KAUST Artificial Intelligence Initiative
مبادرة كاوست للذكاء الاصطناعي

Mindstorms in Natural Language-Based Societies of Mind

@ The NeurIPS 2023 Workshop on Robustness of Zero/Few-Shot Learning in Foundation Models

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology


Mingchen Zhuge


Haozhe Liu


Francesco Faccio


Dylan R. Ashley

USI/SUPSI
IDSIA
Istituto Dalle Molle di studi sull'intelligenza artificiale


Róbert Csordás


Anand Gopalakrishnan


Abdullah Hamdi


Hasan Abed Al Kader Hammoud


Vincent Herrmann


Kazuki Irie


Louis Kirsch


Bing Li


Guohao Li


Shuming Liu


Jinjie Mai


Piotr Piękos


Aditya Ramesh


Imanol Schlag


Weimin Shi


Aleksandar Stanić


Wenyi Wang


Yuhui Wang


Mengmeng Xu


Deng-Ping Fan


Bernard Ghanem


Jürgen Schmidhuber

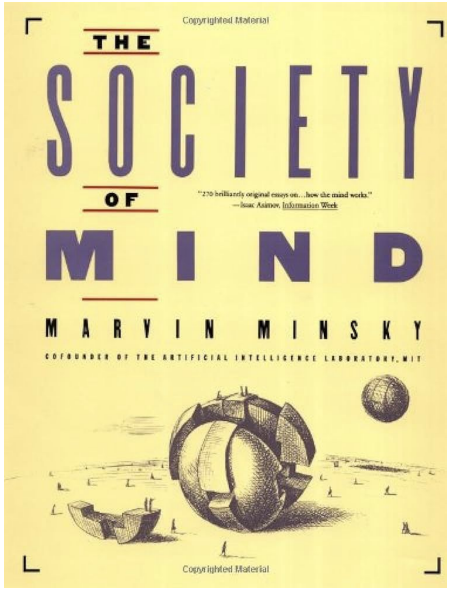
UNIVERSITY OF OXFORD

HARVARD UNIVERSITY

北京航空航天大学 BEIHANG UNIVERSITY 1952

ETH zürich

Minsky's Society of Mind

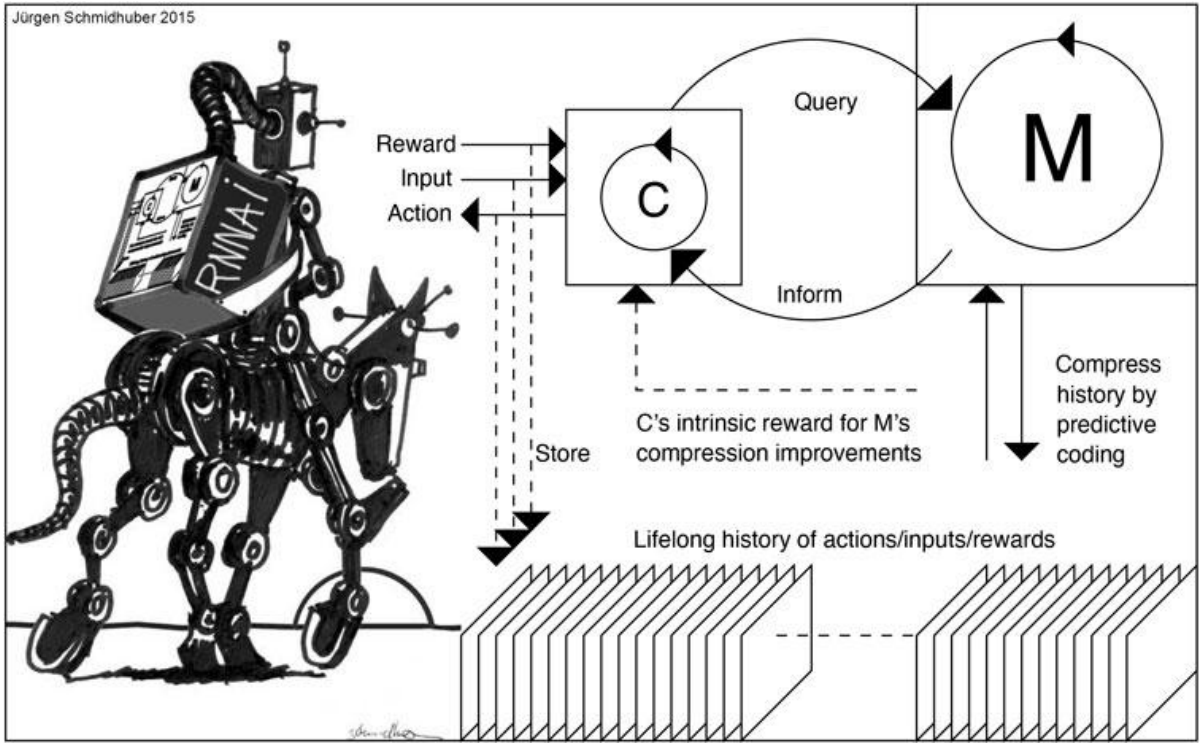


What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.

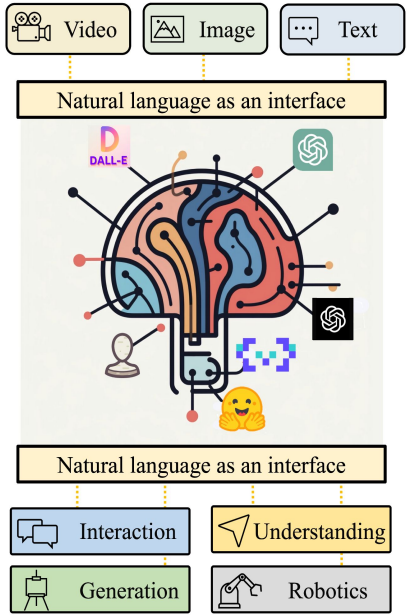
- Marvin Minsky in *The Society of Mind* (1958)

Learning to Think

- Two networks: A controller and a world model
- Controller learns in a free-form manner how to exploit the world model to maximize RL signal
- Kinda obvious architecture nowadays (e.g., foundation models) but pretty revolutionary back in 2015



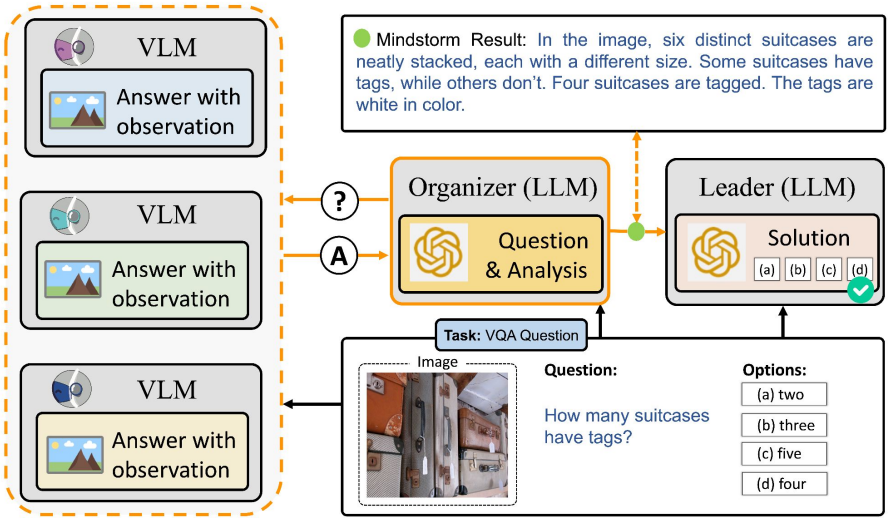
A Natural-Language–Based Society of Mind



