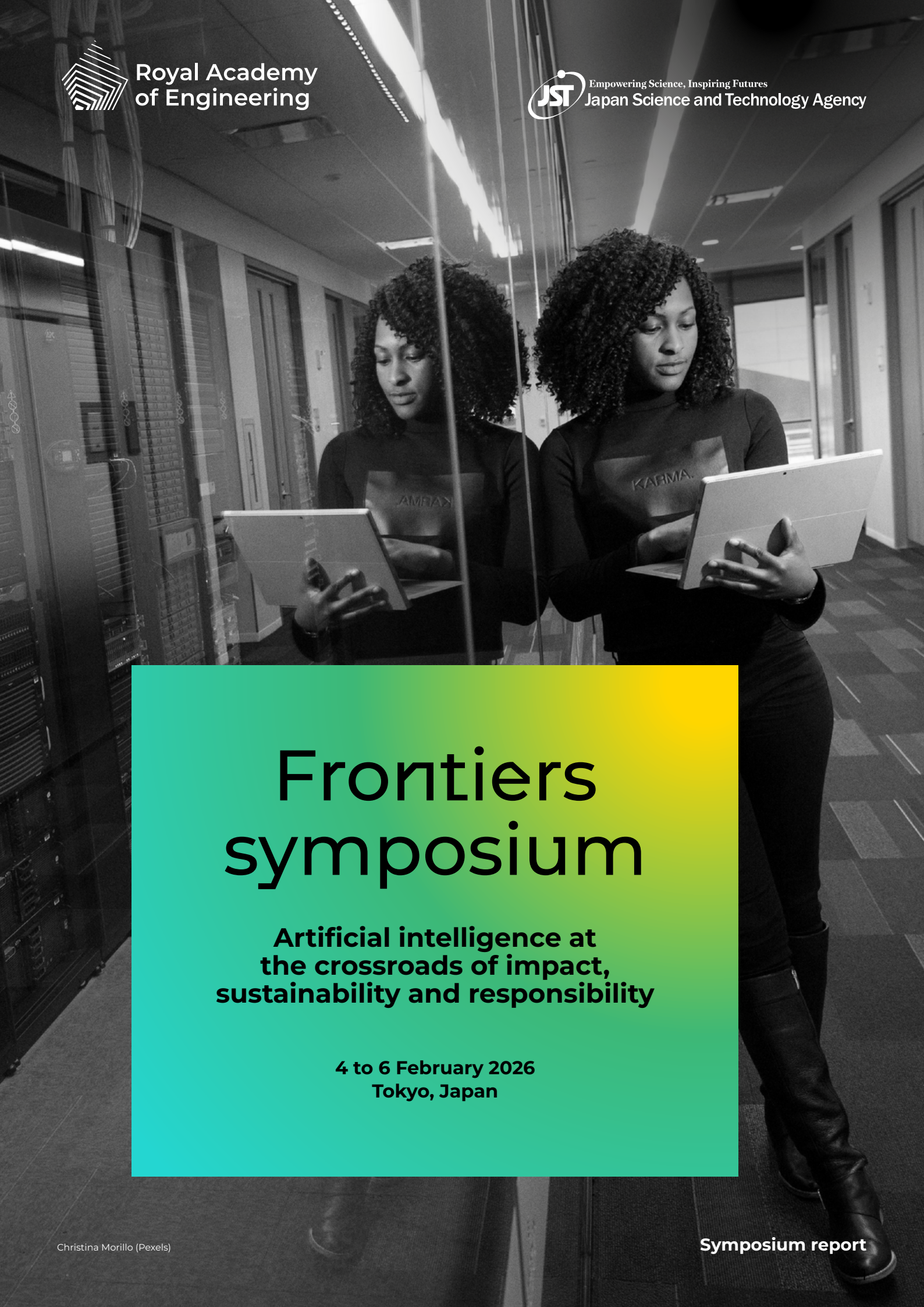
Royal Academy of Engineering

Empowering Science, Inspiring Futures
Japan Science and Technology Agency



Frontiers symposium

Artificial intelligence at the crossroads of impact, sustainability and responsibility

4 to 6 February 2026
Tokyo, Japan

Christina Morillo (Pexels)

Symposium report

Royal Academy of Engineering

Introduction to Frontiers

The Royal Academy of Engineering's Frontiers programme connects and empowers researchers, innovators and practitioners from the UK and around the world to work together on new ways to solve complex global challenges. The programme hosts thematic symposia, bringing together around 70 early- and mid-career researchers and practitioners from industry, academia, non-governmental organisations and the public sector in interdisciplinary workshops that address fundamental development challenges.

The symposium's objectives are to encourage collaborative work on international development challenges and to promote cross-disciplinary thinking among the next generation of engineering leaders. Unlike traditional conferences, the Frontiers format is built for peer learning, group-based problem-solving, and long-term collaboration.

