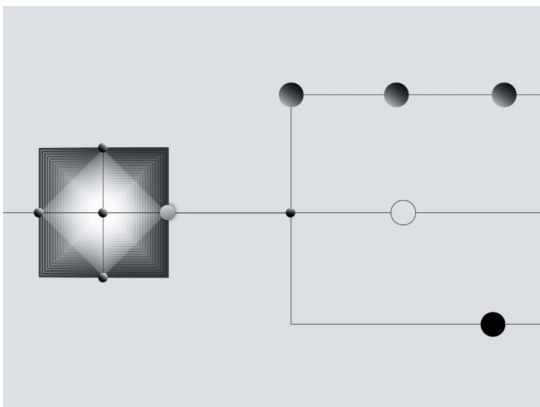# In the Lab

An air traffic controller for LLMs

Analyzing incoming queries and handing them off in real time to the model most likely to provide a cost-effective response

Each LLM comes with its own set of performance metrics; sometimes the largest model isn't the most optimal for a task. Examining the economics of generative AI and automation potential, the Lab teams of Justin Solomon, Neil Thompson, Kate Soule, and Mikhail Yurochkin trained a routing algorithm to identify, from a model suite, the model with the best predicted accuracy and cost.
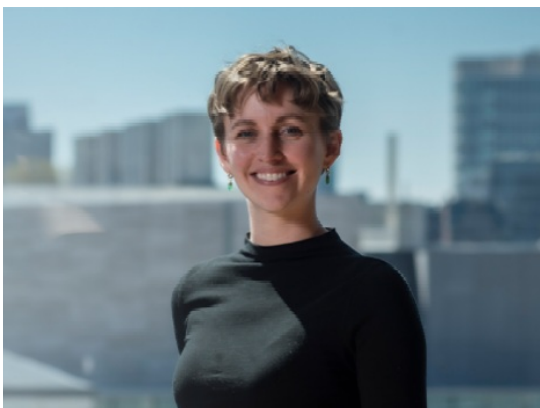
Participatory AI highlights paths to sustainability

**Combining machine learning with human expertise to better understand a rapidly changing planet**

Lab researcher Sara Beery grapples with environmental and ecological data to discover ecosystem trends and biodiversity losses and help wildlife management target solutions, and companies to be more sustainable.

How generative AI is boosting manufacturing design

[Towards a "general artificial designer" that can deliver novel and highly practical results](#)

Lab researcher Faez Ahmed is rethinking design tools that can learn from many different designs, transfer ideas across domains. By combining generative AI with optimization methods, his group looks to innovate in the engineering and manufacturing space.
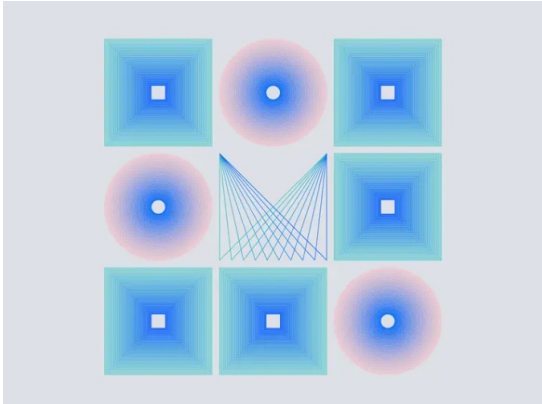
How memory augmentation can improve LLM efficiency, flexibility

[Creative strategies to reduce the memory footprint of generative AI models](#)

Unlike humans, LLMs lack long-term memory. However, work from the Lab groups of Leonid Karlinsky, Dmitry Krotov, and Rogerio Feris are looking to neuroscience to improve the capacity of LLMs with their method, CAMELoT (Consolidated Associative Memory Enhanced Long Transformer).
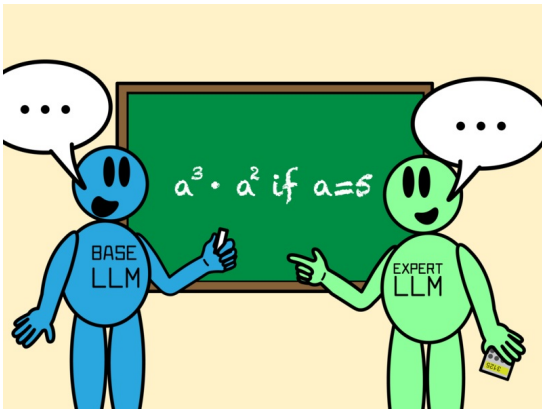
Enhancing LLM collaboration for smarter, more efficient solutions

[“Co-LLM” algorithm helps a general-purpose AI model collaborate with an expert LLM.](#)

The Lab groups of Yoon Kim and David Sontag have developed a framework that uses machine learning to train a "switch variable," to find areas in a response when the specialist LLM should intervene with a better answer. The method improves the accuracy of response to medical prompts and math and reasoning problems.

In the Media

Balancing fairness of item display and user preferences in online marketplaces

Algorithms that market online items for purchase don't prioritize goods and services equally. A first-of-its-kind framework from the Lab groups of Negin Golrezaei and Dajallel Bouneffouf aims to balance fairness of items displayed to the purchaser, user preferences, and the platform's objectives, reports [Morningstar](#).

Lab Highlights

Lab researcher Song Han has provided his [course on efficient machine learning and systems](#) online for

viewing.

Lab researchers had more than 20 papers and workshops accepted to the Conference on Neural Information Processing Systems ([NeurIPS](#)), a premier conference on machine learning and computational neuroscience conference.

Online Learning

[Unsupervised Machine Learning: Unlocking the Potential of Data](#)
A joint MIT Sloan & Schwarzman College of Computing Executive and Professional Course begins November 13.

[Making AI Work: Machine Intelligence for Business and Society](#)
A joint MIT Sloan & Schwarzman College of Computing Executive and Professional Course begins November 20.

[Artificial Intelligence: Implications for Business Strategy](#)
A joint MIT CSAIL and MIT Sloan School of Management Course begins November 27.