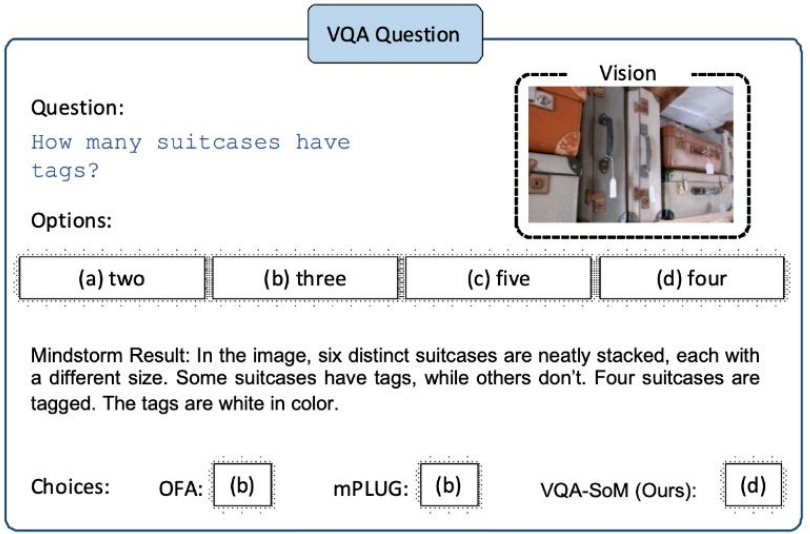
- An NLSOM consists of many heterogeneous agents, each acting according to their own objectives and communicating with one another primarily through natural language according to some organizational structure.
- To solve a problem, agents in an NLSOM communicate, each contributing their own expertise in a *Mindstorm*.

Visual Question Answering & Image Captioning

Architecture:

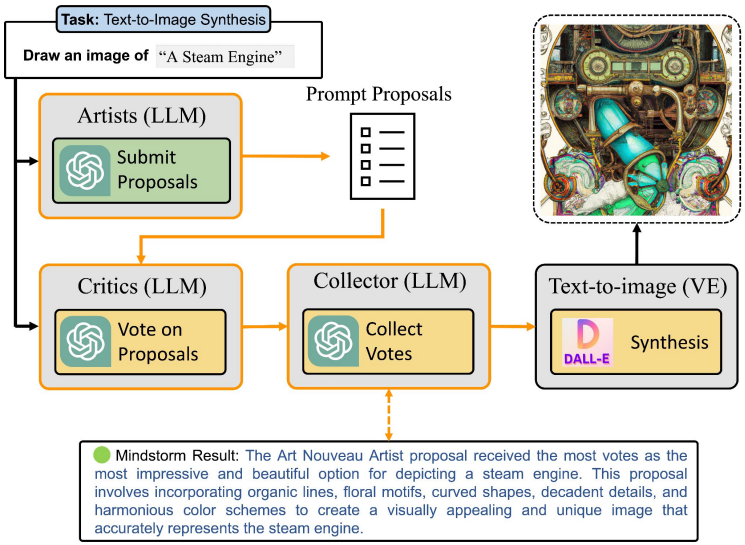


Result:

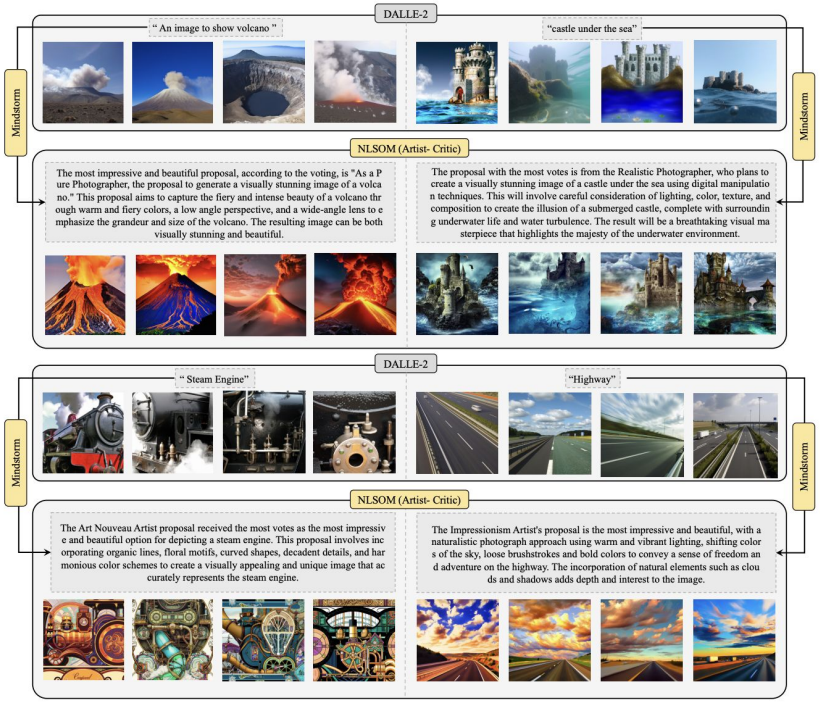


Prompt Generation for Text-to-Image Synthesis

Architecture:

