

Rethinking Research: the Role of Humans in Scientific Discovery in the age of the LLMs*

**This presentation was prepared by humans, augmented by LLMs*



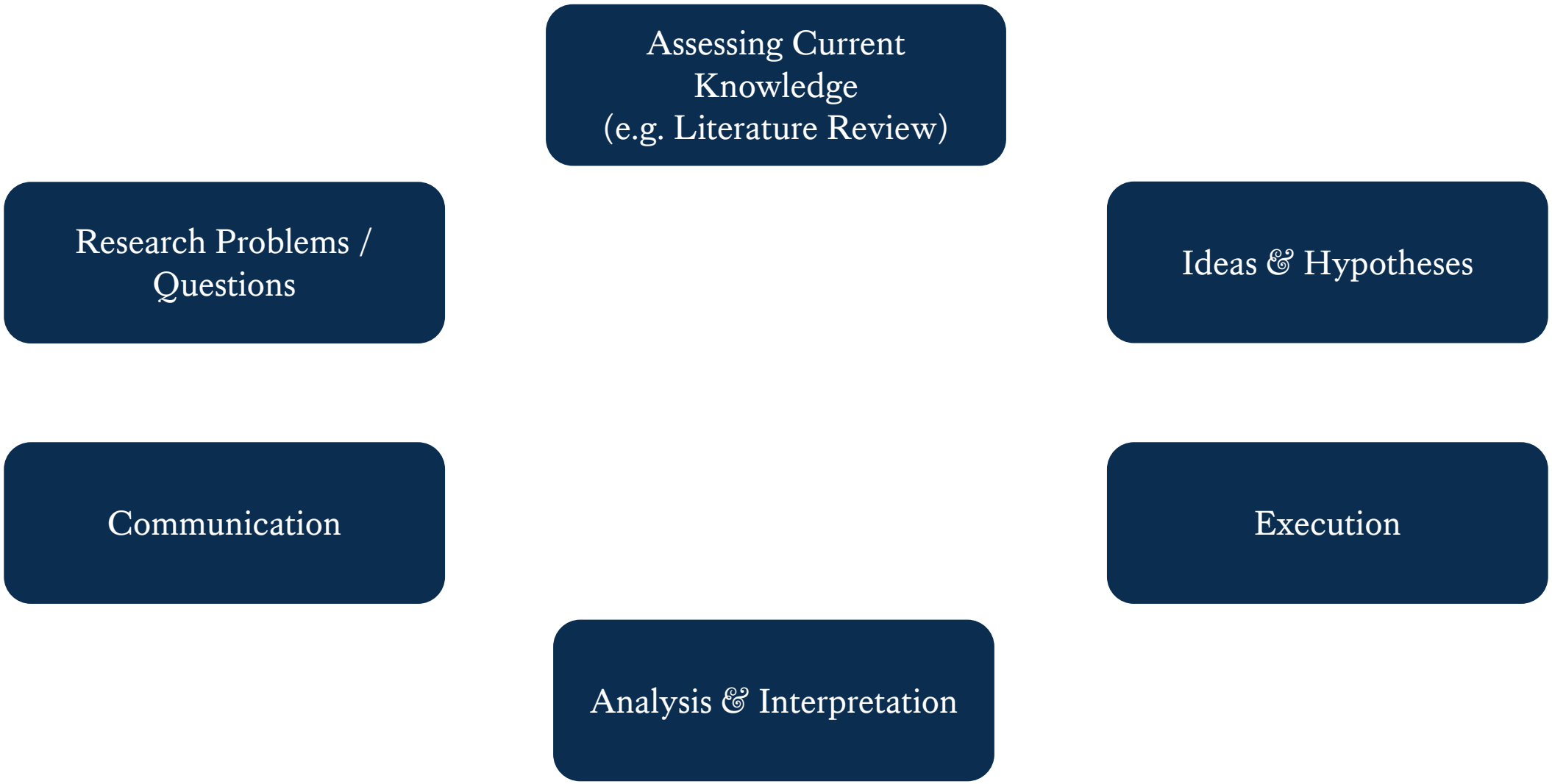
Professor Sir Bashir M. Al-Hashimi CBE FREng FRS

December 2025

Assertion

- AI technologies – particularly LLMs – have the potential to transform research, improve human scientific capability and accelerate knowledge creation
- Without training and appropriate oversight, the use of LLMs has equally the potential to undermine:
  - The human contribution to scientific discoveries
  - The supervision, mentoring, and intellectual development of doctoral students and researchers

Research Workflows



Research Topics and Assessing Current Knowledge



ChatGPT Prompt

What are the two open research problems in Edge AI and list the most influential paper in each problem?

Research Problems and Assessing Current Knowledge



ChatGPT Output

1. Efficient AI Model Deployment on Edge Devices

- **Problem:** Edge devices (e.g. smartphones, IoT sensors) have limited **computing power**, **memory**, and **energy**.
- **Why it's open:** Designing AI models that are **both accurate and lightweight** (e.g. using quantization, pruning, or neural architecture) without sacrificing performance is still a hard balance.
- **Highly influential paper:** *"Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding"* by Song Han, Huizi Mao, and William J. Dally (2016).

2. Data Privacy and Federated Learning

- **Problem:** Training models without centralizing data (to preserve privacy) is difficult.
- **Why it's open: Federated learning** introduces problems like **communication overhead**, and **security vulnerabilities** (e.g. model poisoning).
- **Highly influential paper:** *"Communication-Efficient Learning of Deep Networks from Decentralized Data"* by H. Brendan McMahan et al. (2016)

Observations

Observation 1: Research problems appear plausible but without **domain expertise***, won't be able to establish if the open problems are worthy

Observation 2: These are indeed highly cited papers, not everyone agrees on these being the most influential: **domain expertise is vital**

*Experimental and theoretical knowledge and skills developed through study and research

Research Questions



ChatGPT Prompt

Regarding the first open research problem, what are the two most promising research questions I can research that are worthy and leading to publications?

Research Questions

ChatGPT Output

Here are **2 promising, researchable questions** with brief background and why your contribution would matter:

1- Adaptive Compression for Edge Deployment of Transformer Models

🔍 **Problem Statement:**
"Design an adaptive compression framework for Vision Transformers (ViTs) or TinyBERTs to dynamically reduce compute/memory cost during inference on heterogeneous edge devices."
✨ **Why it's valuable:**
Transformer models are replacing CNNs in edge vision and language tasks.
They're notoriously heavy for small devices.
Real devices vary in capability — dynamic compression can make deployment flexible.
📈 **Potential Contribution:**
A new compression method that adjusts *at runtime* based on device constraints.
A hardware-aware ViT pruning/quantization framework.

Research Questions



ChatGPT Output

2- Energy-Aware Neural Architecture Search (NAS) for TinyML

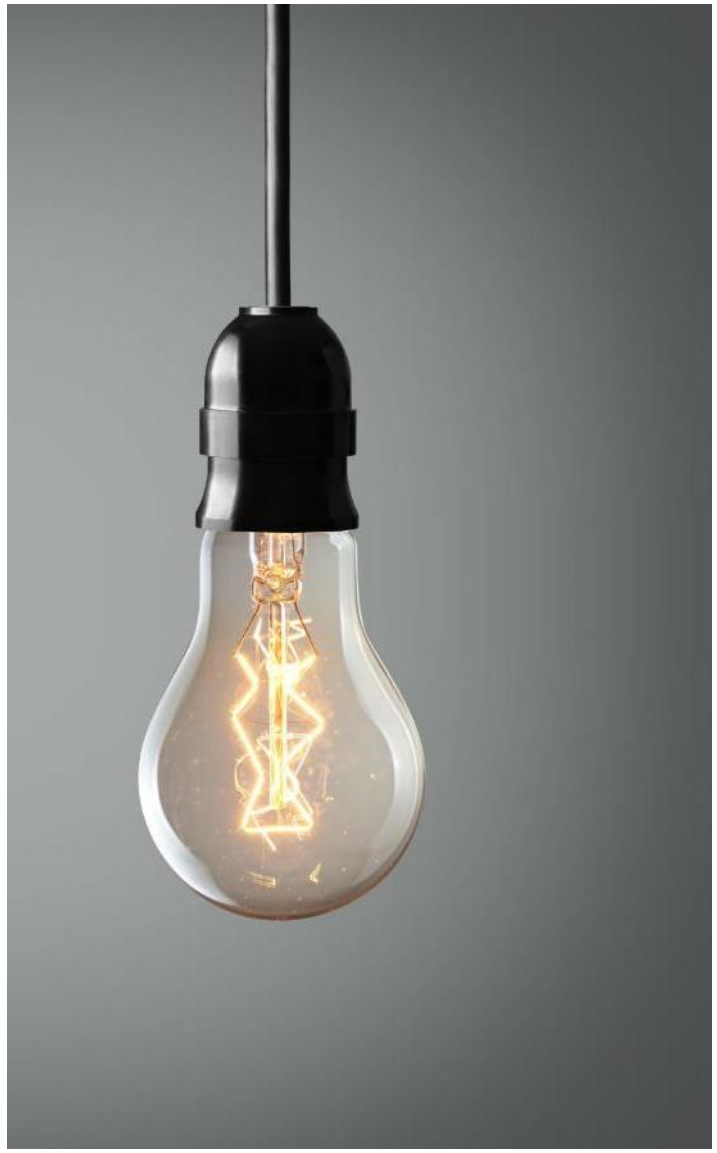
🔍 **Problem Statement:**
"Propose a multi-objective NAS approach that jointly optimizes accuracy, latency, and energy consumption for microcontroller-based edge AI systems."
✨ **Why it's valuable:**
Existing NAS techniques often target FLOPs or latency, not **energy consumption**, which is critical for battery-powered edge devices.
Energy modeling is difficult but under-researched.
The field of **TinyML** (ML on microcontrollers) is rapidly growing.