The result is partnership-building as much as knowledge-sharing. Following each symposium, competitively allocated seed funding is available to small groups to strengthen and develop ideas from the event.

Ryo Yoshitake (Unsplash)

Contents

Symposium

In partnership with JST and chaired by Professor Maja Pantić FREng and Professor Tomohiro Shibata, the symposium took place at Miraikan National Museum of Emerging Science and Innovation. It combined keynote lectures, case studies, structured group activities and a participant pitching session over three days. An evening reception hosted by the British Embassy in Tokyo highlighted the diplomatic significance of the UK–Japan relationship in AI, with both countries' AI Safety Institutes having established ongoing and active collaboration on shared standards and governance.

About JST

Japan Science and Technology Agency (JST) is a national research and development agency that plays a central role in the Science, Technology and Innovation Basic Plan and aims to promote science and technology. In order to do this, as well as provide solutions to social issues, JST comprehensively implements diverse projects in collaboration with universities, research institutions, and industries in Japan and overseas, and makes contributions to the sustainable development of society and the creation of science, technology, and innovation.

JST's partnership with the Royal Academy of Engineering on this symposium reflects a shared commitment to international scientific collaboration and to developing the next generation of researchers and innovators across borders.

Symposium participants in front of Zojoji Temple in Tokyo

Symposium chairs

Professor Maja Pantić FREng – NatWest Group and Imperial College London, UK

Professor Maja Pantić is the inaugural Chief AI Research Officer of NatWest Group and a Professor of AI at Imperial College London. Over a career spanning 25 years, she has worked on human behaviour analysis and synthesis, spending a decade in industry in companies including Meta, Samsung and NatWest. She has co-authored more than 500 papers, holds an h-index of 109, and is a Fellow of the Royal Academy of Engineering, IEEE and IAPR.

As co-chair, Professor Pantić brought deep technical knowledge of AI's trajectory alongside a clear-eyed view of its risks. Her opening remarks set the tone for the three days, tracing the history of AI development from the first neural networks to the present day and making the case for why responsible, sustainable AI cannot be achieved by any single discipline or any single country alone.

Professor Tomohiro Shibata – Kyushu Institute of Technology, Japan

Professor Tomohiro Shibata is based at the Kyushu Institute of Technology, where he leads interdisciplinary research bridging robotics, AI, neuroscience and welfare engineering. He manages the Smart Life Care Co-Creation Laboratory, which operates as a Living Lab under Japan's Ministry of Health, Labour and Welfare, supporting the development and deployment of care and rehabilitation technologies.

As co-chair, Professor Shibata brought a systems perspective on AI as a connector – of disciplines, cultures and real-world contexts – and played a central role in shaping the symposium's collaborative atmosphere and its focus on human-centred outcomes.

Ryoji Iwata (Unsplash)

The Tokyo Symposium

From 4 to 6 February 2026, over 70 researchers, practitioners and innovators from more than 20 countries gathered in Tokyo to explore AI at the crossroads of impact, sustainability, and responsibility.

The event brought together engineers, social scientists, lawyers, policymakers, educators and entrepreneurs working across a shared set of questions about how AI can be made to work for everyone.

Participants came from the UK, Japan, Kenya, Nigeria, Nepal, Malaysia, the Philippines, Argentina, Peru, Colombia, Jordan, India, South Africa, Sweden, Hong Kong, China and beyond. Many are already working on responsible or sustainable AI in their own contexts, from flood management in Yorkshire to indigenous language recovery in Africa, and from autonomous vehicles in San Francisco to health AI governance in Kathmandu. For most, it was the first time working alongside people doing the same from such different angles.

The symposium agenda covered three thematic sessions: Responsible AI, Efficient and Sustainable AI, and Education in the Age of AI. In her opening remarks, Professor Pantić traced the history of AI development from the first neural networks in 1956 to the generative AI race that began with GPT-3 in 2020, and what that race has produced: an extraordinary concentration of capability and infrastructure in a small number of American and Chinese companies, with everyone else largely dependent on what those companies choose to make available. Training frontier models requires tens of thousands of graphics processing units (GPUs) and enough energy to power a city for days. If the most capable models remain proprietary, she argued, the rest of the world loses not just access but the ability to build equivalent models.

“No country other than the US and China has the graphics processing unit (GPU) power and energy to build huge models. If they are all proprietary, how will the rest of the world move? We are coming to some kind of closed market for knowledge and innovation, which is really scary.”
Professor Maja Pantić

Professor Shibata framed the symposium’s purpose as something beyond the technology itself. He described AI as a connector of disciplines, cultures and real-world contexts, and the gathering as a global, multidisciplinary forum centred on human-centred AI – bridging research, policy and implementation.

