM-Based Action Planning Module, 3) Digital Twin Integration Module.

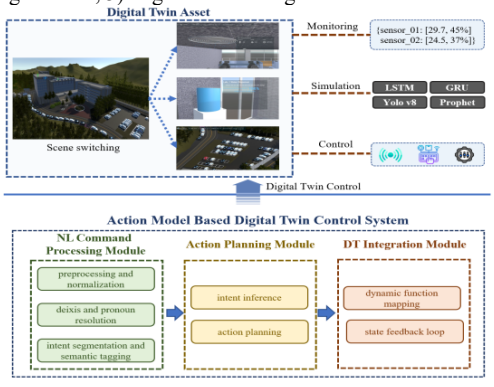

Fig. 1. The overall architecture of Action Model based DT control system.

A. *Natural Language Command Processing Module*

The **Natural Language Command Processing Module** handles diverse linguistic characteristics commonly found in user input, such as colloquial expressions, abbreviations, omissions, typographical errors, and pronouns. To normalize such inputs into a structured format suitable for downstream processing, this module performs a series of steps. First, it conducts preprocessing and normalization**,** including typo correction, removal of unnecessary punctuation, and standardization of non-canonical expressions. Next, it performs deixis and pronoun resolution through context-aware coreference analysis to interpret referential expressions. Finally, the module carries out intent segmentation and semantic tagging**,** decomposing compound commands into meaningful units and assigning semantic labels to each, which are then passed to the action planning module.

B. *LAM-Based Action Planning Module*

The LAM-Based Action Planning Module takes the processed input tokens and leverages the capabilities of the Large Action Model (LAM) to perform two key functions. First, it conducts intent inference, interpreting the user's query in context to identify which functionalities of the digital twin require control. Second, it performs action planning, generating a task sequence based on the inferred intent. This sequence typically consists of a series of API calls or event triggers necessary for executing the desired operation. The LAM is customized through the addition of predefined system prompts that augment its meta-reasoning capabilities, and it is fine-tuned to reflect the API specifications and operational requirements of the target digital twin environment.

C. *Digital Twin Integration Module*

The Digital Twin Integration Module connects the action sequences generated by the LAM to the digital twin system in real time. This integration layer performs several key functions. First, it enables dynamic function mapping, associating each action sequence with the appropriate API endpoints or control points within the digital twin. Second, it establishes a state feedback loop, collecting updated system states based on execution outcomes and, when necessary, providing follow-up commands or corrective feedback to the LAM. The module is designed to support diverse digital twin interfaces such as MQTT and REST, and can be adapted across domains through modular execution controllers for flexible deployment.

III. Simulation Results

This section presents the performance evaluation of the proposed method. The primary evaluation metric is the **query execution success rate**, assessed over a dataset of 100 natural language command instances. The command set was carefully constructed to reflect a variety of linguistic characteristics commonly encountered in real-world scenarios, including typographical errors, pronoun usage, omitted information, and compound sentence structures.

As shown in **Figure 2**, the proposed method achieved the highest success rate of **98%** in processing user commands. In contrast, the LLaMA-based baseline recorded a success rate of **59%**, with notable failures in handling misspellings, pronouns, and scenario branching. The keyword-based system performed the worst, achieving only **32%**, due to its limited ability to handle diverse and unstructured command formats. These results demonstrate that the proposed system excels in context-aware understanding and multi-functional execution, particularly in intent segmentation, inference of omitted details, and dynamic mapping to digital twin functionalities.

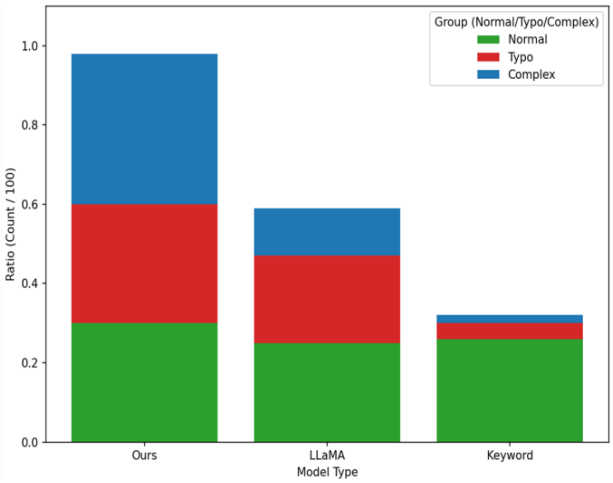

Fig. 2. Performance comparision of query execution success rate.

IV. Conclusion and Future Works

We propose a Action model-based digital twin natural language control system to ease accessibility to digital twins and maximize functional usability. Experimental results show that the proposed system integrates language understanding and action planning to provide flexibility and scalability for complex operations on digital twins in a more realistic environment. In the future, we plan to expand our research so that a digital twin system operating in real time can recognize complex situations and autonomously perform tasks according to user intentions.

Acknowledgment

This research was supported by Institute of Information & communication Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-II220545, Development of Intelligent digital twin object federation and data processing technology).

References

[1] D. M. Botin-Sanabria, A.-S. Mihaita, R. E. Peimbert-Garcia, M. A. Ramirez-Moreno, R. A. Ramirez-Mendoza, and J. d. J. Lozoya-Santos, "Digital Twin Technology Challenges and Applications: A Comprehensive Review," Remote Sens., vol. 14, no. 6, p. 1335, 2022.

[2] L. Wang, F. Yang, C. Zhang, J. Lu, J. Qian, S. He, and Q. Zhang, "Large Action Models: From Inception to Implementation," arXiv preprint arXiv:2412.10047, 2024.

[3] J. Zhang, T. Lan, M. Zhu, Z. Liu, T. Hoang, S. Kokane, and C. Xiong, "xlam: A family of large action models to empower AI agent systems," arXiv preprint arXiv:2409.03215, 2024.