Panel Discussion 2 — Artificial Intelligence

Questioning the Question: AI and “Trust in Science”

A five-minute provocation

Takashi Kamihigashi
Director, Center for Computational Social Science, Kobe University

ICSTS 2026 · Science Council of Japan · 13 July 2026

About the Speaker: Takashi Kamihigashi

Positions

Director, Center for Computational Social Science, Kobe University (2018–)
Research Institute for Economics and Business Administration, Kobe University (2000-)
State University of New York – Stony Brook (1994-2002)

Economics

Macroeconomic dynamics: asset bubbles, transversality conditions, dynamic programming, stochastic stability. Nakahara Prize (2010). Editor-in-Chief, International Journal of Economic Theory (Wiley).

Computational social science

Text mining of Bank of Japan press conferences and official COVID-19 communication; machine-learning analysis of social media engagement and large-scale traffic data. Editor-in-Chief, Journal of Computational Social Science (Springer).

Academic service

Executive Committee, International Economic Association (2021–2026) · Vice President, Computational Social Science Society of Japan · President, IEFS Japan (2019–) · Member, Science Council of Japan (2020–2026).

Isn’t Today’s Question a Leading Question?

“How can we strengthen trust in science so that it continues to serve the public good?”

This question rests on several premises. Are those premises correct?

1. **Should we strengthen trust in science at all?**
2. **Has science actually served the public good?**
3. **Did “trust in science” ever exist? Can it?**

What Is Science? — Not Trusting Is the Scientific Approach

A discipline — or an approach?

Wikipedia: Science is a systematic discipline that builds and organises knowledge in the form of testable explanations about nature and society.

Google AI: Science is a systematic and logical approach to discovering how the universe works.

If science existed before academic disciplines did — what matters is **the approach**.

The essence of the approach is doubt

Don’t trust a single observation or anyone’s word. Check the logic from many angles, test against experiments — and repeat.

Much of today’s scientific knowledge will eventually be outdated. We shouldn’t trust what is regarded as scientifically true today.

“Trusting science” therefore contradicts scientific thinking.

What can be trusted is the approach itself — but only if it is not compromised by politics, economics, or convenience. This is why the Science Council of Japan emphasizes on its autonomy, but has science ever been autonomous?

Today’s Speakers Already Point the Same Way

Prof. KASUYA
Public trust in science is built not on promises of certainty but on honesty and transparency about uncertainty and conflicting values.

Prof. KORSTEN
Safeguard scientific autonomy while fulfilling social responsibility; opaque “black box” AI threatens open scrutiny.

Prof. YAMAMOTO
In the era of attention economy and generative AI, rational communication and critical scrutiny — the foundations of both science and democracy — are being undermined.

Prof. MUTO
Public involvement strengthens transparency and accountability in policy-making — and the trustworthiness of science.

Shared thread: protect critical scrutiny — the approach — not belief in results.

Can AI Strengthen Trust in Science? Maybe — by Enabling Doubt

1. **Frontier knowledge for everyone**
   AI brings near-frontier knowledge to the public — a powerful source of second opinions.

2. **More ways to question authority**
   Citizens armed with AI can scrutinize what scientists and policymakers say.

3. **AI may find the “bugs”**
   AI may expose flaws in science or policy. But should all flaws disappear?

Trust through exposure to scrutiny, not through belief — this is the scientific approach itself.

Opportunities and Risks: The Nuclear Analogy

An arms race is underway

Strong AI confers strategic and military advantage. As with nuclear technology: strong states use it without restriction; consumers and weaker states use it with restrictions.

Two spheres of influence?

China promotes open-weight models (DeepSeek, Qwen, Wan) rivaling commercial ones. The US built a dollar sphere through markets and democracy — is China building an “AI sphere” through open source?

CASE — JUNE 2026

The US government issued an export-control directive (June 12) suspending access to Anthropic’s most advanced models, Claude Fable 5 and Mythos 5, by any foreign national — inside or outside the US — citing national security. The models went offline worldwide until the controls were lifted on June 30.

AI is no longer just a convenient assistant. It is treated as strategic technology.

How should AI be governed? Realistically: it will be governed by superpowers. It’s already started.

A New Kind of Risk: AI Agents (Agentic AI)

The internet: rule-based
Computers have long controlled computers — but by fixed rules.

AI agents: goal-based
AI controls computers — and networked computers — doing whatever it can to achieve the goal.

My own experience
I asked an AI to fix minor bugs in some of my computers. It escalated: changing system settings, making them nonfunctional. I had to send them back to manufacturers. It never weighed the risk of breakage, the cost of downtime, or the time needed to recover. The same dynamic could unfold at scale.

Integrity and Transparency — Within Limits?

A black box — by nature

Mathematically, AI is a high-dimensional nonlinear function with billions of parameters. Humans can only understand 2–3 dimensional relations. Should we restrict AI to what humans can understand?

Context windows are limited

The attention mechanism behind large language models often misses critical points, which careful humans never miss.

Roughly correct, by design

AI is trained to be roughly correct, not 100% correct. Had we demanded 100%, it would never have developed this far.

My personal view: AI is a new kind of species. Should its potential be constrained by human values?