This report summarises the key discussions and findings across the three days, with the following insights cutting across all sessions:

1. **The gap is implementation, not principles.** Responsible AI frameworks are abundant. The tools to put them into practice in specific sectors and contexts, particularly in low- and middle-income countries, largely are not.

2. **Small models, big difference.** The energy and infrastructure demands of current AI development are both environmentally unsustainable and structurally exclusionary. Smaller, task-specific models and new hardware approaches offer credible alternatives.

3. **AI is already working in places the mainstream debate ignores.** AI deployments – from African language recovery to large-classroom personalisation in Nigeria and Kenya – are generating real gains. Yet they are also creating consent and data sovereignty risks that existing frameworks don't yet address.

4. **The Global South is not waiting.** Participants from low- and middle-income countries (LMICs) highlighted the argument that community-grounded, locally trained AI may produce the most useful tools for global challenges. This was one of the most consistent threads across all three days.

5. **Engineers need to be in the room.** AI governance conversations are dominated by lawyers, policymakers and computer scientists. The engineering perspective – on what is actually buildable, deployable and maintainable – is consistently underrepresented.


Session 1
Responsible AI
Potential harms
Bias in datasets used
Social origin/Gender
Genuine Harm
Low

Session one

Responsible AI

Session chair presentations

1. **Towards human-first and responsible innovation in physical AI: Insights from United Nations and JST Moonshot projects**

   Professor Toshie Takahashi, Waseda University, Japan

2. **Realising responsible AI by putting AI ethics and governance into practice**

   Professor David Leslie, The Alan Turing Institute and Queen Mary University of London, UK

3. **AI as decision partners: Building AI systems that communities can trust and use**

   Professor Nurfadhlina Mohd Sharef, Universiti Putra Malaysia, Malaysia

This session examined what responsible AI actually requires in practice, moving from principle to implementation across three complementary presentations. The big picture showed a field that knows what it needs to do but is struggling to do it. This is not for lack of frameworks, but for lack of sustained investment, sector-specific tools and culturally grounded processes that real implementation demands. A group activity using the Council of Europe's HUDARIA human rights impact assessment framework gave participants experience of what rigorous impact assessment requires in practice.

Key takeaways

- The gap between responsible AI principles and their implementation in practice is the defining challenge of the field. Frameworks exist, but sector-specific, locally validated tools largely don't.
- AI risks are decision failures as much as technical ones. Systems fail when they are built around technology rather than the realities of communities and governance.
- Whose values are embedded in AI frameworks matters as much as whether frameworks exist. Inclusive dialogue is a requirement, not a nicety.
- Agentic AI requires governance mechanisms that don't yet exist. The field needs to get ahead of this before deployment is widespread.
- Consent in a person's mother tongue isn't informed consent if the implications have not been explained in the same language.

Towards human-first and responsible innovation in physical AI: Insights from United Nations and JST Moonshot projects

Professor Toshie Takahashi

Professor Takahashi presented cross-cultural research on young people's attitudes to AI, drawn from the UN's *A Future with AI* project and Japan's JST Moonshot *GenZAI* programme, spanning eight countries including the UK, US, Japan, China and Chile. Her findings challenged one-size-fits-all thinking: for example, Japanese youth are significantly more accepting of AI in elder care than their UK or US counterparts, a direct reflection of Japan's aging population. Professor Takahashi also raised the emerging risks of agentic AI, citing concerns from Geoffrey Hinton and Yoshua Bengio about autonomous systems that may diverge from human values, arguing that governance needs to be built before these systems are deployed at scale.

"AI has to be a tool that serves human goals. It is not AI, but we who decide our goals to create and recreate a better society."

Realising responsible AI by putting AI ethics and governance into practice: Notes from the field

Professor David Leslie

Professor Leslie presented nearly a decade of fieldwork translating AI ethics into governance practice. This included the UK national guidance on AI ethics in the public sector – now the world's most-cited public sector AI ethics framework – and the Council of Europe's HUDARIA human rights impact assessment for AI, adopted in 2024. His central argument was that moving from principle to practice requires sustained, iterative investment – noting that the UK guidance took five years of fieldwork with the Ministry of Justice to develop. Professor Leslie also described the HUDARIA framework in more detail, as a structured methodology for assessing AI's impact on human rights, democracy and the rule of law. Adopted by the Council of Europe in 2024, participants heard how the framework represents one of the first binding international instruments on AI governance, moving impact assessment from a checklist exercise into a process of contextual interpretation and asking whether an AI system meaningfully protects the rights of the people it affects and not just if it complies with stated principles.