Observation

Observation 3: Are these research questions that will lead to new knowledge and publications?

Towards novel research ideas

- Prompting mechanisms: Multi-Perspective Critique/Chain of Thought/Recursive Reflection
- Propose cross-discipline, e.g. HW-SW co-design
- List assumptions, deliberately stress test and invert assumptions
- Find contradictions in current literature



Domain-Specific LLMs



ChatGPT Prompt

Give me an example of LLM optimised for Energy-Efficient Computing?

Domain-Specific LLMs: Scientific Models



ChatGPT Output

Large Language Models fine-tuned or designed to **help** in developing **low-power architectures, energy-efficient chips, and computing systems**.

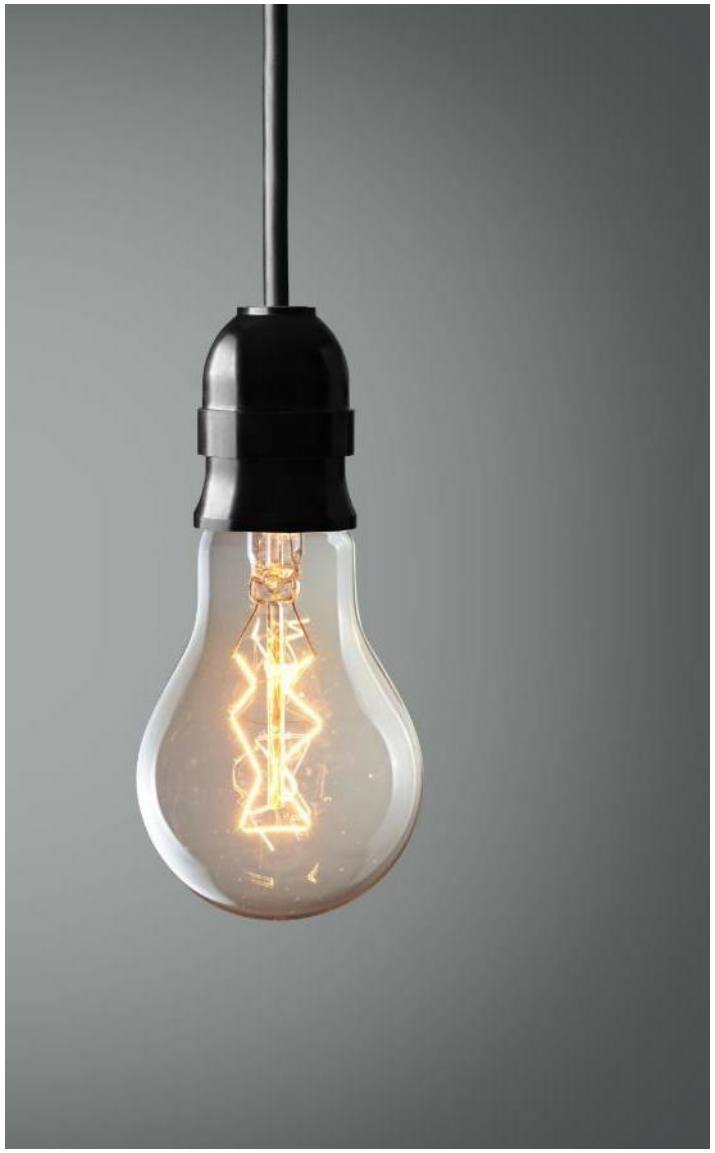
EnergyGPT ⚡ 🔋

- **Developer:** Lawrence Berkeley National Laboratory (LBNL)
- **Purpose:** Designed to **optimize energy usage** in **HPC (High-Performance Computing)** and **data centers**.
- **Capabilities:**
- Analyzes **power consumption logs**.
- Suggests **energy-saving configurations** for supercomputers.
- Assists researchers in selecting **low-power algorithms**.

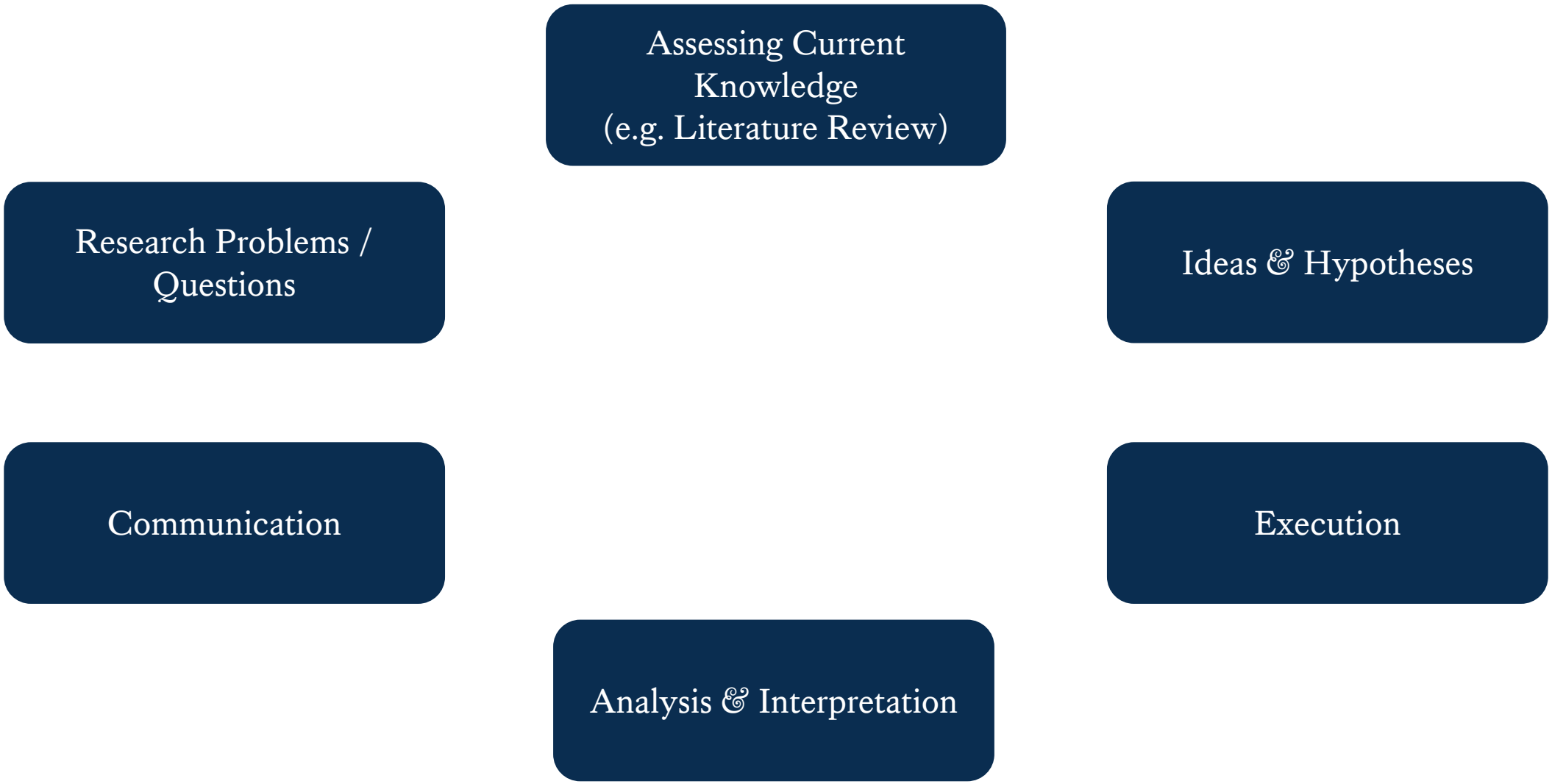
Observations

Observation 4: LMMs are remarkable in speed and efficiency at carrying out specific tasks: literature review, research ideas…, but need human domain expertise and supervision

