



KAUST Artificial  
Intelligence Initiative  
مبادرة كاوست  
للذكاء الاصطناعي

# Mindstorms in Natural Language-Based Societies of Mind

---

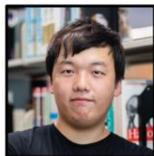
@ The NeurIPS 2023 Workshop on Robustness of  
Zero/Few-Shot Learning in Foundation Models



جامعة الملك عبد الله  
للعلوم والتقنية  
King Abdullah University of  
Science and Technology



Mingchen Zhuge



Haozhe Liu



Francesco Faccio



Dylan R. Ashley



Istituto  
Dalle Molle  
di studi  
sull'intelligenza  
artificiale

IDSIA



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



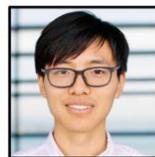
Aleksandar Stanić



Wenyi Wang



Yuhui Wang



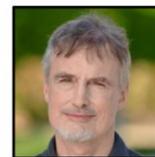
Mengmeng Xu



Deng-Ping Fan



Bernard Ghanem



Jürgen  
Schmidhuber



UNIVERSITY OF  
OXFORD



HARVARD  
UNIVERSITY

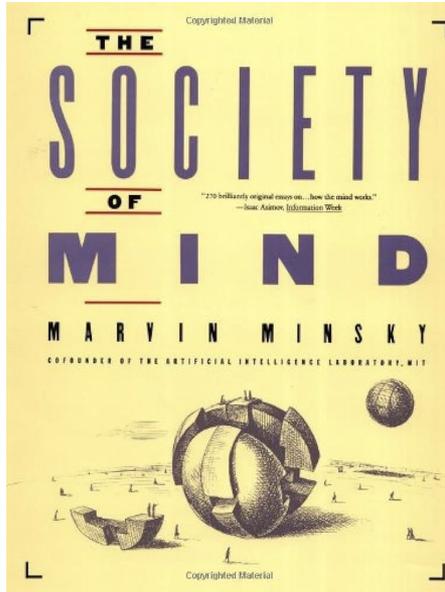