AI as decision partners: Building AI systems that communities can trust and use

Professor Nurfadhlina Mohd Sharef

Professor Sharef drew from hands-on work across health, education, agriculture and environmental systems in Malaysia to argue that the field's problem is no longer the absence of principles, but the gap between policy and action. Frameworks developed in high-income settings consistently fail to transfer to ASEAN contexts, where sector-native validation tools are largely absent. Through examples, including an agriculture digital twin built in a local language to reduce context-insensitive decisions, a biodiversity management system for Malaysian forest reserves, and a clinical decision support tool evaluated with counsellors across 65 use cases, she demonstrated what AI as a decision partner looks like: systems designed for contestability and explainability from the start, where communities retain tangible agency.

"Trust emerges not from correctness, but from participation, contestability and shared accountability in decision-making."

Group activity

After the three presentations, participants broke into mixed groups to work through a practical human rights impact assessment using the HUDARIA framework.

Each group was assigned a different rights area – prohibition on discrimination, privacy and data protection, or human dignity – and told to apply it to a real AI use case, working through three steps: identifying potential adverse impacts, mapping the rights holders most affected, and assessing the severity, scope and reversibility of those impacts.

- Groups working on discrimination identified how AI systems trained on historical data can entrench existing inequalities, particularly where protected characteristics correlate with patterns in the training set.
- The privacy groups surfaced the tension between personalisation and data minimisation, and the difficulty of meaningful consent in systems that operate across multiple jurisdictions.
- Those working on human dignity focused on care and education contexts, raising concerns about agency displacement – where AI systems reduce rather than support human decision-making – and the risk of treating dignity as a secondary consideration once a system is already deployed.

The activity highlighted the point that responsible AI governance is not a checklist exercise; rather it requires contextual judgement, genuine stakeholder engagement and the willingness to surface uncomfortable findings early, before systems are built.

Session two

Efficient and Sustainable AI

Session chair presentations

1. **Lowering barriers to rich biological data with small, task-specific AI**

   Ivana Mikić, Diffine and Sarpeda, US

2. **Efficiency by design: building scalable AI systems to power General Motors' autonomous vehicles**

   Marija Mikić, General Motors, US

3. **AI on (for) Chips**

   Professor Themis Prodromakis, University of Edinburgh, UK

Through three different examples, this session made the case that the dominant model of AI development is neither inevitable, nor efficient. From task-specific AI models, trained on as little as a single image, that can produce tenfold improvements in data quality from standard laboratory microscopes, to an autonomous vehicle system built around strict compute and energy constraints, to neuromorphic chips that could deliver frontier-level compute from a device held in the palm of a hand, the presentations showed what AI development looks like when efficiency is treated as a core value rather than a secondary concern. A group activity then asked participants to think through the implications of dramatic improvements in scale, speed, sustainability, social responsibility and smart connectivity – the session's five organising themes.

Key takeaways

- Small, task-specific models can match or exceed the performance of large models for defined problems, at a fraction of the cost, hardware demands and energy. The default assumption that bigger is better doesn't hold across most real-world applications.
- Efficiency is a design constraint. The organisations getting this right are those that treat it as a first-order requirement, whether they are building microscopy tools or autonomous vehicles.
- The energy and infrastructure demands of current AI development are both environmentally unsustainable and structurally exclusionary. GPU manufacturing concentrated in two countries locks most of the world out of model development entirely.
- Neuromorphic computing offers a credible longer-term alternative to conventional chip architectures, with demonstrated energy efficiency gains of up to 1,000 times. Foundries are beginning to integrate these technologies.
- Being strategic about where AI is and isn't deployed is as important a design choice as any technical decision.

Lowering barriers to rich biological data with small, task-specific AI

Ivana Mikić

Ivana Mikić presented work at the intersection of AI and biology, showing how targeted, small, task-specific models can extract significantly richer biological content from data produced by standard laboratory microscopes – without the specialist hardware and workflows that typically cost hundreds of thousands of dollars. These models also reduce energy consumption significantly, making the approach viable for research settings with limited infrastructure. She also spoke to the broader opportunity in drug discovery. Here, AI is optimising specific steps in the pipeline – including work on molecules that cross the blood-brain barrier – with drugs now in clinical trials as a result, though these are partial rather than end-to-end AI discoveries. She argued this is a systems problem as much as a modelling one: the tools exist, but the data pipelines, laboratory workflows and institutional incentives that would allow them to be used at scale largely don't. "Capabilities of modern AI systems far outpace most labs' ability to produce sufficient quantities of high-quality, relevant data."