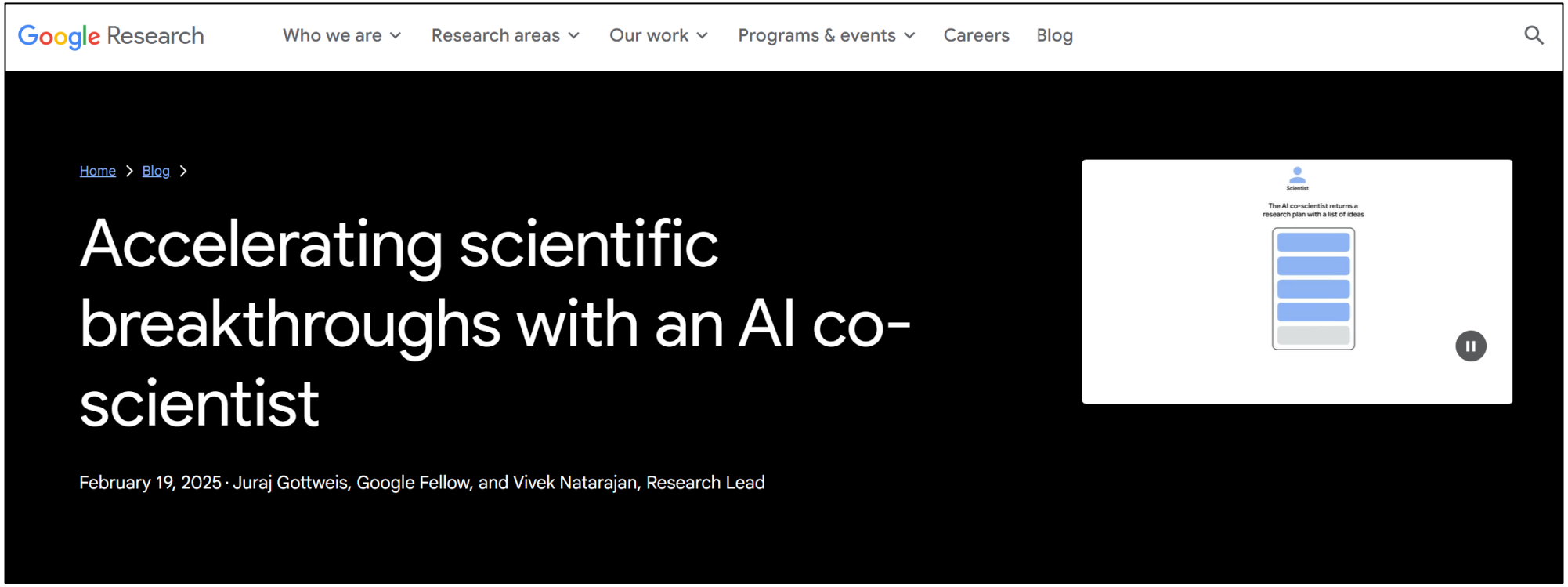
Observation 5: The trajectory of LLMs, scientific and frontiers models is nothing short of extraordinary. Technology doesn't just assist but amplify the humans cognitive. In hardware, we have the Moore's Law. In AI, we have the Kapland (OpenAI) and Chinchilla (DeepMind) scaling laws for LLMs performance



Research Workflows



LLMs for Research Collaboration (AI Co-Scientist)

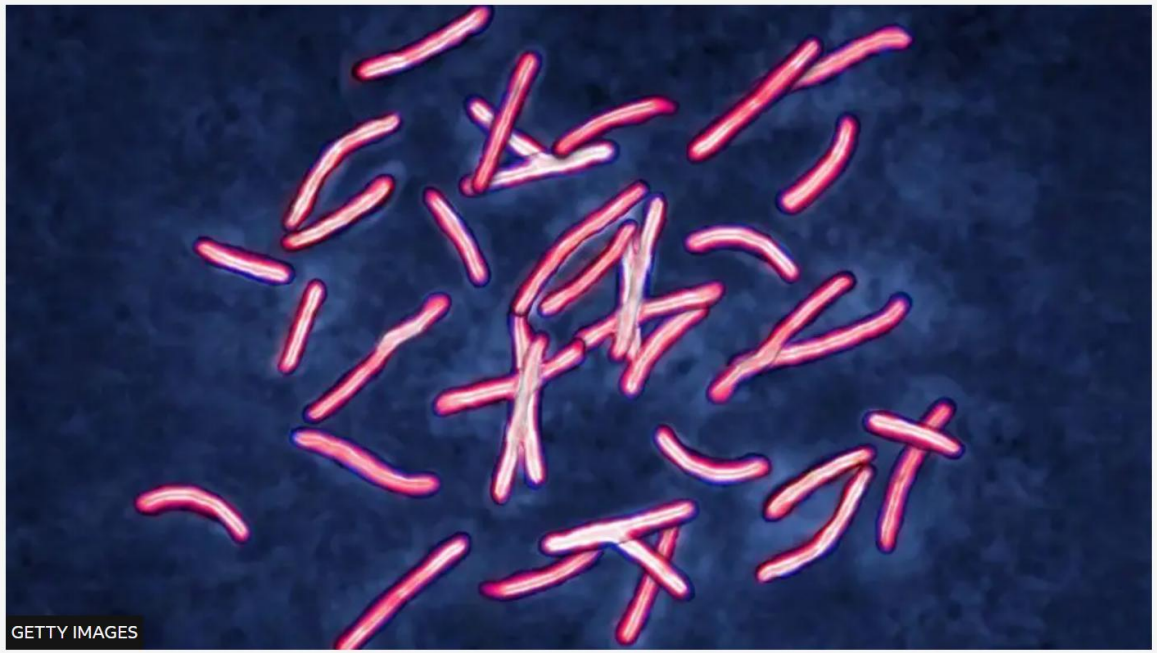


Built for collaboration to help **experts**, not aimed at automating the scientific process

DeepMind's **AI Co-Scientist** is not publicly available for download or open-source use; access is currently limited to a **Trusted Tester Program** involving select research organizations and academic institutions

AI Co-Scientist: Case Study

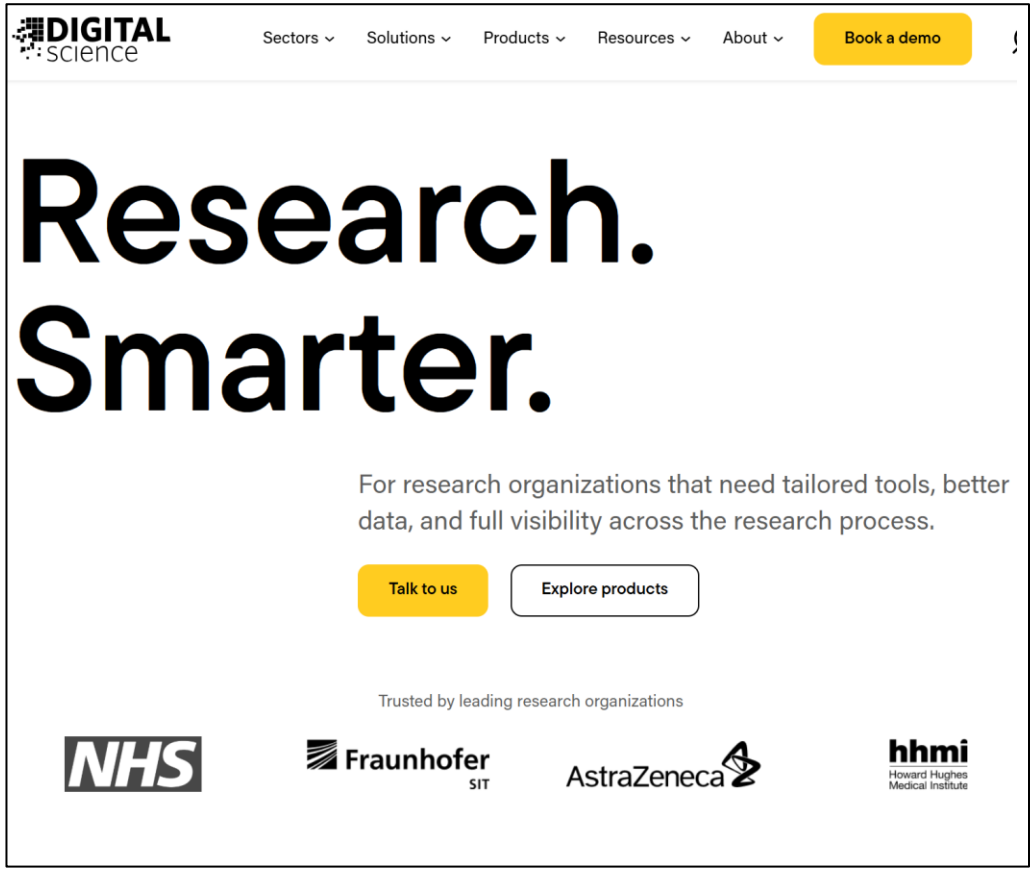
AI cracks superbug problem in two days that took scientists years