ETH zürich



---

# Minsky's Society of Mind

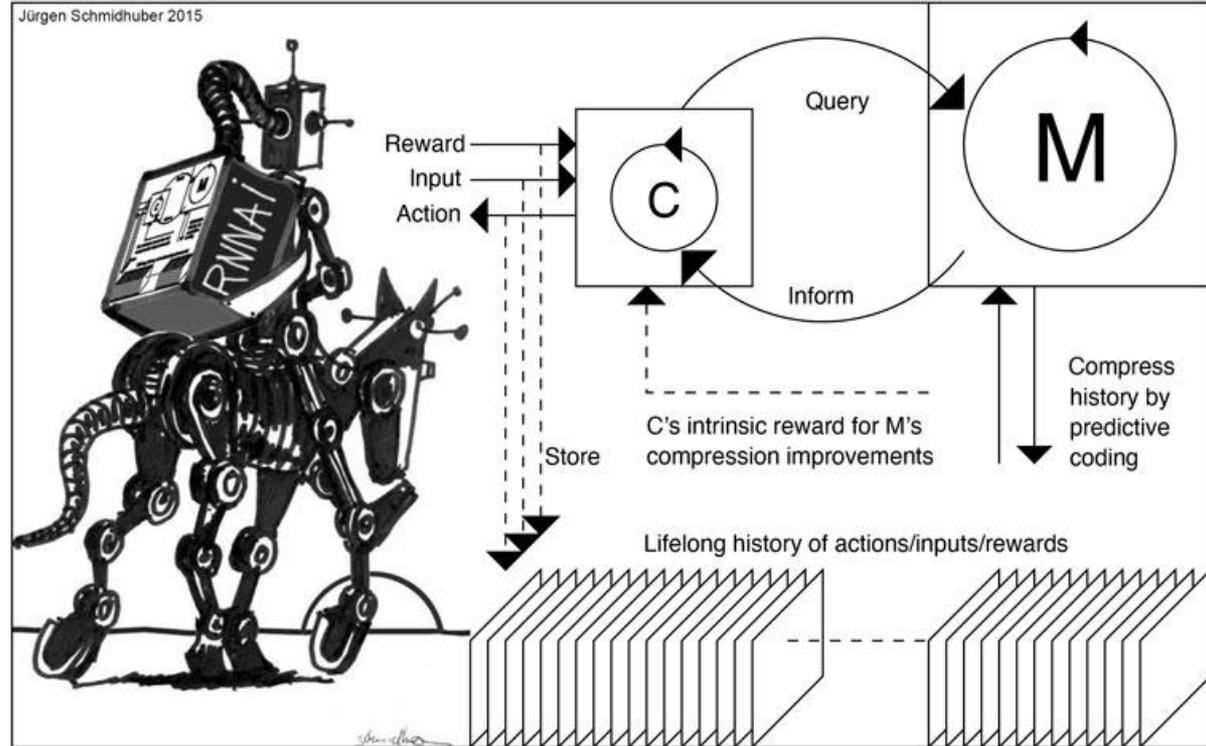


*What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.*

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

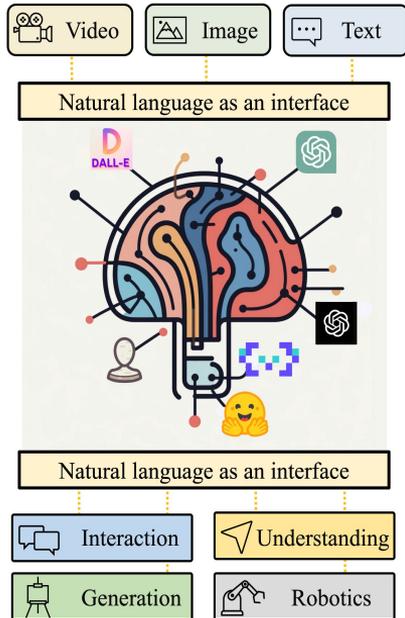
# Learning to Think

- 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



---

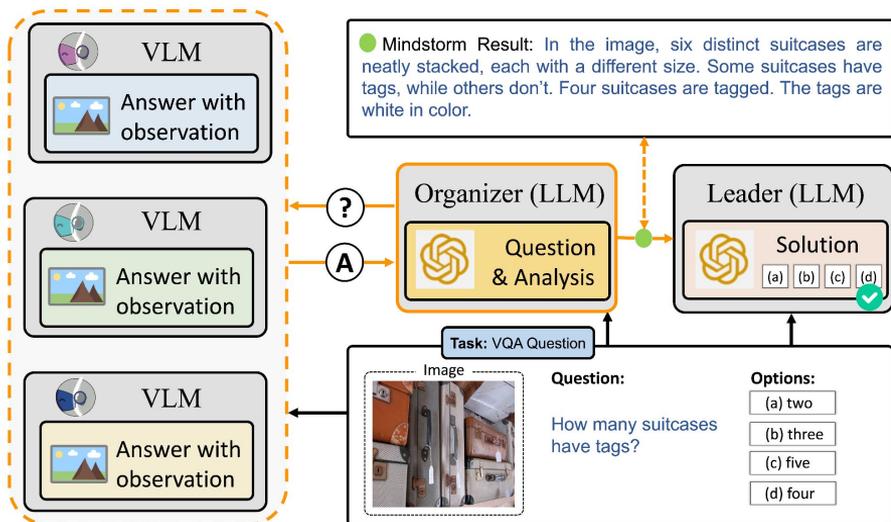
# A Natural-Language–Based Society of Mind



- 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:

**VQA Question**

**Question:**  
How many suitcases have tags?

**Options:**  
(a) two (b) three (c) five (d) four

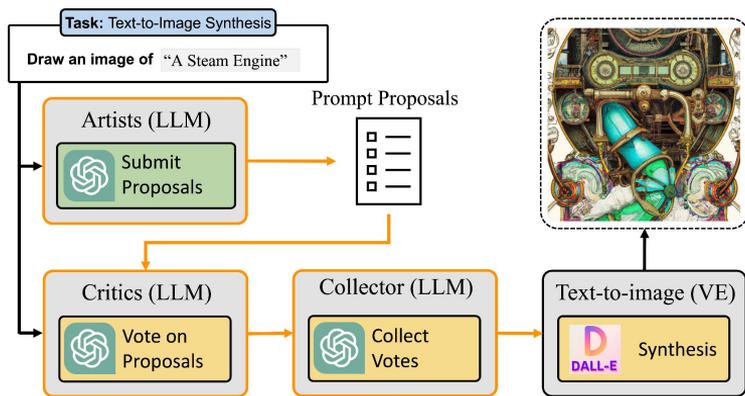
**Mindstorm Result:** In the image, six distinct suitcases are neatly stacked, each with a different size. Some suitcases have tags, while others don't. Four suitcases are tagged. The tags are white in color.

