

Building Resilient and Scalable AI Value Chains: A Nexus Strategy

INSIGHT REPORT
MAY 2026



Contents

Foreword	3
Executive summary	4
Introduction	5
1 AI today and the value chain forces steering its future	7
1.1 AI infrastructure	8
1.2 AI factory and data centre operations	9
1.3 AI applications	9
1.4 AI transformation: A platform shift	11
1.5 Early stage AI application growth amid infrastructure constraints	11
1.6 AI infrastructure and access to capital	12
2 A call to action	15
2.1 Energy: A highly visible choke point	17
2.2 Water: The overlooked AI growth limiter	17
2.3 Minerals: Supply chain and geopolitical vulnerability	19
2.4 Land: Spatial pressure on communities and ecosystems	20
3 What key stakeholders can do	22
Conclusion - The way forward	23
Contributors	25
Endnotes	26

Disclaimer

This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 2026 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Foreword



Sam Hsenghung Hsu
Member, Global Future
Council, World Economic
Forum; Board Member,
Fastenal



Lauren Smart
Member, Global Future
Council, World Economic
Forum; Global Head of
Sustainable Finance,
Bloomberg

Artificial intelligence (AI) is reshaping businesses, communities, nations and the global landscape at a speed and scale that few anticipated. What began as a wave of high-computational exploratory deployments in late 2022 has rapidly evolved into a structural transformation touching every sector of the economy. AI is no longer a discrete technological advancement; it is becoming a foundational capability that will define competitiveness, productivity and societal progress for decades to come.

Yet this transformation is unfolding in systems that were never designed for the magnitude of demand AI now places on them. The accelerating buildout of AI infrastructure, the rise of industrial scale AI factories and the proliferation of AI enabled applications are creating unprecedented pressures on energy grids, water systems, mineral supply chains, land use and community infrastructure. These pressures are converging into a new strategic reality: the AI-energy-water-minerals-land nexus will shape the trajectory of the future of AI as much as any breakthrough in compute capacity.

This insight report is written at a pivotal moment. The world is experiencing the early waves of AI adoption, but far larger waves are forming behind them. The economic signals – historically high capital investment, soaring demand for compute and the rapid ascent of AI driven enterprises – indicate that we are still in the early innings of a multi decade, transformative platform shift. The question is no longer whether AI will scale, but whether the systems that support it can scale sustainably, equitably and resiliently.

Our aim is to provide clarity on this challenge and a framework for action. By examining the full AI value chain – its infrastructure, its operational backbone and its application layer – we highlight the energy, water, minerals and land interdependencies that will determine the long term viability of AI. We also surface the system stressors that must be addressed to ensure that AI becomes a net positive force for innovation, economic growth and human well being.

This report is grounded in a simple belief: the future of AI will be shaped not only by technological breakthroughs, but by the choices leaders make today across industry, government, finance and civil society. Integrated thinking, shared accountability and coordinated action are no longer optional. Instead, they are the prerequisites for building an AI ecosystem capable of meeting global demand without compromising the ecological and social systems on which we all depend.

This report is part of the work of the Global Future Council on Energy Nexus, which aims to equip decision-makers with insights and frameworks for better energy planning. We offer this work as a contribution to that collective effort – an invitation to think more broadly, plan more holistically and act with greater urgency as we enter the next chapter of the AI era.

Executive summary

The AI boom is dependent on energy, water, minerals and land systems. Artificial intelligence (AI) is entering an unprecedented infrastructure boom. Capital spending by major technology firms on AI infrastructure is projected to reach \$700 billion in 2026, compared to \$410 billion in 2025,¹ as companies race to build the computing capacity for advanced AI systems.

The scale and speed of this investment reflect earlier transformative infrastructure booms – from the nineteenth-century railroad expansion to the fibre-optic buildout that enabled the modern internet. Beneath this digital expansion lie critical dependencies. AI runs on electricity, requires water for cooling and power generation, uses land for data centres and transmission corridors, and relies on critical minerals for chips, batteries and electrical systems. These systems – energy, water, minerals and land – form the resource backbone of AI infrastructure. While investment in AI infrastructure is accelerating at unprecedented speed and scale, it is often without full visibility of the pressures emerging across these interconnected domains. Without more integrated planning, AI infrastructure risks getting locked into unsustainable dependencies that could increase systemic risks.