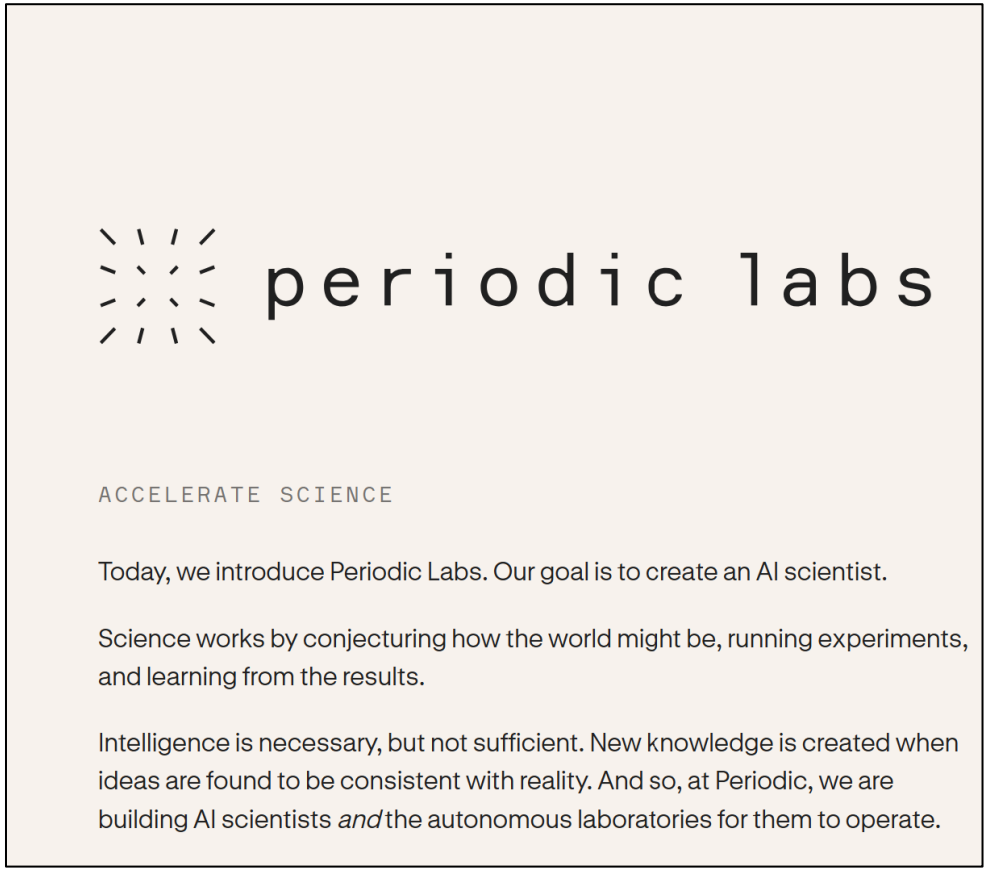


Cases of tuberculosis (pictured) have increased in the UK and worldwide as the disease increases its resistance to antibiotics

- Why are some superbugs immune to antibiotics?
- Spending considerable time identifying scientific questions, to ask, designing the experiments to answer them, and evaluating the results.
- Imperial researchers: Presented AI Co-Scientist with Qs and a hypothesis they have been investigating for a decade and know the answers to from lab experiments
- The tool generated the same hypothesis (how superbugs get created), and provided another four, all of them made sense (one completely novel) in just two days!
- Remarkably the research was not published, so could not have been found by the tool in the public domain

Direction of Travel: LLMs and Tools for Scientific Research







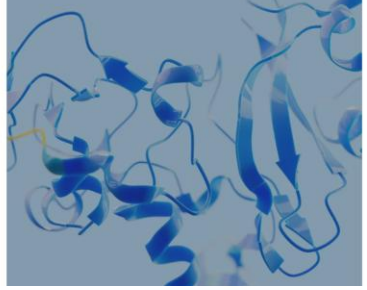


Frontier AI Tools for Science

Google DeepMind

Biology | Climate & sustainability | Mathematics & computer science | Physics & chemistry

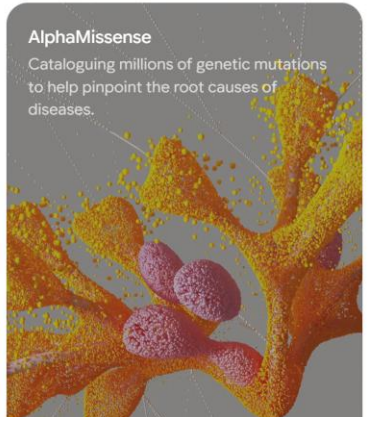
AlphaFold
Revealing millions of intricate 3D protein structures, and helping scientists understand how life's molecules interact.



AlphaGenome
A unifying DNA sequence model that advances regulatory variant-effect prediction.

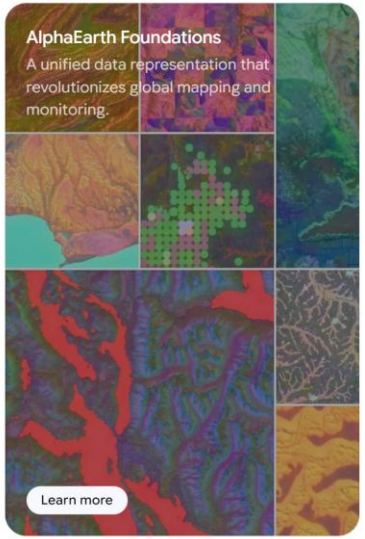


AlphaMissense
Cataloguing millions of genetic mutations to help pinpoint the root causes of diseases.



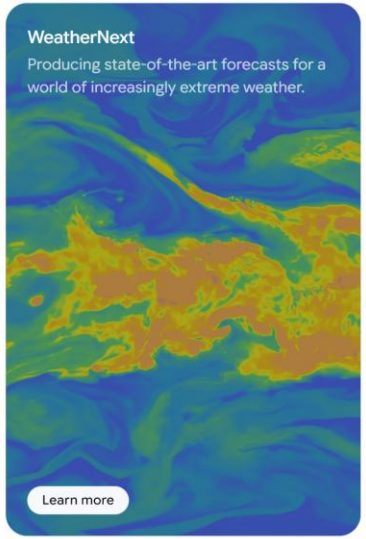
Biology | Climate & sustainability | Mathematics & computer science | Physics & chemistry

AlphaEarth Foundations
A unified data representation that revolutionizes global mapping and monitoring.



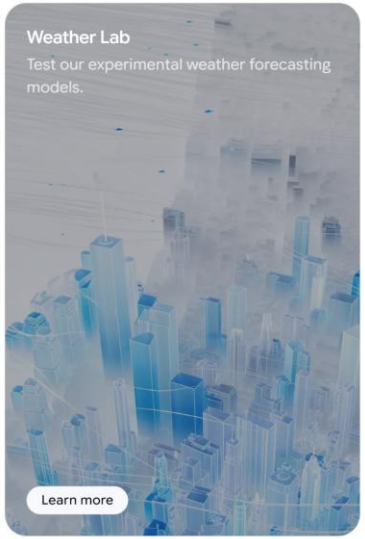
Learn more

WeatherNext
Producing state-of-the-art forecasts for a world of increasingly extreme weather.



Learn more

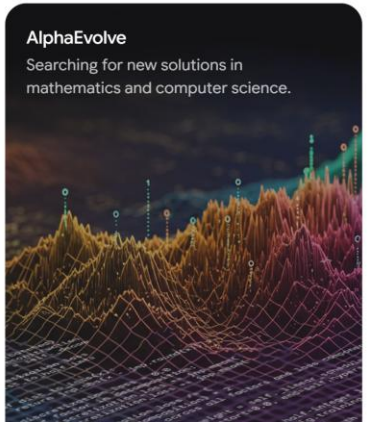
Weather Lab
Test our experimental weather forecasting models.