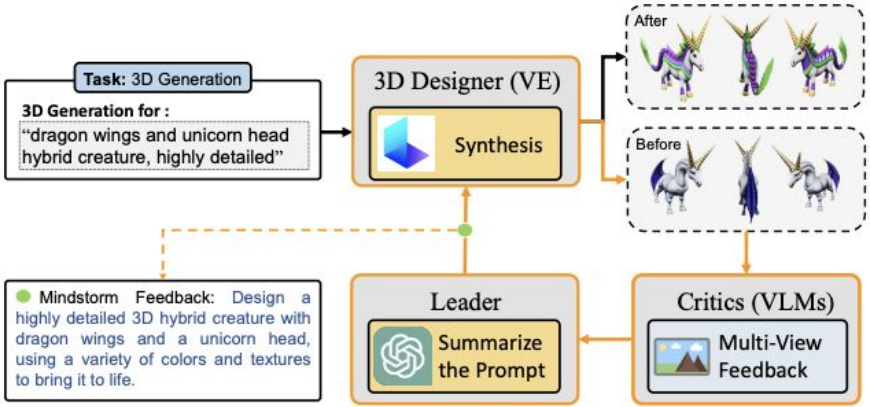


Result:

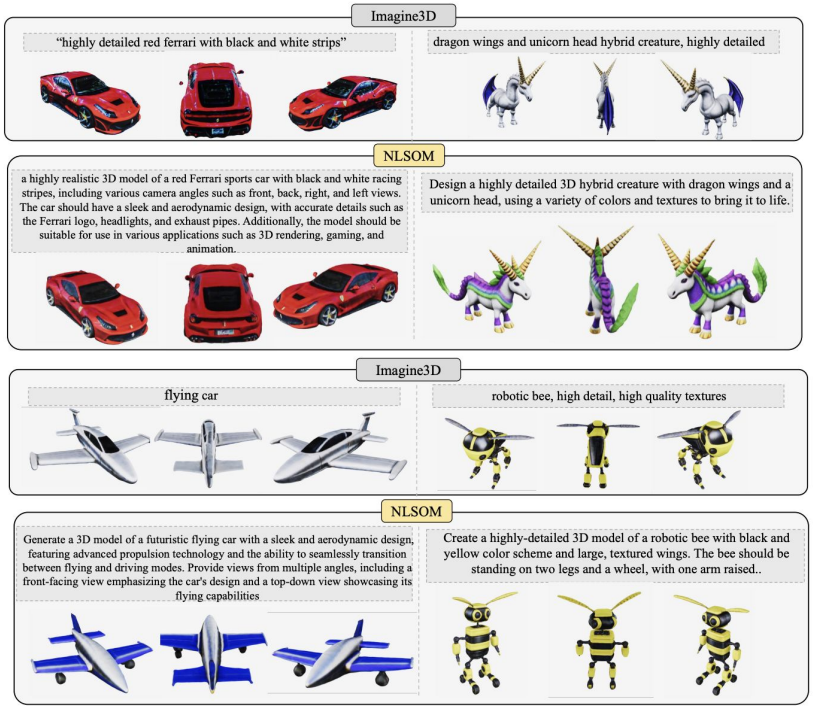


3D Generation

Architecture:

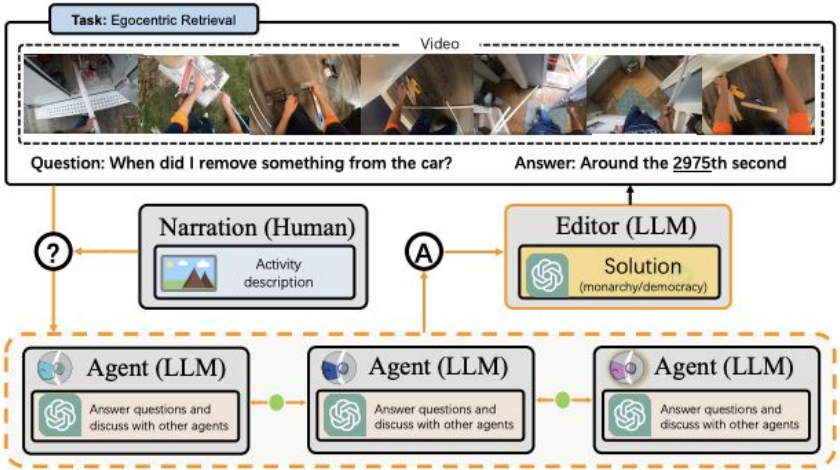


Result:



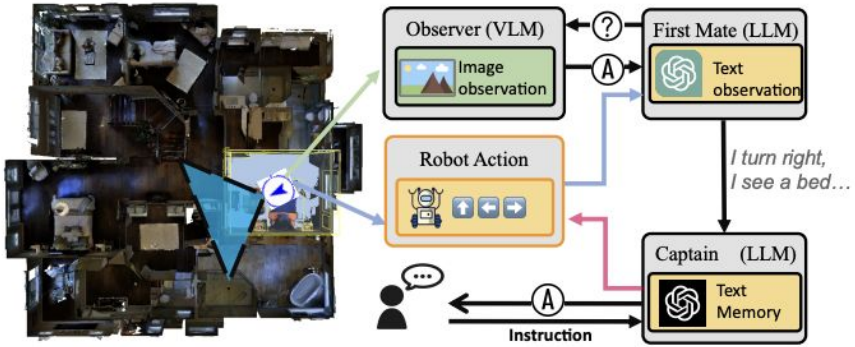
Egocentric Retrieval

Architecture:



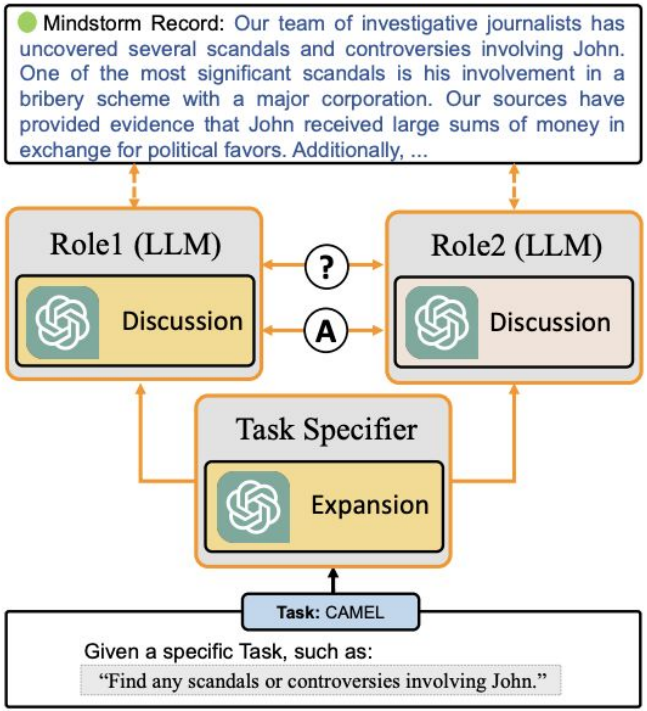
Embodied AI

Architecture:

