GSIS® HAMBURG
9-10 December 2026
Global Security and Innovation Summit

organised by Hamburg Messe + Congress IISS

Monthly Executive Briefing

The AI ecosystem: managing fragility and building resilience

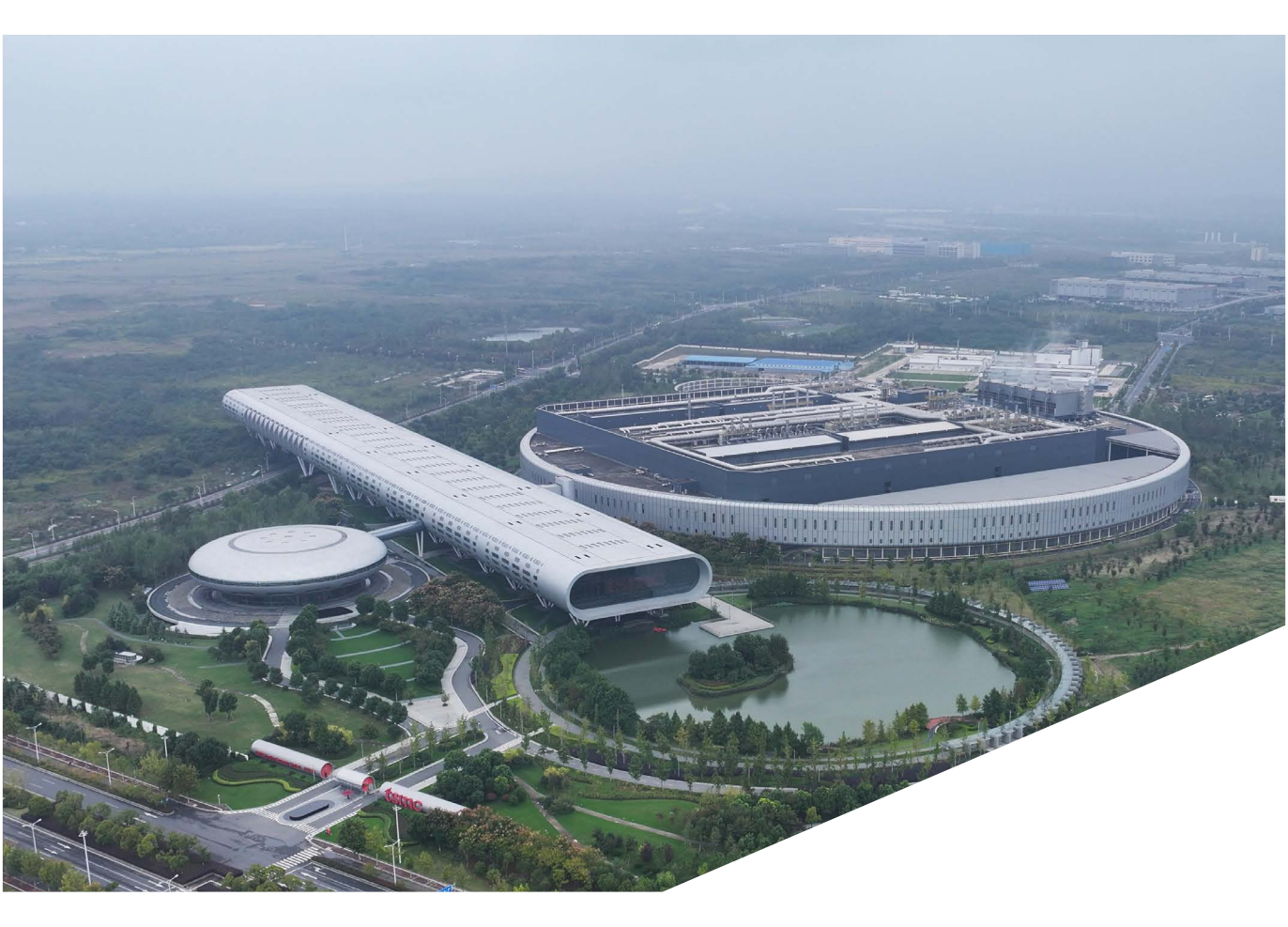
Artificial intelligence (AI) is often seen as just software, but its foundations are physical, global and fragile. From advanced semiconductors and hyperscale data centres to energy, water and talent, the AI ecosystem depends on scarce and interlinked resources. As governments and firms race to harness its power, the question is no longer just how to innovate, but how to build resilience across a system under growing strain.