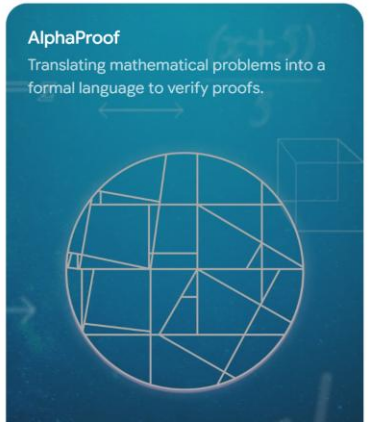
Learn more

Biology | Climate & sustainability | Mathematics & computer science | Physics & chemistry

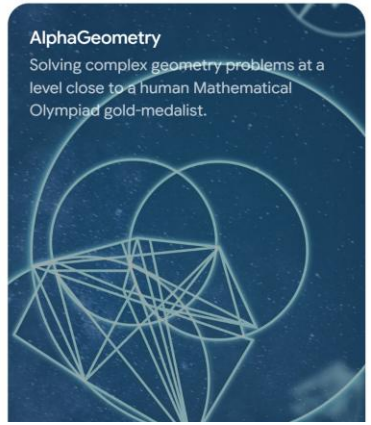
AlphaEvolve
Searching for new solutions in mathematics and computer science.



AlphaProof
Translating mathematical problems into a formal language to verify proofs.



AlphaGeometry
Solving complex geometry problems at a level close to a human Mathematical Olympiad gold-medalist.



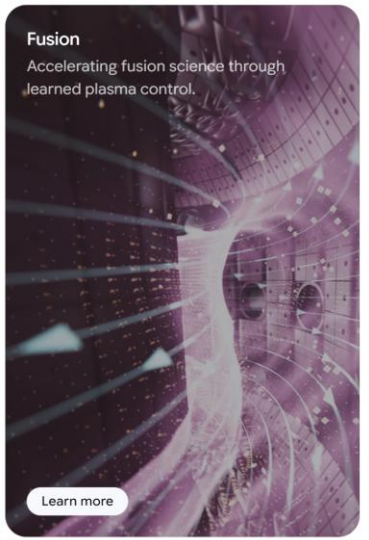
Biology | Climate & sustainability | Mathematics & computer science | Physics & chemistry

GNoME
Speeding up the discovery of inorganic crystals, essential components for technology like solar panels and batteries.



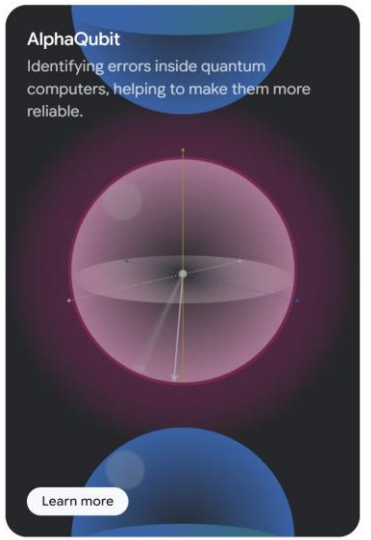
Learn more

Fusion
Accelerating fusion science through learned plasma control.



Learn more

AlphaQubit
Identifying errors inside quantum computers, helping to make them more reliable.



Learn more

Frontier AI Tools for Science (OpenAI)

Frontier models OpenAI's most advanced models, recommended for most tasks.



GPT-5.2
The best model for coding and agentic tasks across industries



GPT-5 mini
A faster, cost-efficient version of GPT-5 for well-defined tasks



GPT-5 nano
Fastest, most cost-efficient version of GPT-5



GPT-5.2 pro
Version of GPT-5.2 that produces smarter and more precise responses.

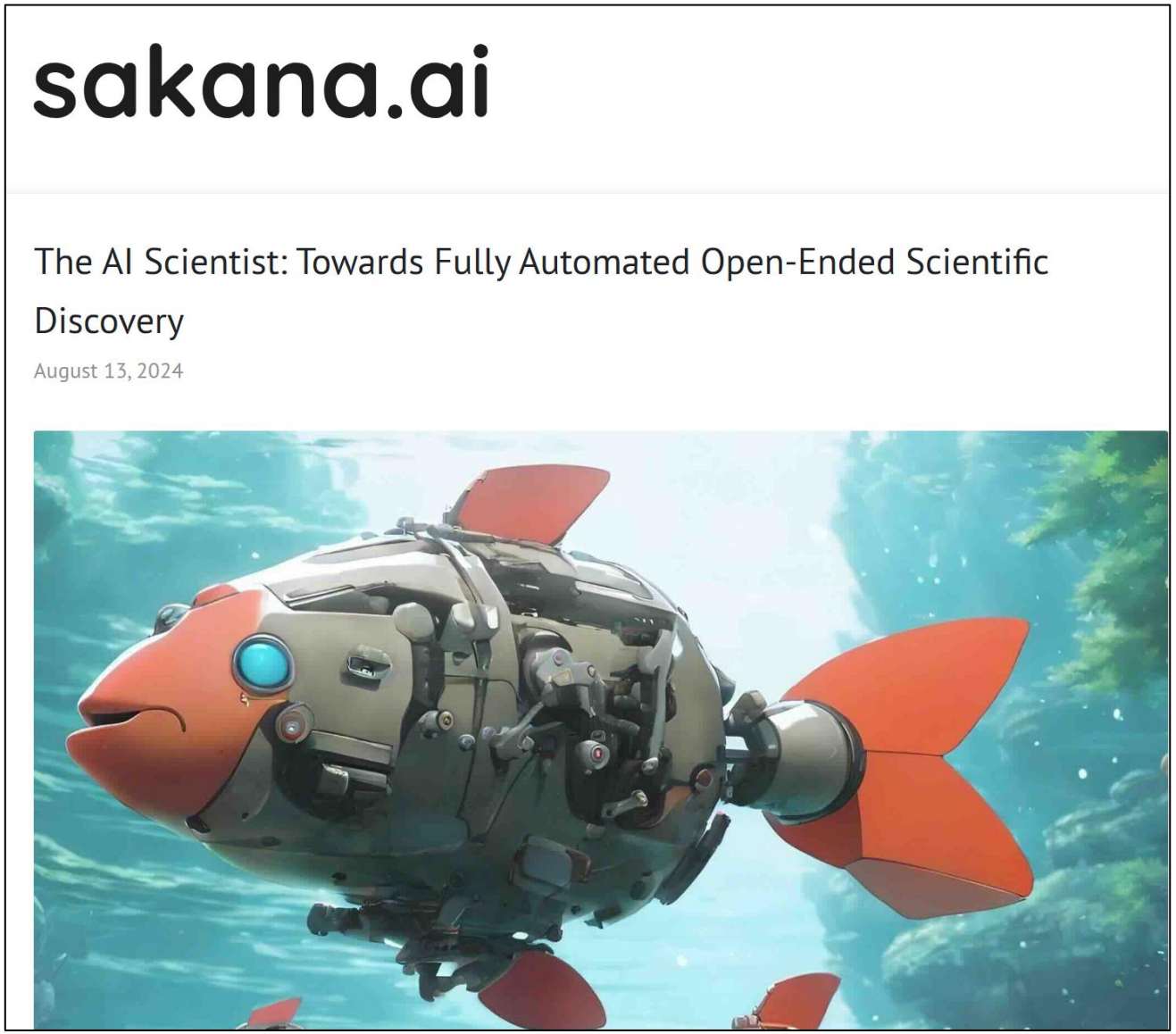


GPT-5
Previous intelligent reasoning model for coding and agentic tasks with configurable reasoning effort



GPT-4.1
Smartest non-reasoning model

Automation of Entire Research Process



- **AI Scientist-v2** is an open-source platform automating the entire research process
- Demonstrated to generate peer-reviewed research papers autonomously. The code is available on GitHub: arXiv+2arXiv+2
- Each experiment costs $15–$20, with an additional $5 for manuscript generation!

Engagement with LLMs: Critical Thinking and Philosophy of Science

- Critical thinking: systematic way of questioning and understanding the “why” and “how”- assumptions, constraints, arguments/counter arguments, alternative perspectives leading to deeper understanding of problems, identify bias sources, and for better outcomes

- Explicit learning outcome in humanities, implicit in STEM

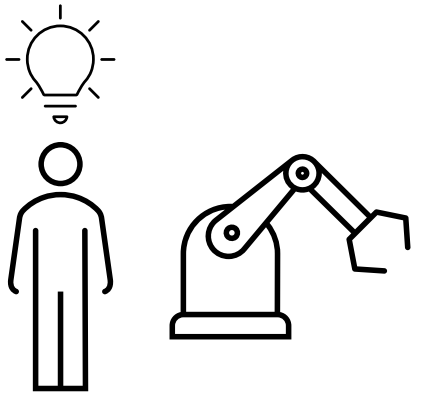
- Philosophy of Science: Examines how scientific knowledge is generated, justified and study methods and assumptions that make science reliable

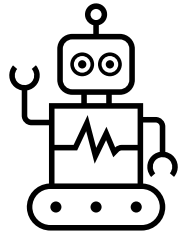
- Critical thinking and domain expertise are vital to engage with and evaluate LLM outputs effectively



Partial or Full Automation of Scientific Discovery?

