Active Learning ⋆

Zhifeng Han * **Claire Walton** **

* *University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: zhifeng.han@utsa.edu).*
** *University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: claire.walton@utsa.edu).*

Abstract: As AI agents (LLMs and Robots) accumulate "crossed paths" of knowledge, internal world models become prone to "pre-convicted" illusions—internal biases that override reality. This framework proposes a Focus Mode that uses Active Learning to dynamically balance internal predictions with direct environmental interactions, ensuring the agent remains grounded in the "Working Sense."

Keywords: LLM, robotic, Reinforcement Learning, world model, Active Learning, Focus Mode, Working Sense

1. INTRODUCTION

We use world model to predict the future based on historical data. Hafner et al. (2019) world models allow for planning and behavior learning given only small amounts of real world interaction

2. BACKGROUND

2.1 Main challenges

3. ARCHITECTURE

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used NotbookLM and ChatGPT in order to help English as secondary language speaker. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

Hafner, D., Lillicrap, T., Ba, J., and Norouzi, M. (2019). Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*.

⋆ This work was supported in part by the U.S. National Science Foundation (NSF) under Award No. 1826086.