- Because of the fragilities visible across all its layers – from chips and cloud infrastructure to data, algorithms, software and recruitment – AI must be managed as an ecosystem, not a stand-alone technology. Disruptions or shortages in one layer can cascade through an entire system.
- The risks vary: frontier labs face supply, regulatory, infrastructure and reputational shocks, while downstream firms remain vulnerable to cascading disruption. Hidden costs in energy, water and carbon emissions are mounting, as are the costs of recruitment and retraining.
- The key to success is to build resilience, invest in efficiency and embed responsibility, enabling states to gain capability advantage, long-term resilience and a reputation for trustworthy, responsible AI leadership.

GSIS® HAMBURG
9-10 December 2026
Global Security and Innovation Summit

GSIS 2026 News:

- *Take a moment to review the Closing Report for the full GSIS 2025 story.*
- *Experience the highlights of GSIS 2025 in our recap video*
- *Exhibitor and partnership packages for GSIS 2026 are now available. Connect with your partnership contact to find the best fit for you.*



>60%
COMBINED MARKET SHARE FOR AMAZON, MICROSOFT AND GOOGLE IN GLOBAL CLOUD INFRASTRUCTURE

4
THE DOMINANT COMPANIES IN CRITICAL AI HARDWARE: NVIDIA, TSMC, ASML AND BROADCOM

1.5%
DATA CENTRES' ESTIMATED SHARE OF GLOBAL ELECTRICITY CONSUMPTION, 2024

AI's fragile foundations

Artificial intelligence (AI) is best understood not as a stand-alone technology but as a global ecosystem built on interdependent layers. Control rests in the physical infrastructure where critical inputs such as computing hardware are scarce, and sourced from a tightly concentrated supply chain. This concentration shapes the capabilities of AI models while its effects on downstream applications are more diffuse as diversity expands further along the chain. This analysis focuses on frontier (including generative), large-scale AI models whose training and deployment depend on global computing, cloud and data infrastructures, where fragilities and hidden costs are most pronounced. With AI now at the centre of geopolitical competition and industrial policy – from the US–China contest over advanced semiconductor chips to the UK's Bletchley Declaration – the key question is how governments and industry can build resilience across an increasingly fragile AI ecosystem.

At the foundations of the AI ecosystem are the design of graphics processing units (GPUs) and the fabrication of advanced semiconductors, which are concentrated amongst only a few companies: US-based NVIDIA and Broadcom, Taiwan-based TSMC and the Dutch multinational corporation ASML. This concentration creates chokepoints vulnerable to export controls, tariffs or technical delays. The 'cloud', comprising vast interconnected data centres, is also central to the AI ecosystem. Three US-based hyperscalers – Amazon, Microsoft and Google – together account for more than 60% of cloud infrastructure worldwide, creating both vendor lock-in and geopolitical exposure. The severity of these supply-chain vulnerabilities varies: the US dominates cloud services but relies on global supply chains for semiconductors and the critical materials needed to produce them; Europe depends heavily on foreign providers across both layers; and China, while constrained by export controls on advanced semiconductors, retains a strategic advantage in upstream inputs such as rare-earth minerals and component manufacturing. That said, these players depend on the same complex, globally distributed supply chains that underpin AI development and deployment.