- Humans decide which questions matter and worth solving so far!
- Breakthroughs come from challenging assumptions, connecting fields in unexpected ways
- Human scientists are accountable for decisions, implications, communications
- Even if full automation (black-box, brute force) of scientific discovery becomes possible, it is neither realistic nor desirable*
- **Observation 6**: Scientific discovery is hybrid: human-led (intuition and creative hypothesis), AI-accelerated (pattern recognition and data processing) enabling trust and avoiding a future where science must choose between innovation and the environment





*OWKIN, BIOPHARMA'S AGENTIC AI MOMENT

....Industry's focus on data curations, accelerate regulatory processes and scientific communication, rather than breakthrough scientific work due to trust in results.....

The Future: Understanding the Role of AI

Artificial Intelligence and Philosophy BSc

UCAS code: G704

Overview | Teaching & Modules | Entry requirements | Fees | Careers

Overview



Artificial Intelligence for Science MSc

Overview | Teaching & Modules | Entry requirements | Fees | Careers

Overview

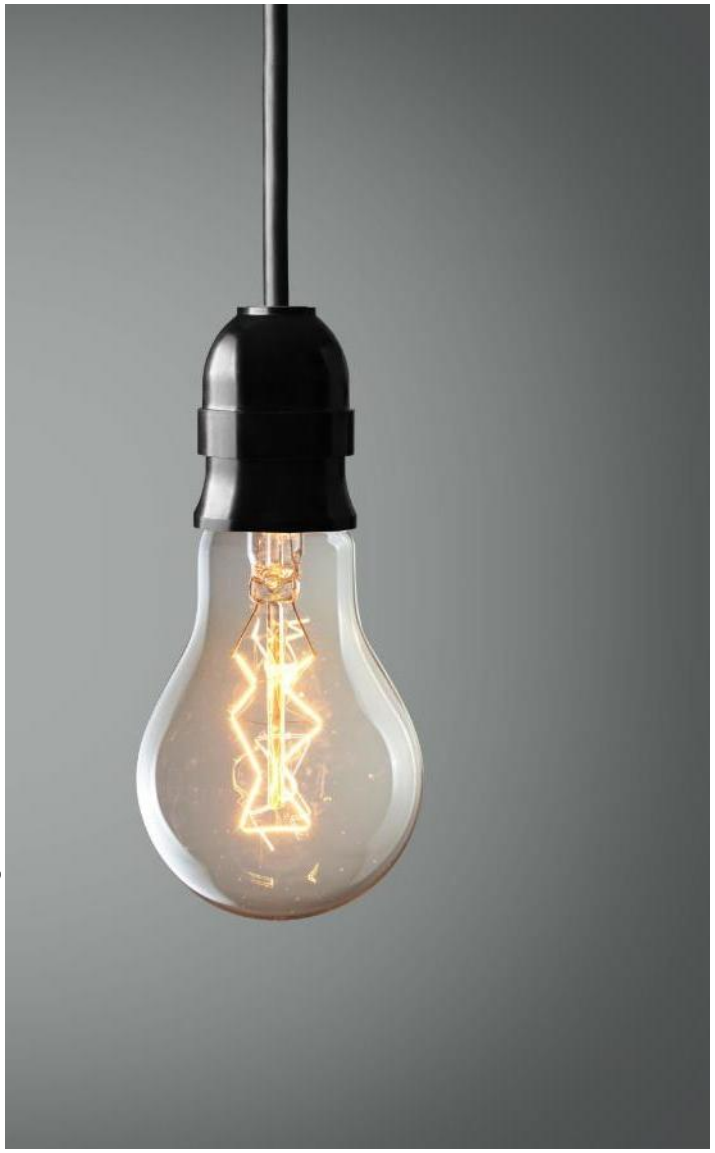


Observations

Observation 7: Humans as the Filter: humans assess LLMs outputs for usefulness and accuracy using expertise, reasoning and judgment, serving as LLMs peer reviewers

Observation 8: LLMs as Critical Friend: treat LLMs like a collaborator who challenges thinking to critique our work to stimulate different perspectives (novelty, assumptions validity..), identify flaws in explanation and design…

Observation 9: LMMs can accelerate domain expertise (how is domain expertise established?), LLMs are not the absolute truth and their values lie in their role as force multipliers



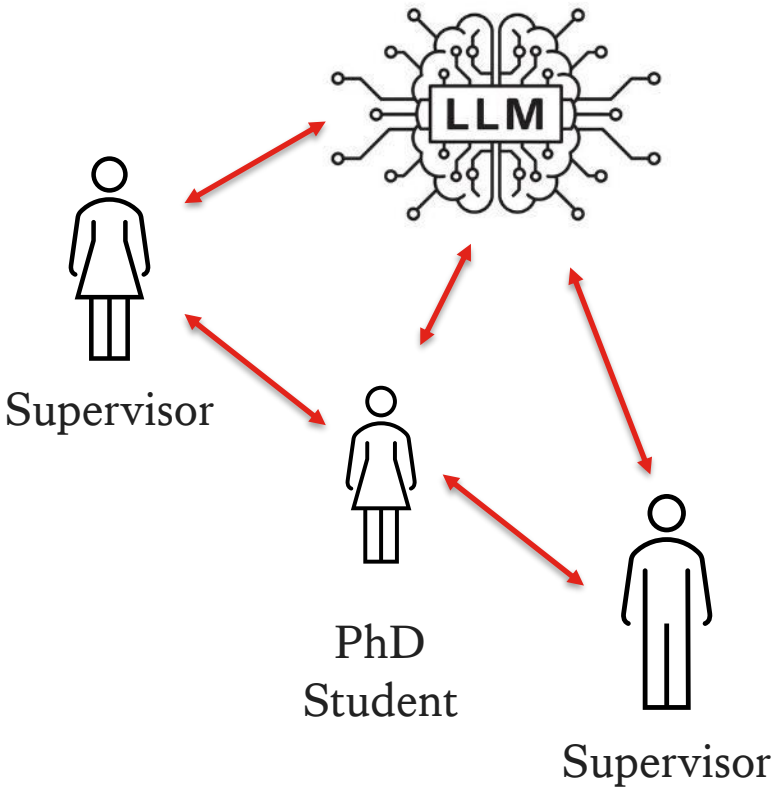
LLMs are Stochastic

- LLMs outputs have randomness with tokens predicted with probability distributions
- Randomness can be controlled through model parameters
- Computational tools and LLMs **both** have limitations
- Computational tools limitations or **errors** are explainable and grounded in explicit assumptions, approximation, and model known physical laws
- LLMs limitations or **uncertainties** are often not explainable, inaccuracies arise from training data gaps, probabilistic approach, but errors in language undermine **trust**

Parameter	Purpose
`temperature`	Randomness
`top_p`	Sampling diversity
`max_tokens`	Output length
`presence_penalty`	Topic diversity
`frequency_penalty`	Repetition control
`stop`	Stop sequences
`logit_bias`	Token control
`seed`	Reproducibility

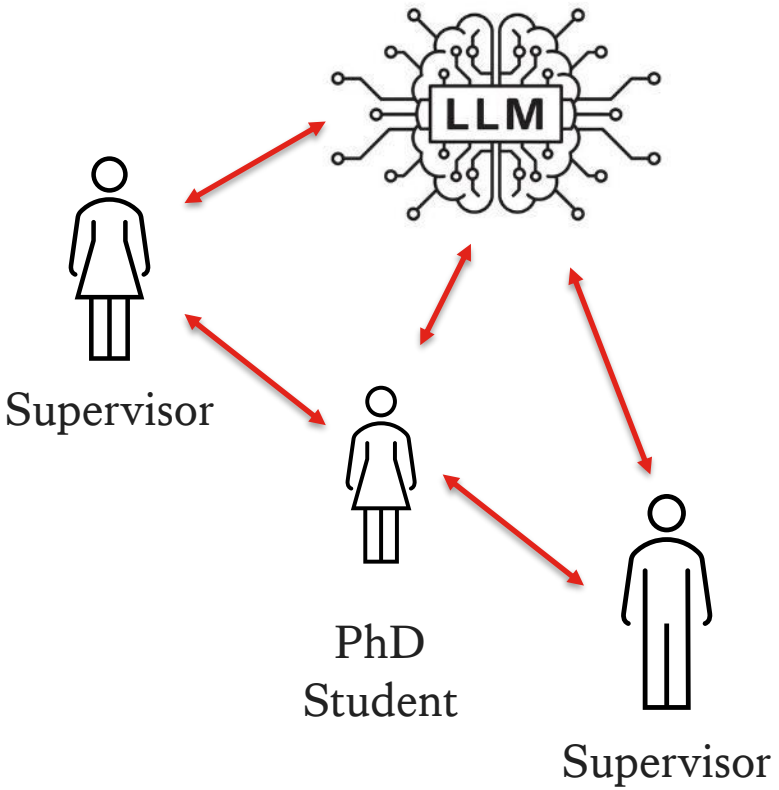
LLMs as PhD Co-Supervisors, Co-Authors, Co-Creators…