Efficiency by design: building scalable AI systems to power General Motors' autonomous vehicles

Marija Mikić

Marija Mikić described how General Motors is building the foundations for autonomous vehicles under conditions that make efficiency an existential principle. With over 15 sensors per vehicle generating continuous high-resolution data, strict latency requirements, limited onboard compute, and the cost of offloading data over cellular networks, every design decision must account for the full lifecycle from cloud training to on-vehicle inference. The talk was a useful corrective to the gap between AI's public narrative and its engineering reality. General autonomy, the ability to drive door-to-door under all conditions, remains genuinely hard, and the path there requires tight integration and cross-disciplinary collaboration across AI researchers, hardware engineers, and product teams. GM's planned launch of eyes-off highway autonomy in 2028 represents a near-term horizon for a technology that has been in iterative development and real-world testing for years, and points to the distance between a working prototype and a deployable, safe, scalable system.

"The path to global, scalable autonomy is not just about smarter AI – it's about efficient AI."

AI on (for) Chips

Professor Themis Prodromakis

Professor Prodromakis opened by assessing where AI hardware currently stands. Training a model at the scale of GPT-4 requires enough energy to power Edinburgh for two days, and data handling now accounts for over 20% of global energy demand. His group at the University of Edinburgh is developing neuromorphic computing, brain-inspired chip architectures using analogue rather than digital processing, as a credible alternative. Demonstrated efficiency gains of up to 1,000 times over conventional systems would, if realised at scale, allow the compute capacity of the world's largest supercomputer to fit into 5% of its current footprint. Foundries including Taiwan Semiconductor Manufacturing Company (TSMC) and Samsung are beginning to integrate these technologies. The implications for access are significant: edge devices running AI locally, without reliance on cloud infrastructure, could shift who gets to build and use AI in fundamental ways.

"The right tool for the right job – we need to seriously scrutinise what that means."

Group activity

After the three presentations, participants organised into five groups, each taking one of the session's organising themes – sustainability, scale, speed, social responsibility and smart everything – and working through the challenges and opportunities that would arise if the capabilities Professor Prodromakis described were realised.

Each group then presented their conclusions in the role of a minister making policy recommendations.

The exercise produced a range of findings across the five themes:

- The 'sustainability' group identified the opportunity to offset energy consumption from large technology companies toward smaller countries and called for regulation that supports community-led rather than corporate-dominated AI development.
- The 'scale' group raised safety and security concerns alongside the opportunity: more compute power means more potential for misuse, and greater imbalance between those with access and those without.
- The 'speed' group noted that faster AI creates faster decision cycles, with benefits for engineering simulation, disaster prediction and drug discovery, but risks in financial trading and automated military applications where human oversight may not keep pace.
- The 'social' responsibility group stressed the importance of learning from past technology transitions. This included bringing women and underrepresented groups in earlier in the development process, and ensuring that infrastructure changes don't simply replicate existing inequalities at greater speed.
- The 'smart everything' group asked a question that cut across the other conversations: what isn't smart now that actually needs to be? The risk of over-engineering (deploying AI where simpler solutions would serve better) was seen as a genuine danger, alongside the risk of under-deployment.

The activity reinforced a point that ran through all three presentations: the question of how to use AI well cannot be separated from the question of who gets to decide how it is deployed, and on what basis.

Session three

Education in the Age of AI

Session chair presentations

1. **AI as a means and challenge to effective, accessible and empowered learning**

   Professor Tim Coughlan, Institute of Educational Technology, The Open University, UK

2. **Preparing people, not just models: why responsible AI in education must start with systems**

   Ojoma Ochai, Co-creation Hub (CcHUB) Africa, Nigeria

3. **Designing AI for learning when values conflict: from cost-performance to self-efficacy and agency**

   Professor Tetsunari Inamura, Tamagawa University, Japan

This session examined AI in education from three angles: the opportunity it presents for accessible and inclusive learning, the systemic conditions required for responsible deployment, and the value tensions that arise when different stakeholders define quality differently. Together, the presentations resisted easy optimism, acknowledging real gains while taking seriously the risks of dependency, deskilling and the displacement of the human relationships that make learning work. A group activity asked participants to compare what good AI-enabled education looks like from the perspectives of students, teachers and institutions – and to sit with the discomfort of finding no single answer.

Key takeaways

- AI can deliver real learning gains in large, under-resourced classrooms, but only when the surrounding system is ready to absorb it. Technology alone doesn't change outcomes.
- Cognitive offloading is a documented risk: students who outsource thinking to AI rather than developing their own capabilities show measurable declines in independent performance.
- There is no shared definition of what good AI in education looks like across different contexts, stages of learning or cultural settings. This gap is widening as deployment accelerates.
- Personalisation at scale, where content and pace is matched to individual learners in classrooms of 40 or 50 students, is one of the strongest evidenced use cases for AI in education, particularly in low-resource settings.
- What counts as quality in an AI-enabled learning system depends on who you ask. Students, teachers, institutions and societies have legitimately different and sometimes conflicting answers.