Data access is becoming increasingly uneven. The amount of data produced worldwide continues to grow rapidly, but access to the high-quality, large-scale datasets needed to train advanced AI models is narrowing. Large, vertically integrated firms such as Amazon and Google generate, collect and control vast proprietary datasets through their search, retail and cloud platforms, giving them a structural advantage over smaller players. Smaller developers often cannot match this access, and the ongoing tightening of data-protection and sovereignty rules further restricts what can be collected, shared or exported. Copyright disputes such as Anthropic's US$1.5 billion settlement with authors, privacy restrictions such as Italy's temporary ban on ChatGPT, and national data-sovereignty regimes in China (Personal Information Protection Law–PIPL; Data Security Law–DSL; Cybersecurity Law–CSL), the EU (General Data Protection Regulation–GDPR; AI Act) and India (Digital Personal Data Protection Act–DPDP) illustrate how the space for freely gathering and using data is shrinking.

Model development is also highly concentrated. A handful of frontier AI labs such as OpenAI, Anthropic, Google DeepMind and China's DeepSeek dominate the development of the most capable generative models. While these firms currently lead in scale and resources, the question of who leads truly cutting-edge research remains open, as new entrants and national research programmes compete to close the gap.

Most organisations will not build their own models but will use AI through commercial applications built by these frontier firms. However, these downstream systems depend on the same hardware, cloud infrastructure and data pipelines, so when problems arise at the top of the stack, their effects are inherited by every user. Costs and friction are already arising as energy demand strains grids, water use sparks local opposition, carbon emissions draw regulatory scrutiny, and talent shortages slow deployment. The lesson is clear: frontier AI is not plug-and-play; it depends on an ecosystem that must be managed with the same resilience, efficiency and governance as any other strategic infrastructure.

The real costs of AI

Debates about electric-manufacturing capacity – especially in Europe – often treat it as a matter of
The growth of AI depends on a vast physical infrastructure. In 2024, data centres worldwide consumed around 415 terawatt-hours of electricity, about 1.5% of total global consumption, which represents an average rise of 13% annually since 2020. Efficiency gains have tempered growth in power use – for example, despite a 550% increase in US computing workloads from 2010–18, US electricity consumption rose by only 6%. However, overall demand continues to surge. As computing becomes more efficient and accessible, it also becomes cheaper and more widely used, driving up total energy consumption.

In response, countries such as China, Singapore and the Netherlands have introduced measures to balance economic growth with environmental limits on data-centre expansion, including temporary construction moratoriums and Power Usage Effectiveness (PUE) requirements to improve energy efficiency. Hyperscale AI facilities are generally more efficient than those of smaller enterprises, yet a single hyperscale facility can have a capacity of 100 megawatts or more, consuming as much power annually as 100,000 households. Chip manufacturing adds further strain as each new generation of GPUs and server-grade semiconductors – smaller, faster and more complex than the last – requires more energy per wafer to produce. Training frontier systems can consume hundreds of gigawatt-hours of electricity, while inference costs, though variable, continue to add substantially to overall energy demand.

Water is a parallel concern, and here too efficiency does not automatically mean a lower impact. Roughly three-quarters of a data centre's water footprint is indirect, coming from its electricity consumption, but direct use for cooling can still place real strain on local resources, particularly in regions already facing scarcity. As data centres expand globally, their water usage adds up even if that of individual centres is modest, and efficiency improvements can encourage further growth overall, mirroring the effect seen with energy. Local conditions matter, as volumetric totals can be misleading if scarcity and competing community needs are not taken into account. This has already sparked community opposition and regulatory pressure in water-stressed areas in Chile, Uruguay and the United States, prompting firms to redesign facilities or relocate projects. Shifting to 100% renewable energy could reduce overall water footprints by more than half, but such transitions take investment and time. Carbon emissions mirror these dynamics, with estimates that generative AI could add tens to hundreds of millions of tonnes of CO_2 annually by 2035 if left unmanaged.