**Choices:** OFA: (b) mPLUG: (b) VQA-SoM (Ours): (d)

**Vision**

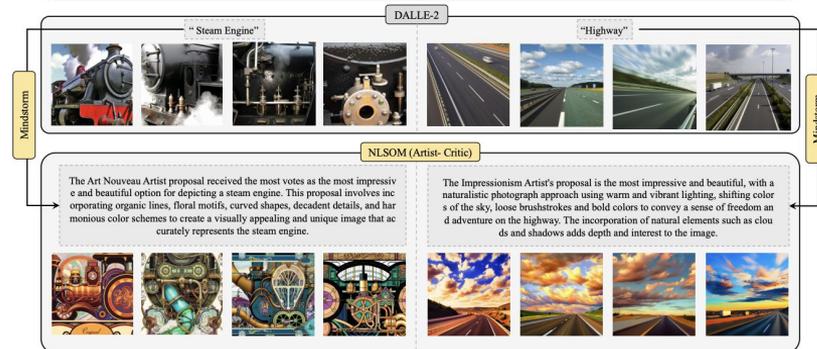
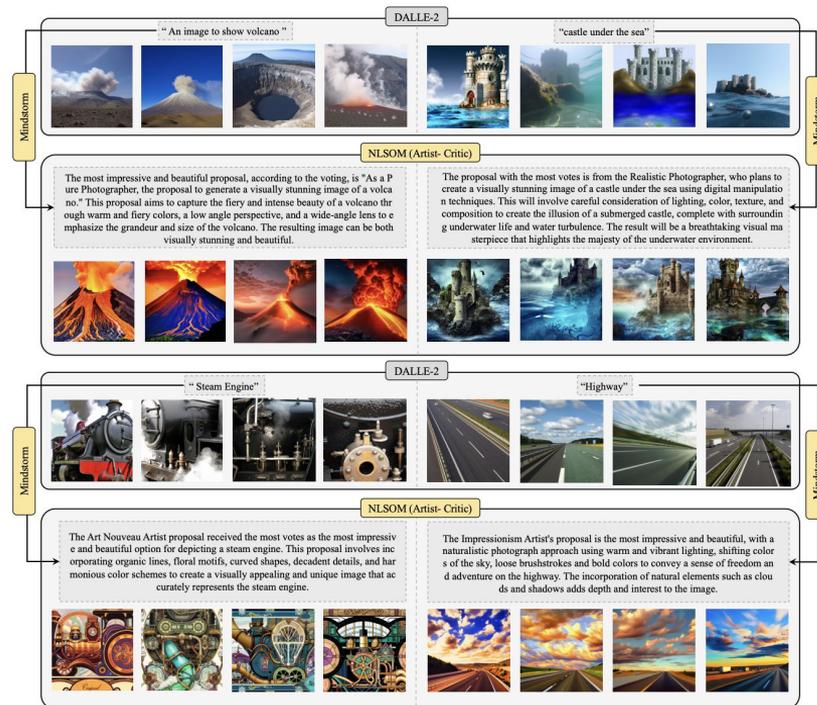
# Prompt Generation for Text-to-Image Synthesis

Architecture:



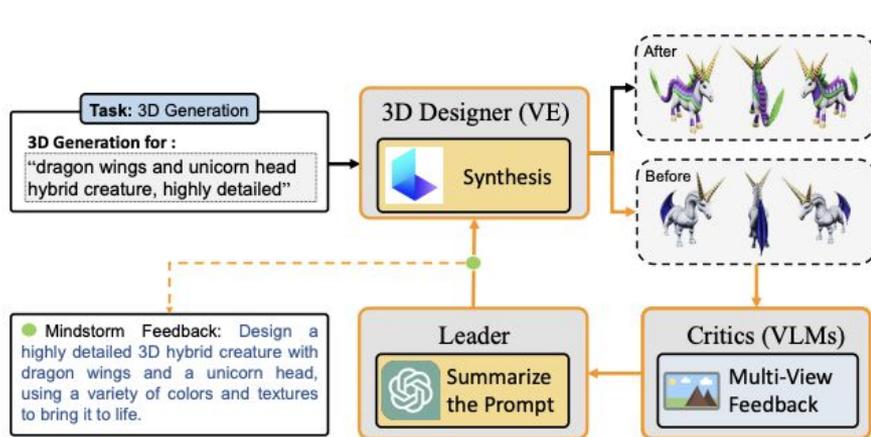
● **Mindstorm Result:** The Art Nouveau Artist proposal received the most votes as the most impressive and beautiful option for depicting a steam engine. This proposal involves incorporating organic lines, floral motifs, curved shapes, decadent details, and harmonious color schemes to create a visually appealing and unique image that accurately represents the steam engine.

