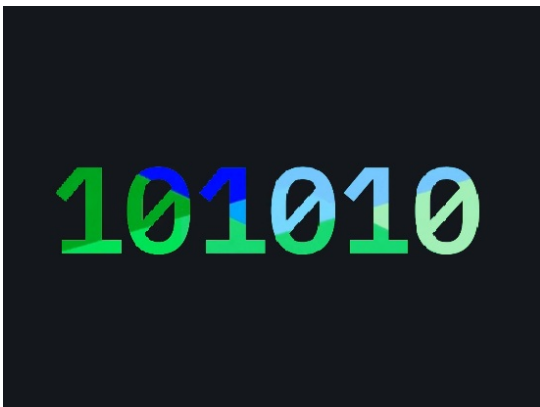




In the Lab



IBM Granite now has eyes

[New vision-language model can extract knowledge locked away in tables, charts, and other graphics](#)

IBMers and many Lab researchers developed Granite Vision — an open-source vision-language language model (VLM) for enterprise AI that can analyze natural images and other data visualizations people encounter daily and outperforms other VLMs.



Like human brains, LLMs reason about diverse data in a general way

[LLMs represent data types based on their underlying meaning and reason about data in their dominant language.](#)

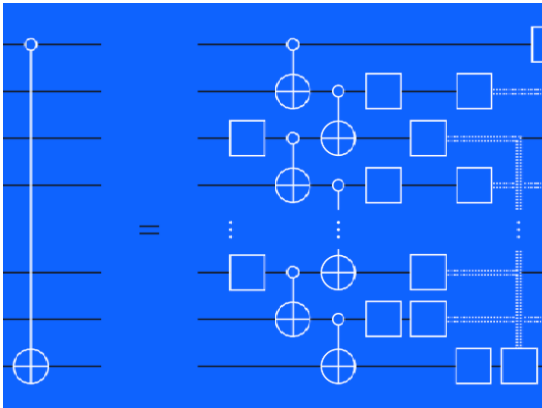
A team of researchers working with the Lab's Yoon Kim examined the inner workings of LLMs and found that they processed assorted data like the human brain in a centralized way — through a "semantic hub."

21249	09511	21249
8 5 7 6	5 9 9 9 0	7 8 5 7 6
9 3 3 0	4 2 6 8 0	7 9 3 3
5 5 6 8	7 7 4 9 3	4 5 5 6
11704	96459	11704
55411	23586	55411
9 9 6 4	9 2 4 4 7	2 9 9 6 4
8 8 7 2	7 8 1 8 1	1 8 8 7
3 2 8 2	8 3 3 3 0	2 3 2 8

New benchmark changes monthly to avoid teaching to the test

[LiveXiv evaluates vision-language models on questions and images they've likely never seen before.](#)

A multimodal benchmark, developed by a team including Lab researchers Leshem Chosen, Leonid Karlinsky, and Mikhail Yurochkin, provides an evolving set of test questions corresponding to images in the latest papers posted on arXiv for evaluating VLMs on their ability to analyze charts, tables, diagrams, and other images.

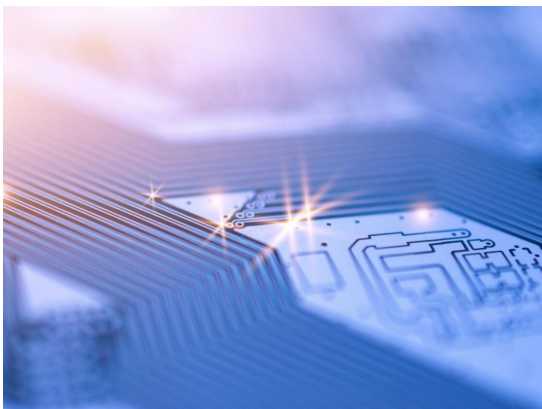


Using dynamic circuits to implement quantum states with long-range entanglement

[An efficient new strategy for connecting distant qubits](#)

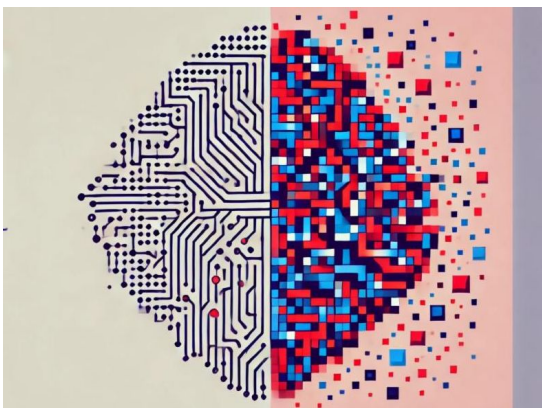
Work from a team including Lab researcher Patrick Rall shows how dynamic circuits can generate long-range entanglement between qubits using shallow-depth quantum circuits, overcoming the limited connectivity between qubits in superconducting quantum chips while also making the qubits more resilient against environmental noise.