Scarcity of the required skills compounds these pressures, especially for firms developing advanced AI systems. Elite machine-learning researchers can now command multi-million-dollar packages, in some cases receiving signing-on bonuses exceeding US$100 million as companies compete for a limited pool of expertise. While few firms disclose full research and development (R&D) spending, public estimates suggest that building frontier-scale models can cost tens to hundreds of millions of dollars, even though recent entrant DeepSeek claims to be able to train competitive models for under US$300,000, a figure that is contested. For firms that use rather than build AI, the challenge is different but no less significant – training staff, integrating systems and ensuring their effective use are all processes that demand new investment and expertise.

Beyond the need for AI-related skills, AI is shaping labour markets in other ways. The UK estimates that 10–30% of current jobs are automatable with AI, with professional occupations in finance, law, business management and education the most exposed. However, other studies differ on the sectors most likely to be affected and the probable level of disruption across the economy as a whole. Entry-level roles are declining in AI-user sectors, even as new opportunities emerge in developing and deploying frontier systems. These pressures will only become more visible. In addition, adopting AI effectively requires an AI-literate workforce capable of using tools safely and productively, without leaking sensitive data or misapplying outputs, and AI-proficient engineers who can tailor models, build integrations and rework back-office processes to capture real productivity gains. These skills are increasingly prerequisites for competitiveness and come with their own business costs.

For any country hoping to prosper in the AI era, all types of skills matter, including those of elite researchers, proficient engineers who integrate and maintain the systems, and data-literate workers who use AI effectively and safely. Each underpins a different layer of the ecosystem, from frontier research to responsible deployment to broad workforce readiness. As with energy and water, the increasing demand for talent will become more evident, and for consumers it will manifest in higher subscription fees, premium pricing for advanced features and slower rollout of new capabilities as companies absorb or pass on the rising costs of remunerating highly skilled employees.

"We are crossing humanity's most profound technological frontier. Our lives, our world, our politics are about to be flooded with super-powerful AI. There is only one way forward. Resilience."

David Lammy, Deputy Prime Minister, United Kingdom, September 2025

"You're not going to lose your job to an AI, but you're going to lose your job to someone who uses AI."

Jensen Huang, CEO and co-founder of NVIDIA, May 2025

"AI will be extremely energy-intensive. We are really not taking that into account."

Jonas Gahr Støre, Prime Minister, Norway, September 2024

When pressures collide

The fragility of the AI ecosystem is most visible when supply-chain chokepoints impact on environmental or operational constraints. Hardware shortages have affected projects worldwide: in May 2024, a South Korean supplier to NVIDIA, SK Hynix, announced that almost all the high-bandwidth memory chips it could produce in 2024 and 2025 had already been prepurchased. In February 2025, OpenAI delayed the rollout of GPT-4.5 to ChatGPT Plus users due to GPU scarcity.

The strain on cloud infrastructure is another pressure point. In Ireland, where data centres consumed over 20% of national electricity in 2024, community opposition has intensified, forcing the government to weigh the economic gains of being Europe's leading hub against environmental limits. Similar pushback, over water use, has derailed Google projects in Uruguay and Chile, while in Malaysia nearly one-third of applications to establish data centres in the state of Johor have been rejected for failing to meet sustainability standards. To manage these pressures, firms such as Microsoft are now building or financing their own energy-generation capacity, including on-site gas plants and agreements to secure zero-emissions green-hydrogen power for one of their European data centres. While this may ease grid strain in the short term, it raises broader questions about who controls critical energy infrastructure, how emissions from private generation are accounted for, and whether national regulators can keep pace with the scale and speed of AI-driven demand.