- PhD supervision involves academic and non-academic guidance and support
- **Observation 10**: Supervision shift from 2-way to 3-way partnership (Trialogue) based on trust and collaboration
- Supervisor provides wisdom and values, student provides curiosity and work, and LLM provides computational intelligence and knowledge access at scale
- Supervisors play crucial role in balancing the partnership ensuring intellectual growth and safe and responsible use



LLMs as PhD Co-Supervisors, Co-Authors, Co-Creators…

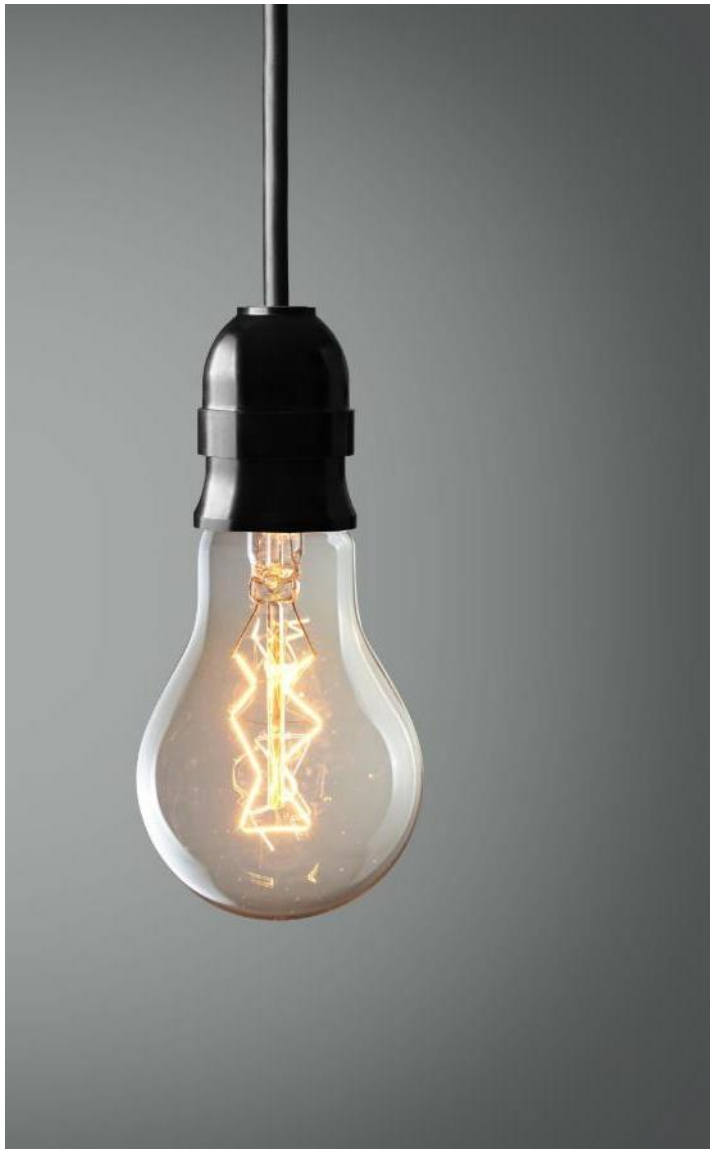
- PhD is based on making intellectual contribution: scholarly outputs to advance knowledge or practice
- Can LLMs assist with intellectual contribution? This depends on how "intellectual" is defined
- **Authorship** is about intellectual contribution, accountability and responsibility, LMMs can't take responsibility and held accountable, this does not meet authorship criteria



Observations

Observation 11: Research training should evolve as LLMs become integral to how research is conducted. With training, policies, and oversight, LLMs can support research and supervision rather than weaken them, despite people are reluctant to admit using LLMs!

Observation 12: With more automation and higher LLMs capabilities, humans gain more time to push the boundaries of discovery thinking deeply about research questions, and to maximise the innovation impact



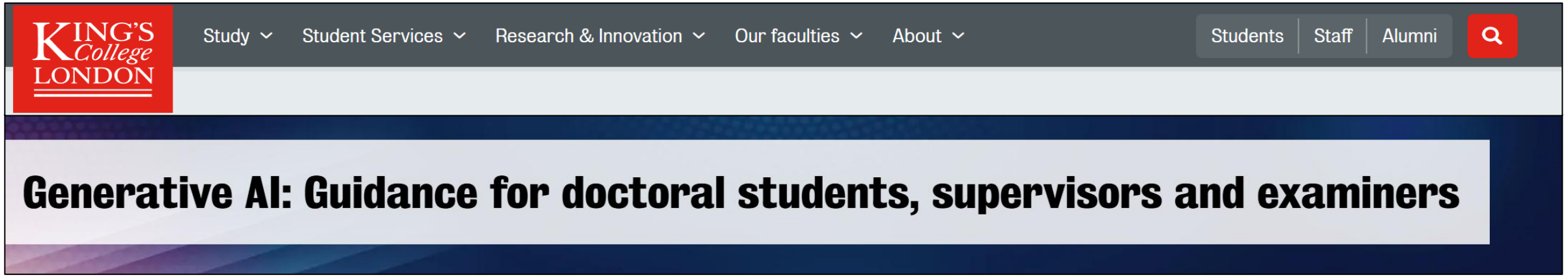
Working with LLMs: possible outline

- **Critical thinking** embedded in prompt engineering and hands-on exercises (e.g. counter-questioning, test reasoning, explore multiple perspectives and assumptions)
- **Technical**: Use different LLMs, APIs, Python programming. Statistics and ML fundamentals. For advanced users how to access AI compute and secure data storage
- **Policy**: Understanding and application of research ethics, research integrity, and university AI policies



Russell Group Principles on the Use of Generative AI Tools in Education

1. Universities will support students and staff to become AI-literate.
2. Staff should be equipped to support students to use generative AI tools effectively and appropriately in their learning experience.
3. Universities will adapt teaching & assessment to incorporate ethical use of generative AI and support equal access.
4. Universities will ensure academic rigour and integrity is upheld.
5. Universities will work collaboratively to share best practice as the technology and its application in education evolve



Are students allowed to use generative AI tools when writing their thesis?

Yes, students are permitted to make use of generative AI tools in their thesis writing processes. Students should declare use of generative AI tools and take full responsibility for the content of their submitted thesis.

Position on the use of AI in...

ELSEVIER

Academic & Government Health Industry Insights About Support

Home > About > Elsevier policies > Generative AI policies for journals

Generative AI policies for journals

About For authors For reviewers For editors Publication process Frequently asked questions

These policies were initially triggered by the rise of generative AI* and AI-assisted technologies, which were expected to increasingly be used by researchers and have now been updated to reflect evolving good practice. These policies aim to provide greater transparency and guidance to authors, reviewers, editors, readers and contributors. Elsevier will continue to monitor developments in this area and will adjust or refine policies as appropriate.