In the Media



Adapting large language models for specialized tasks

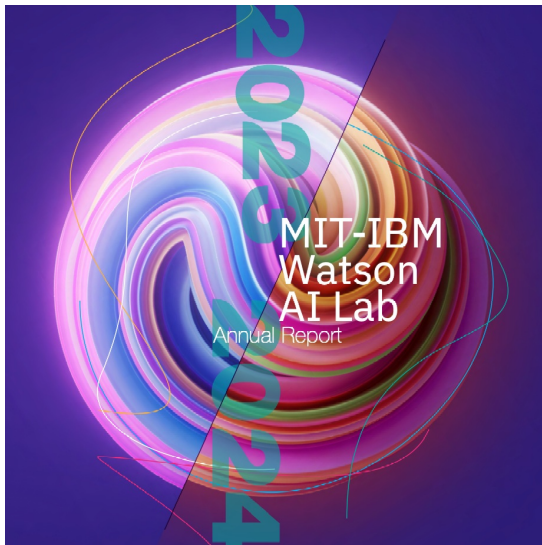
Research from the Lab group of Xin Zhang developed SOLOMON, a neuro-inspired LLM reasoning network architecture. The work, [EngiSphere](#) reports, uses prompt engineering and in-context learning methods along with several components — thought generator, thought assessor and a steering subsystem — to adapt to domain-specific tasks, like semiconductor layout design.



Diffusion models offer keys To success for resource-scarce systems

Work out of the Lab team of Song Han has produced SVDquant, a 4-bit quantization for diffusion, reports [Forbes](#), resulting in reduced memory usage and latency.

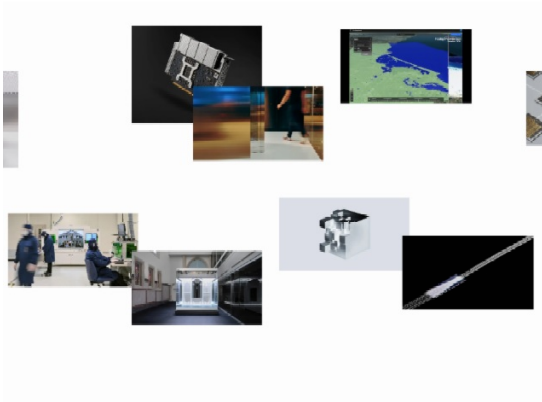
Annual Reports



MIT-IBM Watson AI Lab

[2023-2024](#)

The accomplishments the Lab has achieved in the past year are varied and exciting — from the development of the Lab's summer internship and the publication of over 1,100 peer-reviewed papers over the life of the Lab to key innovations for our corporate members and contributions to IBM's watsonx, and collaborative projects targeting foundational challenges. Here, we highlight how we are shaping the future of the field and its players.



IBM Research annual letter

[2024](#)

Over the last year, IBM Research made advances in AI, quantum computing, semiconductors, and fundamental research. This includes Lab contributions to the Granite suite of models.

Event Recordings

[Expanding Horizons in Computing](#)

Hosted by the MIT Schwarzman College of Computing, MIT faculty explored essential computing topics over four immersive, daylong sessions. From deep learning [and societal impacts](#) to cryptography, security, and quantum technology, the sessions offered a compelling look at the opportunities and challenges shaping the future of computing.

Lab Highlights

Lab researchers Tamara Broderick and Nuno Loureiro were awarded the [Presidential Early Career Award for Scientists and Engineers \(PECASE\)](#). Broderick was nominated by the Office of Naval Research for her project advancing "lightweight representations for decentralized learning in data-rich environments," and Loureiro was nominated by the NSF for his work on the generation and amplification of magnetic fields in the universe.

Lab researcher Marzyeh Ghassemi was awarded a [2025 Sloan Research Fellowship](#), which honors exceptional researchers at U.S. and Canadian educational institutions, whose creativity, innovation, and research accomplishments make them stand out as the next generation of leaders.

Online Learning

[Making AI Work: Machine Intelligence for Business and Society](#)

A joint MIT Sloan & Schwarzman College of Computing Executive and Professional Course begins
March 12

[Unsupervised Machine Learning: Unlocking the Potential of Data](#)

A joint MIT Sloan & Schwarzman College of Computing Executive and Professional Course begins
March 19

[Transforming Healthcare with AI](#)

A joint MIT Sloan School of Management & MIT Jameel Clinic Course begins
March 24

[Artificial Intelligence in Health Care](#)

A joint MIT Sloan School of Management & MIT Jameel Clinic Course begins
March 26

[Quantum Computing Fundamentals](#)

An MIT xPRO Course begins
April 7

[Artificial Intelligence in Pharma and Biotech](#)

An MIT Sloan School of Management Course begins
April 9

[Machine Learning in Business](#)

A joint MIT CSAIL and MIT Sloan School of Management Course begins
April 9

[Driving Innovation with Generative AI](#)

An MIT xPRO Course begins
April 14