Another problem is data access. The EU's AI Act now requires model providers to disclose sufficiently detailed summaries of their data, while regulators in Singapore, India and China are, for varying reasons, tightening rules on personal-data use and cross-border data transfers. These measures primarily affect access to large-scale datasets for model training, where restrictions on source material and data movement can slow or fragment development. Local firms that use customer-feedback loops to fine-tune or adapt existing models within national boundaries may be less affected, but they still face higher compliance costs and reduced flexibility in sourcing or sharing data internationally. The overall effects are rising costs, operational fragmentation and uneven access to AI capabilities across regions.

Security risks compound these vulnerabilities. Recent research has shown that even trusted enterprise systems can be exposed to prompt-injection weaknesses (i.e., where hidden instructions mislead the model), with demonstrations revealing how Slack AI and OpenAI connectors could be manipulated to leak private data. While there have been no confirmed large-scale prompt-injection attacks, cyber criminals are increasingly using uncensored or jailbroken (i.e., modified to bypass built-in safety and content restrictions) large language models, such as WormGPT and Mixtral, to develop and distribute malicious code. These trends highlight how vulnerabilities across the ecosystem – from insecure integrations to misuse of open models – can expose firms to legal, operational and reputational risks, while placing consumers at risk of privacy breaches or manipulated outputs. Without proactive security design, continuous testing and robust governance across the infrastructure, data and application layers, there is a risk that AI ambitions will be slowed or fragmented by cascading vulnerabilities. Emerging efforts such as the Bletchley Declaration and the UK's Safe AI initiative mark early attempts to build shared standards for model safety, security and transparency, but they are works in progress and largely voluntary. Embedding these principles across the global AI ecosystem will be essential in order to move from mere aspiration to enforcement.

From fragility to resilience

Because of the fragilities visible across all its layers – from chips and cloud infrastructure to data, algorithms and software – AI must be managed as a global ecosystem, not a stand-alone technology. Hidden costs in energy, water and carbon emissions are mounting and increasingly visible, as are the costs of recruitment and retraining. Diversifying across vendors and jurisdictions can help reduce exposure to hardware, software and data chokepoints, but will not solve deeper resource and workforce pressures. Addressing those will require parallel efforts to improve energy efficiency, invest in sustainable infrastructure and build AI-ready workforces through stronger skills development and retention.

Risk exposure is uneven. Frontier labs developing the world's most advanced AI models face acute vulnerability to hardware shortages, regulatory shocks and reputational scrutiny, while governments, AI-using firms and downstream service providers are more shielded but still partially exposed to cascading disruption.

The employment implications make foresight essential. In the US, since the widespread adoption of generative AI, the number of jobs for early-career workers (aged 22–25) in the most AI-exposed occupations has decreased while employment for their counterparts in other sectors has expanded. Growth in tech industries such as cloud, web search and computer-systems design stalled at the end of 2022, highlighting early signs of disruption. The balance between displacement and augmentation will shape national labour markets, making training and reskilling critical to resilience.

Whether building AI or deploying it, firms and governments that treat it as an ecosystem by securing resilience, improving efficiency, embedding responsibility and fostering an AI-ready workforce will gain not only competitive strength but also long-term resilience and strategic advantage. Achieving this requires embedding responsibility through mechanisms for trust, explainability, validation, verification and security.

Our Partners 2025

AIRBUS – Founding Partner
Microsoft – Founding Partner
Peregrine – Alliance Partner
Special Competitive Studies Project – Alliance Partner
eraneos – Alliance Partner
EDGE – Official Partner
SPEER GROUP – Venture Partner

YOUR CONTACT

Hamburg Messe und Congress GmbH · Messeplatz 1 · 20357 Hamburg · Germany
Phone +49 40 3569-0 · Fax +49 40 3569-2203 · info@gsis-hamburg.com · www.gsis-hamburg.com

Follow us:

Management Board: Uwe Fischer (CEO), Heiko M. Stutzinger (CEO)
Hamburg Municipal Court, Reg. No. HRB 12 054
VAT No. DE811214125

Hamburg Messe + Congress