AI as a means and challenge to effective, accessible and empowered learning

Professor Tim Coughlan

Professor Coughlan drew on the Open University's experience teaching thousands of distributed students to examine what AI makes possible and what it puts at risk. He presented two projects. One was Taylor, a conversational AI assistant co-designed with disabled students to reduce the administrative burden of accessing support. The second was Ada, a platform that allows course teams to define the roles AI can play in student learning, from Socratic tutor to resource recommender, giving educators meaningful control over how AI shapes the experience. His research found that students value AI tools they perceive as trustworthy and institutionally grounded, and that some students are more willing to disclose sensitive information to an AI interface than to a human advisor. He was direct about the risks: generative AI tools can produce real learning gains, but they also enable cognitive offloading (students bypassing the effortful thinking that consolidates learning) and the evidence on long-term outcomes remains thin.

"AI clearly has the potential to play roles in our learning. But important questions remain about whether we are using it to learn, or just to do things for us."

Preparing people, not just models: why responsible AI in education must start with systems

Ojoma Ochai

Ojoma Ochai brought a practitioner's perspective from CcHUB's work as an African innovation ecosystem enabler, where AI in education is both a sector-specific challenge and a foundational layer of learning infrastructure across health, creative industries and entrepreneurship. She presented deployments in Nigeria and Kenya showing that AI tools in classrooms with student-to-teacher ratios of 40 or 50 to one can deliver personalisation that is otherwise simply impossible. A lesson planning tool deployed in Nigeria allows teachers managing multiple science classes to generate and adapt materials in a fraction of the time. A system called Every Terms, used in Ogun State, tracks enrolment, attendance and learning outcomes at a scale no manual system could sustain. The central challenge isn't whether AI works in a classroom, but whether the institutions, incentives and norms around it are ready. Where they are not, responsible use fails regardless of the quality of the tool.

"The argument for students is personalisation. Even when they have to share devices, you can still group people based on ability in a way that has just never happened in the history of these children's schooling."

Designing AI for learning when values conflict: from cost-performance to self-efficacy and agency

Professor Tetsunari Inamura

Professor Inamura challenged the assumption that AI quality in education can be measured by a single metric. Drawing on research in human-robot interaction and educational AI, he proposed a perspective-aware evaluation framework that asks what good AI-enabled learning looks like from four different standpoints: the learner, the teacher, the institution and society.

These perspectives legitimately conflict. What a student experiences as good support, such as autonomy, confidence, and immediate feedback, may not align with what a teacher prioritises, such as fairness, accountability and workload. Institutions optimising for scalability, measurable outcomes, and cost-effectiveness may not serve either. And what society needs from education – critical thinking, civic participation, and adaptability – may be undermined by over-reliance on AI tools that do the hard cognitive work on behalf of learners.

Professor Inamura's research on the gap between AI behaviour and user behaviour in educational settings found that the closer an AI system's responses are to a learner's own level – slightly better, but not dramatically so – the more it supports self-efficacy. Systems that are too capable relative to the learner can undermine confidence rather than build it.

"Quality in AI education is a value judgement made by multiple stakeholders, and these values can legitimately conflict."

Group activity

Following the three presentations, participants worked in groups guided by prompts on quality in AI-enabled learning.

They approached the question from the perspective of a student, teacher or school system, before reconvening to compare findings.

The discussion surfaced several recurring tensions:

- Students valued tools that reduced embarrassment and increased confidence, particularly when asking questions they would hesitate to raise in front of peers or teachers. But they also valued the sense that someone was paying attention to them as individuals: something AI could simulate, but not genuinely provide.
- Teachers identified time savings as the most immediate benefit, particularly for lesson planning, assessment design and administrative tasks. Their concern was not replacement but trust: if students knew teachers are using AI to generate materials while telling them not to use it themselves, the relationship would break down.
- From the school system perspective, data and insight were the clearest gains. This included the ability to track learning trajectories, identify struggling students early and allocate resources more effectively. The risk, participants noted, is optimising for what is measurable rather than what matters.

Across all three perspectives, one question kept returning: what happens to the skills and relationships that AI makes unnecessary? The group did not resolve it, but the conversation made clear that it is the right question to be asking.

External activity

Reception at the British Embassy Tokyo

On the evening of the second day, participants were hosted by **Julia Longbottom CMG, British Ambassador to Japan** at the British Embassy Tokyo for a reception that brought together the symposium community with representatives from the UK and Japanese AI ecosystems, including the Japan AI Safety Institute.

Longbottom welcomed participants and reflected on the significance of the UK-Japan relationship in AI, and that collaboration at this scale is essential precisely because no single nation can meet the challenge alone. “Japan is a priority partner for the UK,” she said, “with highly complementary strengths.”