Result:



# 3D Generation

Architecture:



Result:

**Imagine3D**

"highly detailed red ferrari with black and white stripes"      dragon wings and unicorn head hybrid creature, highly detailed




**NLSOM**

a highly realistic 3D model of a red Ferrari sports car with black and white racing stripes, including various camera angles such as front, back, right, and left views. The car should have a sleek and aerodynamic design, with accurate details such as the Ferrari logo, headlights, and exhaust pipes. Additionally, the model should be suitable for use in various applications such as 3D rendering, gaming, and animation.

Design a highly detailed 3D hybrid creature with dragon wings and a unicorn head, using a variety of colors and textures to bring it to life.




**Imagine3D**

flying car      robotic bee, high detail, high quality textures




**NLSOM**

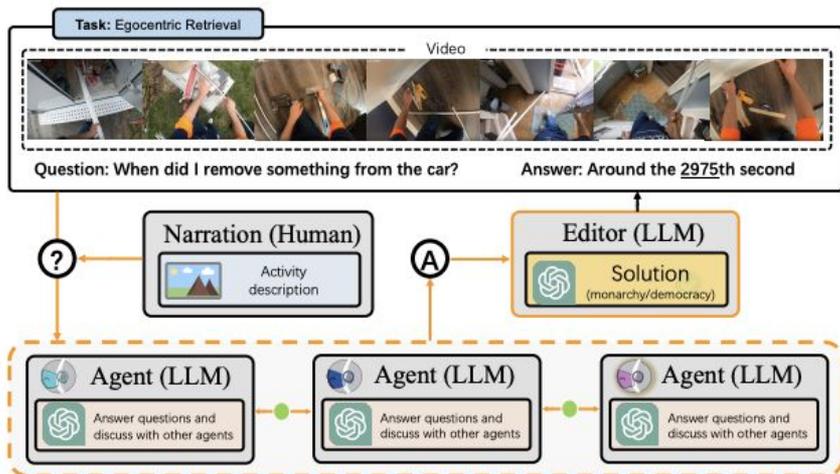
Generate a 3D model of a futuristic flying car with a sleek and aerodynamic design, featuring advanced propulsion technology and the ability to seamlessly transition between flying and driving modes. Provide views from multiple angles, including a front-facing view emphasizing the car's design and a top-down view showcasing its flying capabilities.

Create a highly-detailed 3D model of a robotic bee with black and yellow color scheme and large, textured wings. The bee should be standing on two legs and a wheel, with one arm raised.



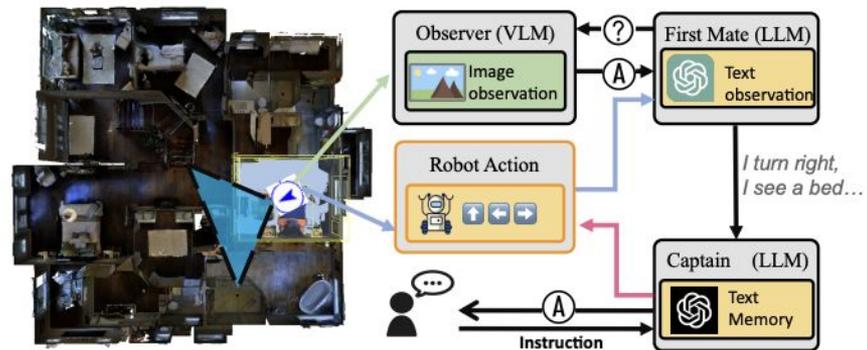

# Egocentric Retrieval

Architecture:



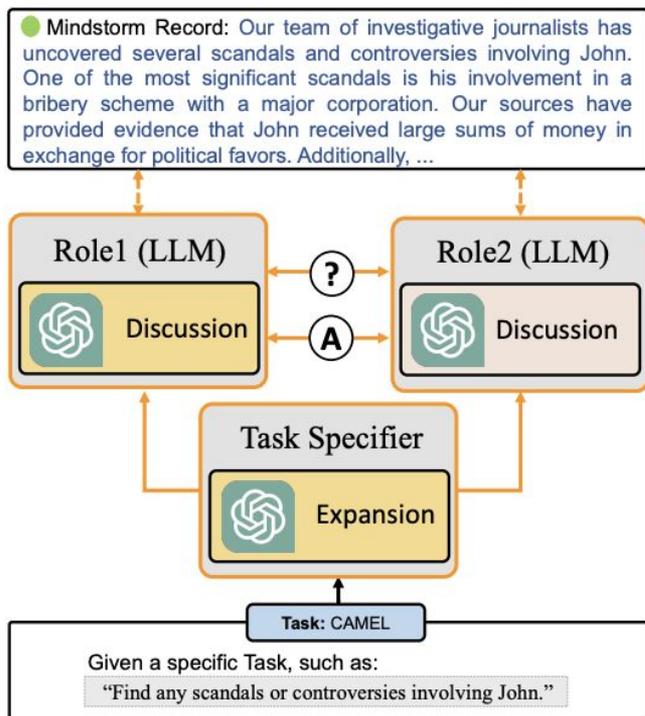
# Embodied AI

Architecture:



# General Language-based Task Solving

Architecture:



Result:



---

IS 2004



---

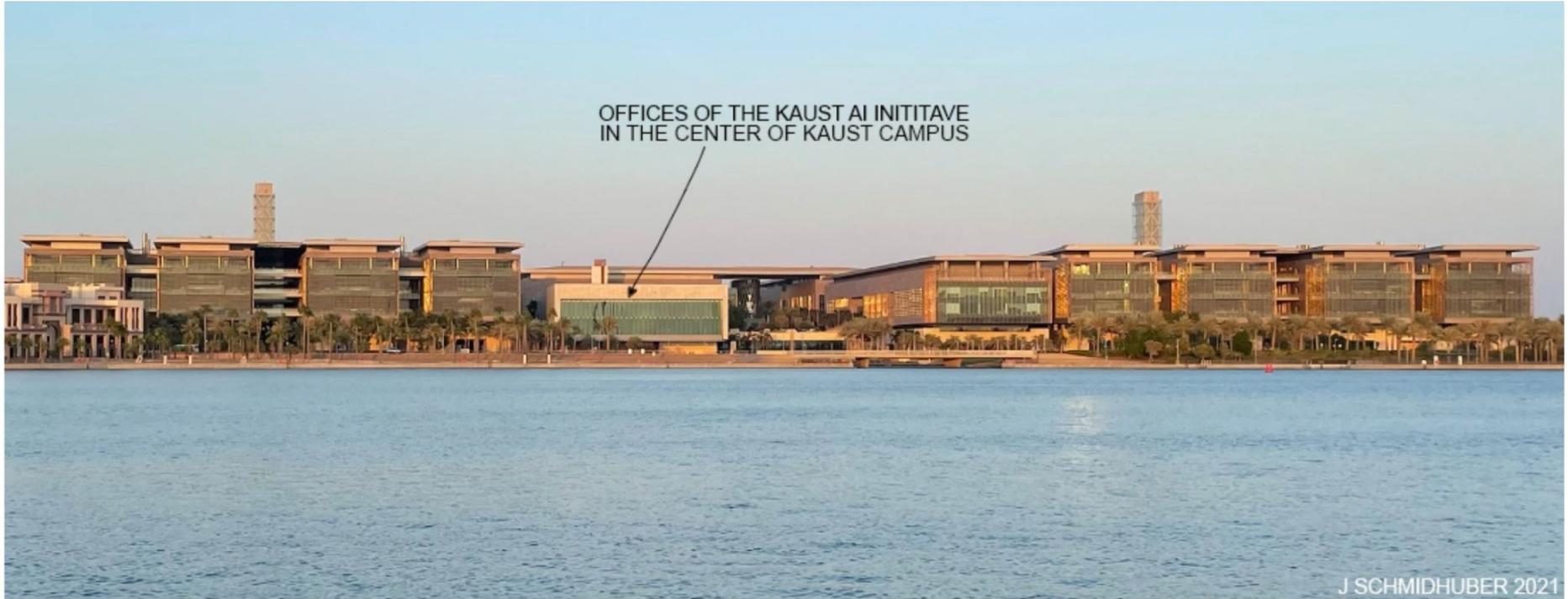
# Outlook

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?

