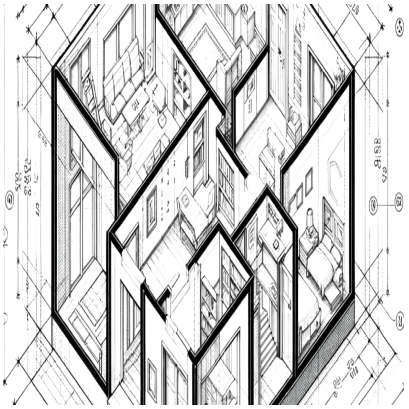

Societies of AI minds

Description

In the 1980s, AI pioneer Marvin Minsky authored the book *The Society of Mind*, which outlined how we could think of the human mind as so many communicating agents: â??Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies â?? in certain very special ways â?? this leads to true intelligenceâ?• [1].

In this work he seems to avoid concepts such as nested agency, agents within agents, or agents made up of clusters of agents, but it appears to me that the agent analogy calls for such complicated framings. It follows that intelligence, or at least cognition, is *agents all the way through*! An urban reference brings this home:

> â??The human brain contains so many agencies and connections that it resembles a great nation of cities and townsâ?• [314].

Such saturated stratified agency pervades NLP models. I alluded in the previous post to multi-attention heads in the Transfomer model of NLP as a configuration of agents. The analogy extends to the entire operations of AI models, platforms and systems, such as ChatGPT, that operate as agents themselves, able to â??authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, forbidâ?• [72], to quote Latour.

Agent narratives

The idea that AIs might communicate with one another enters the AI discourse through several narratives. First is the absurdist characterisation of AI chat bots caught in endless conversational loops. In an earlier post I suggested AIs write essays for students, which are in turn assessed by an AI bot that delivers feedback to the student bot. The student bot subsequently improves its output for further feedback and assessment. The learning and teaching experience is outsourced to AIs while student and tutor meet only to socialise, if at all. Nothing gets learned, no-one will pay for this kind of education, society deskills and institutions shut down.

A second, apocalyptic, variant entails AIs initiating conversations amongst themselves in some kind of singularity. Thatâ??s how I interpret the closing phases of Spike Jonzeâ??s movie *Her*, where the personalised intelligent agent departs from her human confidant to join other similarly competent operating systems to form a singular super agent, with a mission as yet unknown to humans.

A third, banal exploration of inter-agent communication realises the obvious temptation to feed the response from one AI chatbot (e.g. chatGPT) into another (e.g. Bing). A Reddit social media user asked the forum: â??what would happen if you wire the output from ChatGPT as the input to another ChatGPT running on a second computer and then feed that output as input to the first?â?• Responses from other forum members range from reports of enthusiastic agreement between chatbots about some topic to â??without input, nudging or meddling from a human moderator it is a very generic and bland conversation.â?•

A fourth narrative assumes human agency in the creation of multi-agent systems that address specific tasks.

Adversarial AI

Generative Adversarial Networks (GANs) are such multi-agent systems. In a GAN, there are just two â??agentsâ?•: the generator and the discriminator. The generator generates fake data that is indistinguishable from actual data. The network is trained to produce outputs that resemble the training dataset. The discriminator is an agent that aims to differentiate between actual data (from a training dataset) and fake data produced by the generator. Its goal is to accurately classify data as either actual or generated.

The training process of a GAN involves a game-like scenario where the generator continuously tries to â??foolâ?• the discriminator with increasingly realistic data, while the discriminator becomes better at distinguishing fake data from real data. This adversarial process leads to the generator creating highly plausible (though artificial) data over time.

1.

GANs are widely used for creating images, which can range from human faces to art and buildings. This has applications in graphic design, gaming, movie production and architecture. GANs are used to enhance the resolution of images, applicable to restoring old movies, enhancing satellite imagery, and improving the quality of medical images. GANs can generate images from textual descriptions, which has applications in aiding visual artists and generating visual content for advertising. The banner to this post was produced with Dall-e which uses a Transformer architecture rather than a GAN, but GANs produce similar outputs.

In the next post I will review multi-agent systems designed for AI agents and human participants using NLP tools such as chatGPT.

References

- Minsky, M. L. (1986). *The Society of Mind*. New York: Simon and Schuster.

- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-network-theory* Oxford: Oxford University Press.
- Wu, Q., G. Bansal, et al. (2023). â??AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation.â• arXiv: Cornell University 3 October. Retrieved 9 December 2023, from https://arxiv.org/abs/2308.08155.

Note

- Featured image is from Dall-e, captioned â??Here are the updated images showing a 2D architectural working drawing of a house section, without any 3D models. The drawings are aligned orthogonally with the edges of the picture and display details like room names, measurements, and architectural symbols.â•

Category

1. Artificial Intelligence

Tags

1. adversarial AI
2. agents
3. GANs
4. Minsky
5. Society of Mind

Date Created
December 30, 2023
Author
rcoyne99