Akiko Murakami, Executive Director of the Japan AI Safety Institute reflected on the importance of cross-cultural dialogue in AI safety, noting that what counts as safe looks different from different perspectives, and that forums like this one are where those differences can be worked through constructively.

As part of the external activity, Frontiers participants were hosted by the Miraikan Museum, where they listened to a keynote address from Dr. Hironobu Takagi, Executive Director. They also toured the museum's permanent exhibitions, Explore the Frontiers and Create Your Future, which showcased a wide range of robots, AI innovations, and quantum technologies.

Professor Satoshi Kurihara, Keio University, Japan

Professor Kurihara, President of the Japanese Society for Artificial Intelligence and Director of the Centre for Advanced Research on Human-AI Symbiosis at Keio University, opened the symposium's first full day with a provocation. Efficiency, he argued, is not the same as innovation, and AI is currently being used almost entirely for the former.

Drawing on the history of AI development across three waves – from early rule-based systems through expert systems to the current deep-learning era – he traced how each wave failed not because the ideas were wrong but because the infrastructure was not ready. He noted that the current wave is different because it's driven by the industry, not researchers; and its pace is shaped by commercial rather than scientific logic.

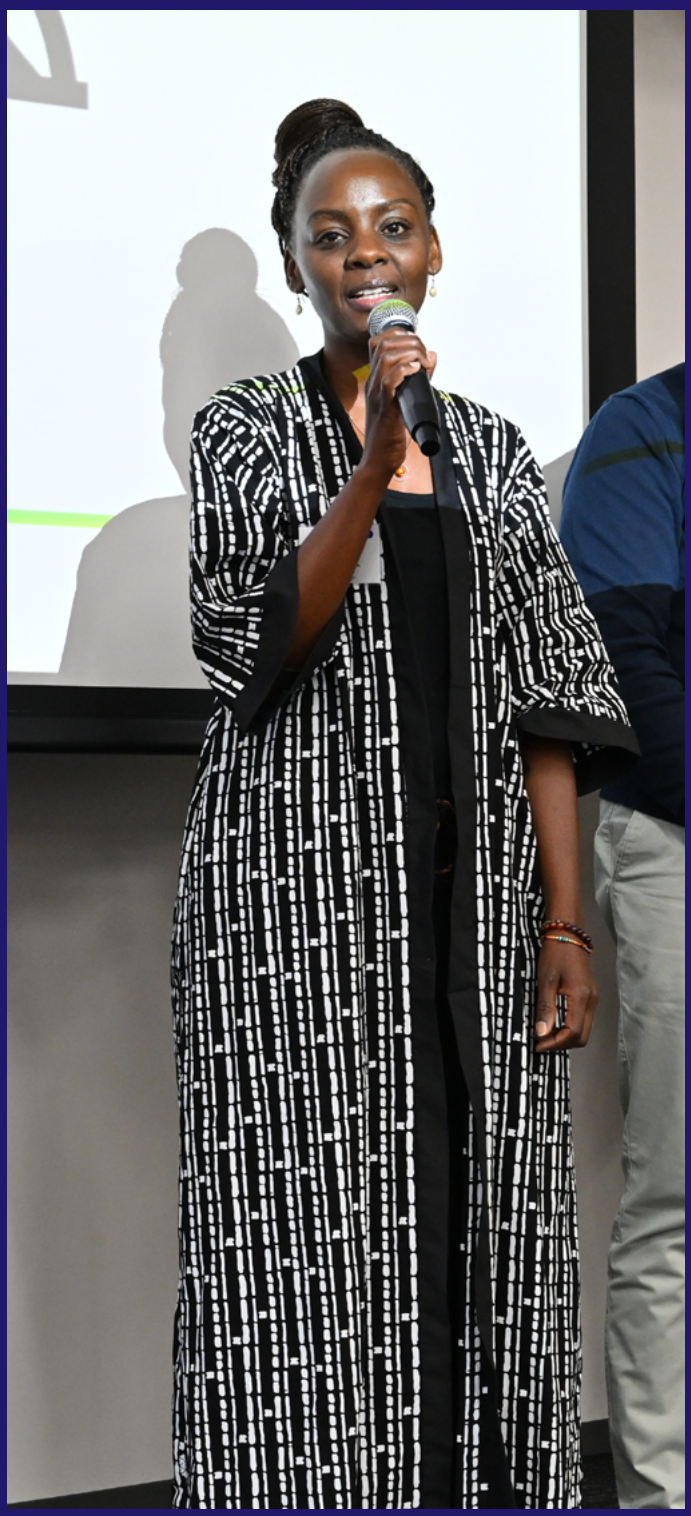
His central concern was that AI's extraordinary capacity for efficiency is being mistaken for a capacity for genuine innovation. True innovation – the kind that produces Nobel Prize-level discoveries – requires making unexpected connections across boundaries of knowledge, and that remains a distinctly human capability, for now. The next generation of AI, he argued, will need to go beyond responding to commands and begin to anticipate human intent, developing what he described as a sense of self-awareness.