Highlight #ARTIFICIAL INTELLIGENCE

Current position of the ERC Scientific Council on AI

UK Research and Innovation

Use of generative AI in application preparation and assessment



Subscribe to UKRI emails

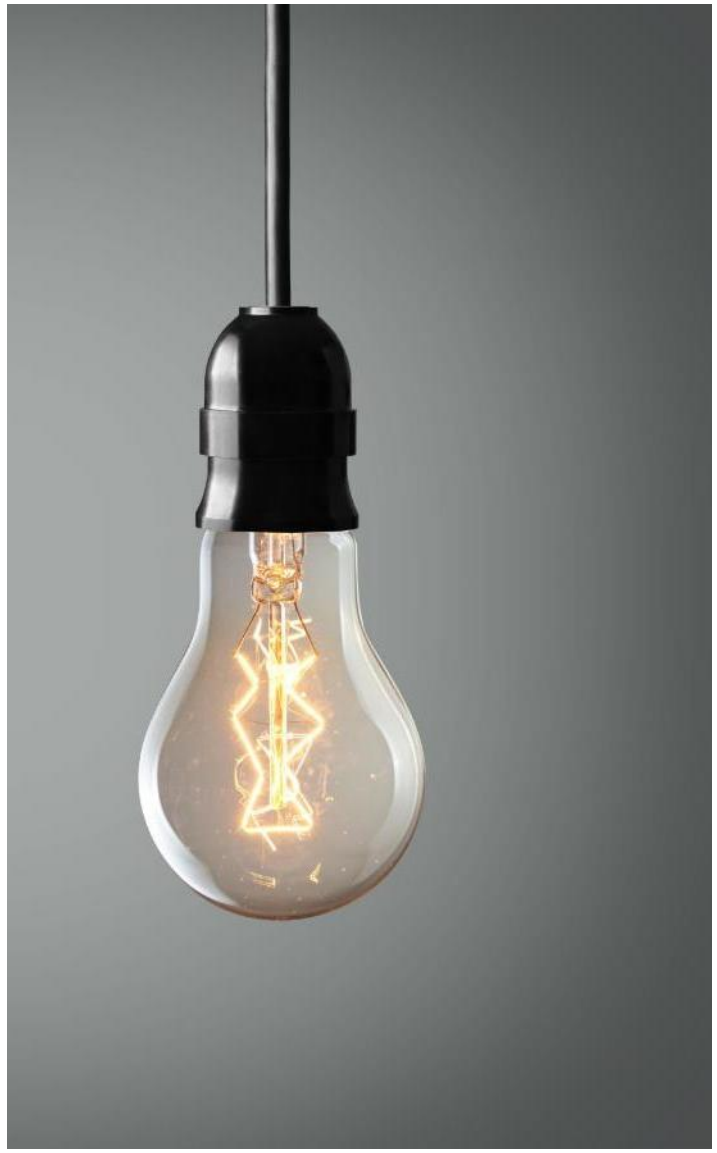
Keep up to date with funding, news and events, and a weekly newsletter.

Subscribe

Further Observations

The boundary between the human and the artificial does not seem so easy to draw

Human-Technology collaboration should be based in framing machine intelligence in the context of critical thinking, scholarly peer review and building of trust

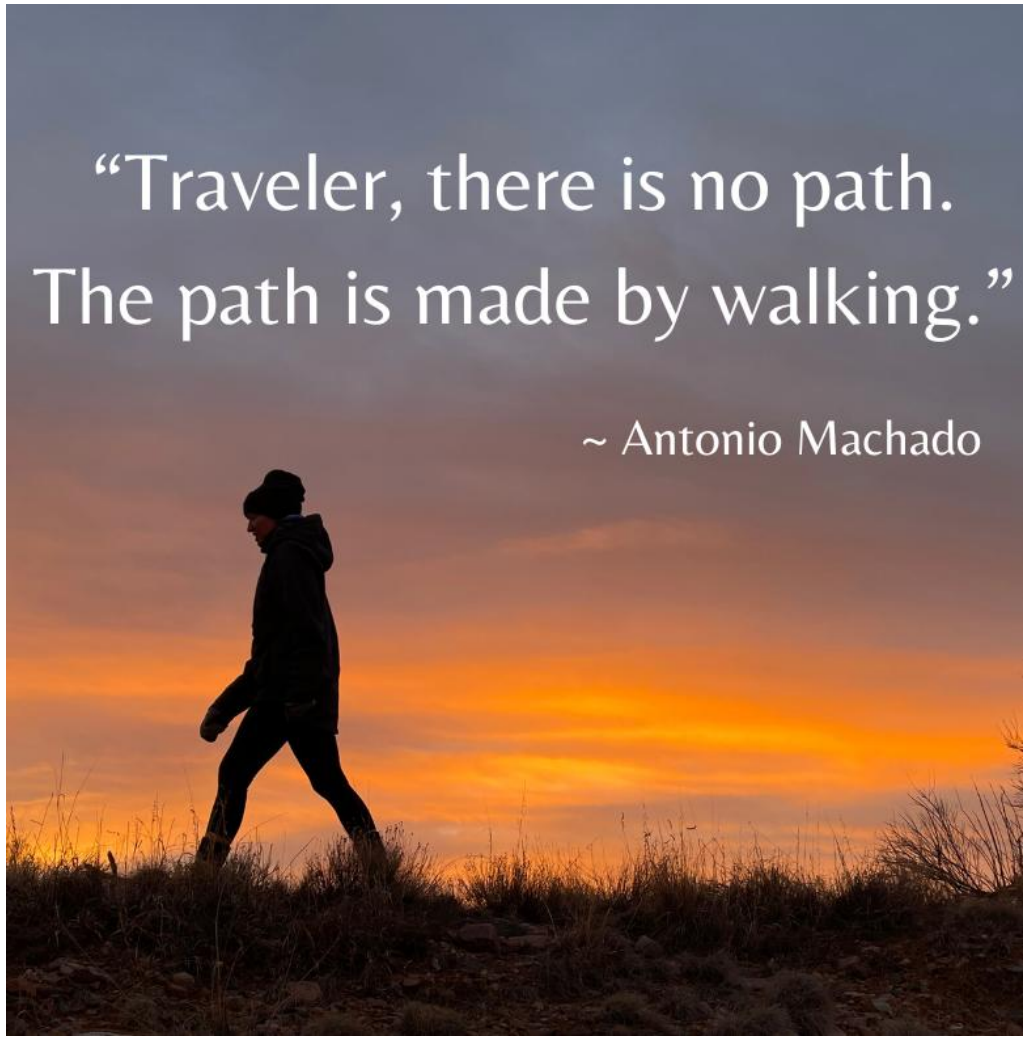


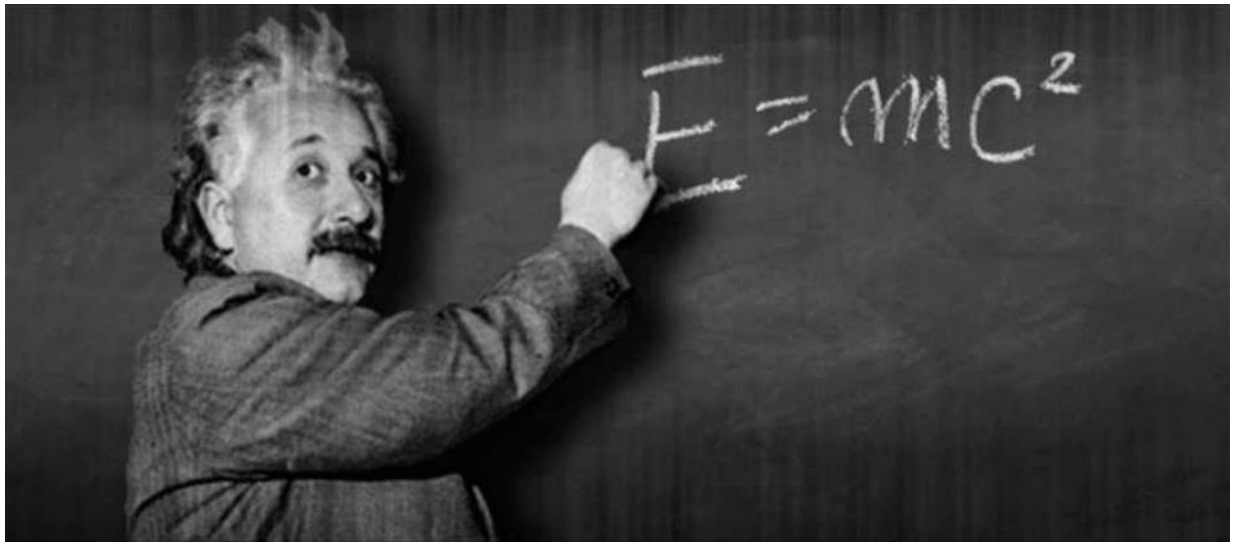
Implications

Combining biological intelligence (human intuitions, hypothesis) with artificial intelligence (pattern recognition, data processing) will lead to:

- PhD co-supervision and research co-creation
- Amplification and acceleration of domains expertise and knowledge
- Discovery science will increasingly be guided by societal wicked problems
- Future of science shouldn't be a choice between innovation and environment







“The world as we have created, it is a process of our thinking. It cannot be changed without changing our thinking.” - Albert Einstein

Is AI in Research a Wicked Problem?

A **wicked problem*** describes complex, ill-defined problems that are difficult or impossible to solve because it has no clear solution, changing and contradictory requirements, and many interconnected factors.

Characteristics of wicked problems include:

- No clear stopping point or final solution.
- Solutions are not true or false but better or worse
- Every solution attempt has consequences that may create new issues.



Wicked problem was coined by **Horst Rittel and **Melvin Webber** in 1973. They introduced the concept in their paper on social policy and planning.*

The Wicked Problem of AI in (Educational) Assessment

- Drawing on interviews with 20 teachers responsible for assessment design within a large Australian University, this paper applies the framework of wicked problems to analyse educators' experiences with assessment design in the context of GenAI.

- The findings demonstrate that the GenAI-assessment challenge exhibits all ten characteristics of wicked problems. It emphasizes the need for flexible and context-specific approaches rather than seeking a one-size-fits-all solution.



https://www.tandfonline.com/doi/full/10.1080/02602938.2025.2553340