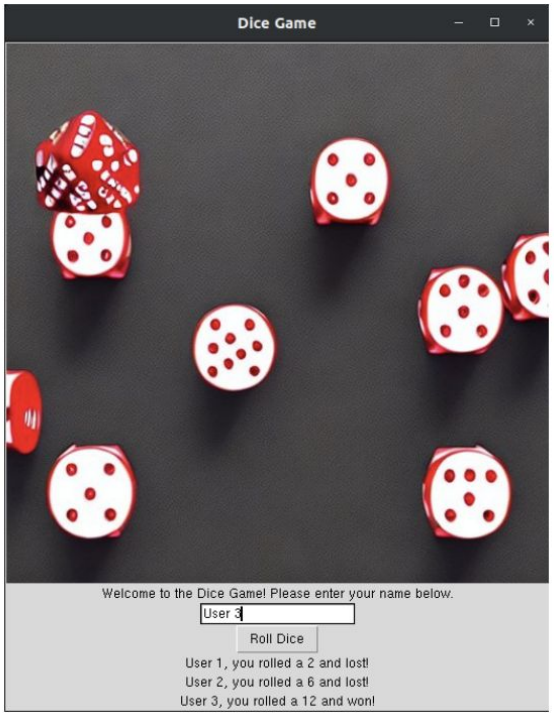


General Language-based Task Solving

Architecture:



Result:



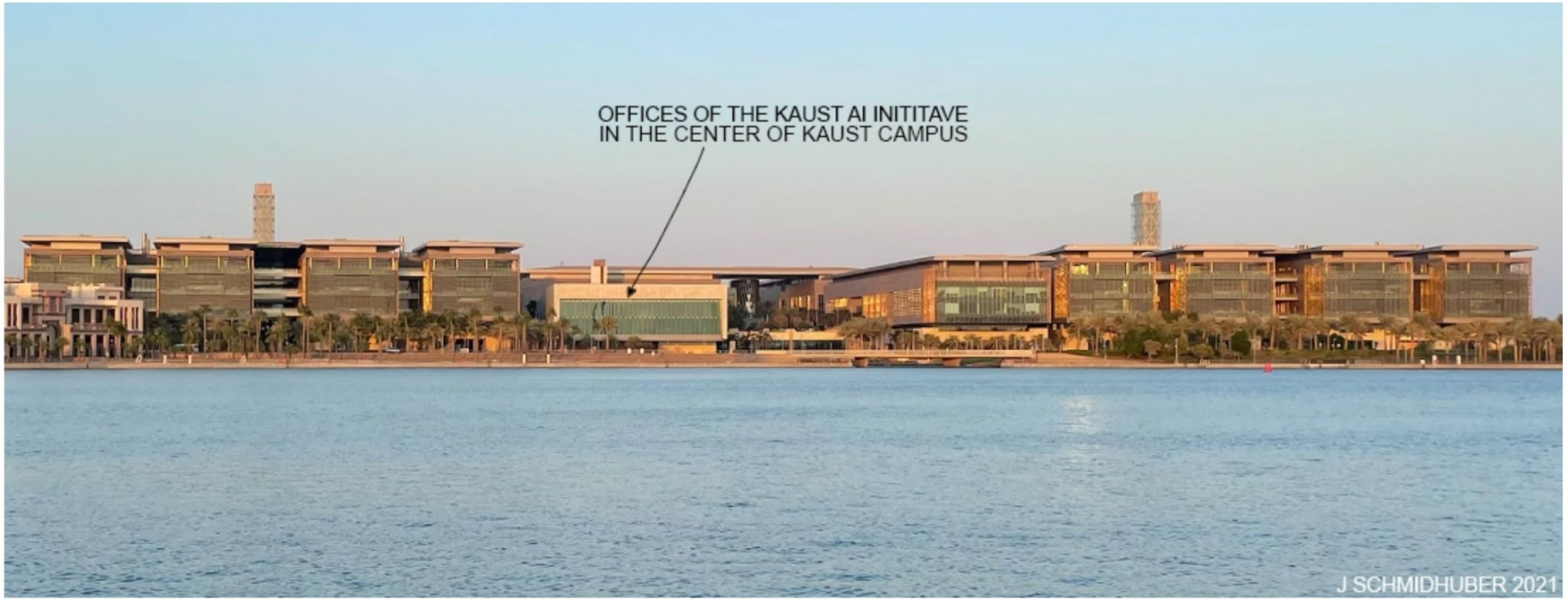
Outlook

This work opens the door to a lot of interesting research questions, such as:

- Self Organization
- Economies of Mind
- NLSOMs in the Physical World

Conclusion

- We present a framework for combining heterogeneous agents to solve a larger class of problems than they could individually.
- We show that this framework achieves strong results in a variety of settings even **without any additional training**.



Hiring PostDocs & PhD students in AI!

Questions?

Made with Midjourney