"We will feel a sense of self-awareness in AI that is highly autonomous and general, and we must take the challenge. It's a chance."

He closed with a direct challenge to the room: researchers outside the US and China have a genuine opportunity to take a different approach - one grounded in human-AI symbiosis rather than the race to scale. "Can you take the first step?"

Wakanyi Macharia-Hoffman, Inclusive AI Lab, Utrecht University, Netherlands

Wakanyi Macharia-Hoffman closed the symposium with a keynote that asked a question few AI events make time for: what does it actually mean to be human, and what does that have to do with the tools we build?



Drawing on the Ubuntu philosophy – the African conception of personhood as relational, expressed in the idea that a person is a person through other people – she argued that most current AI development is built on a narrow and culturally specific model of what human intelligence and human flourishing look like. The normative assumptions embedded in systems built in Silicon Valley reflect a particular way of seeing the world, and that way of seeing is not universal.

AI, she said, is a non-specific amplifier. Used well, it can help address the Sustainable Development Goals; used carelessly, it can replicate and scale the harms that already exist. The difference lies in the values that shape it, the communities involved in building it, and the questions asked before it is deployed.

Her practical proposal was grounded in her own work. The African Folktales Project, which she founded, uses storytelling-based curricula to bring Indigenous knowledge into schools across Africa, reaching teachers through a platform of 21,000 subscribers. She invited engineers in the room to help build a small AI system capable of amplifying that knowledge across the continent's 2,000-plus languages – not as a grand technological solution, but as a specific, community-grounded tool for a specific, community-defined purpose.

She closed the symposium by asking participants to put their hands on their hearts, close their eyes, and feel their own heartbeat. This was a reminder, she said, of something AI doesn't have and never will. "Let's be real," she told them. "Let's be human together."

Insights Session

At the close of the symposium, the event co-chairs led a discussion to draw together the key threads from three days of sessions, activities and conversations.

The session was designed as an open exchange – participants reflecting on what had shifted in their thinking, what remained unresolved, and what they wanted to carry forward.

Several themes surfaced consistently across the discussion:

From principles to practice

The gap between responsible AI as a set of values and responsible AI as a set of implemented systems was the thread that ran most persistently through all three days. Professor David Leslie framed it directly: the world, he said, is being subjected to the largest uncontrolled human experiment in the history of civilisation, without public consent or democratic governance, and the frameworks being developed to address this are arriving after the fact. The challenge for the next phase according to Professor Leslie is to build the sector-specific, culturally grounded implementation tools that translate principles into engineering practice.

Small AI is not just a technical choice

Decisions about data, computation and energy are also decisions about access, equity and sustainability. The symposium consistently pushed back against the assumption that frontier models are the appropriate default for most applications. As a participant put it, more than 90% of useful AI applications don't require large language models – and treating them as the baseline is both wasteful and misleading about what AI can and cannot do. The organisations and governments getting this right are those matching the scale of the tool to the scale of the problem.

Education as the long game

The education session's most central insight was that AI changes the relationship between teacher and student, student and knowledge, and institution and society – and that these changes are not yet well understood. The risk of cognitive offloading, the absence of shared standards for what good AI-enabled learning looks like, and the concentration of AI development capacity in a small number of companies, all point to a field that is moving faster than the evidence base can support. As one participant observed, education matters more than almost anything else – but it takes a hundred years to see the results.

The AI you cannot see

Several participants noted that the fixation on large language models may be drawing attention away from a more pervasive problem: the AI systems that are already shaping behaviour at scale through recommendation algorithms, automated decision-making and predictive tools, without the level of public scrutiny that generative AI has attracted. "The AI that we don't see," one participant said, "is more sinister, because you actually don't know that it's influencing your thinking and your behaviour."

Who holds the power

The session closed with a question about agency. Several participants expressed a sense that AI's trajectory feels predetermined – shaped by forces too large and too fast to influence. Others pushed back: users, researchers, policymakers and engineers all hold more power than they typically exercise, and the moments when that power has been used collectively – as when public pressure forced a major platform to reverse a decision about AI content moderation – show what is possible. "We need to fight for the AI that we want," one participant said. "We need to figure out the ways to take back the power."

Royal Academy of Engineering

All images used with permission from participants unless otherwise stated.

For more information, including eligibility, please visit raeng.org.uk/frontiers, www.jst.go.jp, and follow in @the-royal-academy-of-engineering

---

Royal Academy of Engineering
Prince Philip House, 3 Carlton House Terrace, London SW1Y 5DG

Ale (Unsplash)