The infrastructure choices made today will determine whether AI becomes constrained by these resource systems – or helps transform them. Coordinated action is needed. The future of AI is yet to be determined: the technologies driving these pressures can also help address them if applied strategically. With intentional design, responsible investment and integrated governance, AI can be a **net-positive force** – a resilience multiplier rather than a system stressor – strengthening food systems, improving human

health, protecting water resources, enhancing quality of life and accelerating sustainable development. To achieve this, coordinated action across governments and industry is key to ensure the following:

- **AI infrastructure is sited and designed within ecological and social limits**, accounting for long-term water, energy, land and climate constraints.
- **Investment and incentives prioritize high-impact AI applications** that deliver measurable social and environmental benefits.
- **Cooling, energy and minerals sourcing are built for resilience**, incorporating circular practices and ethical sourcing.
- **Internationally recognized AI governance frameworks** are shaped to guide safe, inclusive development that reflects AI's physical needs across energy, water, land, minerals and labour.
- **Companies adopt nexus-aware strategies and metrics**, moving beyond narrow efficiency indicators to system-wide risk management.

The future trajectory of AI will be determined not only by technological breakthroughs but by the ability of stakeholders across the value chain to build synchronized, resilient and sustainable systems. Stakeholders who recognize and act on the AI-energy-water-minerals-land nexus early will shape both competitive advantage and societal outcomes. The choices made now will determine whether AI becomes a catalyst for long term prosperity or a source of compounding systemic stress.



Introduction

The AI value chain reveals where resource pressures are emerging and the emerging choke points for AI growth. This report examines AI development across three interconnected layers and strategic choke points for AI growth.



1 AI infrastructure: The upstream supply chains and systems that enable AI, including semiconductor manufacturing, computing hardware and the critical minerals used to produce them (e.g. metals such as copper for electrical systems; lithium, nickel and cobalt for batteries and energy storage; and rare earth elements for electronics).

2 AI factories and data centre operations: The large computing campuses where AI models are trained and deployed. These facilities require electricity for computing, water for cooling and land for siting campuses.

3 AI applications: The downstream tools and services built on AI systems – such as digital assistants, predictive analytics and automation platforms – that drive demand for computing capacity. As adoption expands across sectors, these applications increase pressure on resources, yet also offer opportunities to improve resource efficiency and system optimization.

Each has different implications for resource requirements, which, when observed separately, expose vulnerability to AI development as a whole.

Emerging choke points for AI growth

As data centres, power generation and supply chains scale simultaneously, demand for energy, water, minerals and land rises sharply. If unmanaged, these pressures could become **strategic choke points for AI growth**, intensifying trade-offs with communities, ecosystems and other industries. Four emerging pressure points illustrate how these constraints are taking shape.

1 Energy: A visible choke point – and a fork in the road for the energy transition

AI is driving a sharp increase in electricity demand. The International Energy Agency estimates global data centre consumption could approach 1,000 TWh by 2030. In the United States (US), data centres account for 40-50% of incremental electricity demand growth this decade.² AI clusters require high-reliability power, intensifying pressure on electricity grids already facing congestion.

2 Water: The overlooked limiter of AI growth

Cooling high-performance computing infrastructure requires significant volumes of freshwater. Global data centre water consumption is projected to reach 450 million gallons per day by 2030,³ equivalent to the daily needs of roughly five million people, and 43% of global data centres already operate in water-stressed regions.

3 Minerals: A supply-chain and geopolitical vulnerability

AI hardware depends on global supply chains for critical minerals used in semiconductors, electrical systems and energy infrastructure. Demand for

copper, lithium and rare earth elements is rising due to AI expansion. Copper demand alone could increase 30-40% by 2035, while lithium demand could grow five- to seven-fold by 2030.⁴ With new mining projects taking more than a decade to develop and supply chains concentrated, potential bottlenecks are exposed.

4 Land: Mounting pressure on communities and ecosystems

The expansion of hyperscale data centres and the energy infrastructure that supports them is increasing demand for land with power, water and network connectivity. Community concerns about pressure on local infrastructure, diverted resource use, limited economic benefit, and noise and pollution from diesel generators are causing projects to stall or be cancelled.

Together, these pressures highlight how the material foundations of AI – energy, water, minerals and land – are strategic constraints on the growth, profitability and sustainability of future AI development.

Part 1 of this report dives deeper into the critical elements for AI development, Part 2 discusses the material needs of AI, and Part 3 explores what key stakeholders can do to work towards scale and system efficiency within planetary boundaries.



1

AI today and the value chain forces steering its future

The world has only experienced the early waves of AI transformation, but the tide is rising fast and a much larger tsunami is on the horizon.

What is the AI value chain?

AI is not new; its foundational disciplines – machine learning, deep learning, natural language processing, computer vision and generative modelling – have been evolving for decades. Yet for most businesses and consumers, the impact of AI on daily life remained limited until recently. The critical inflection point came in late 2022, when ChatGPT introduced a simple, intuitive interface that made the power of large language models accessible to the general public in many countries.

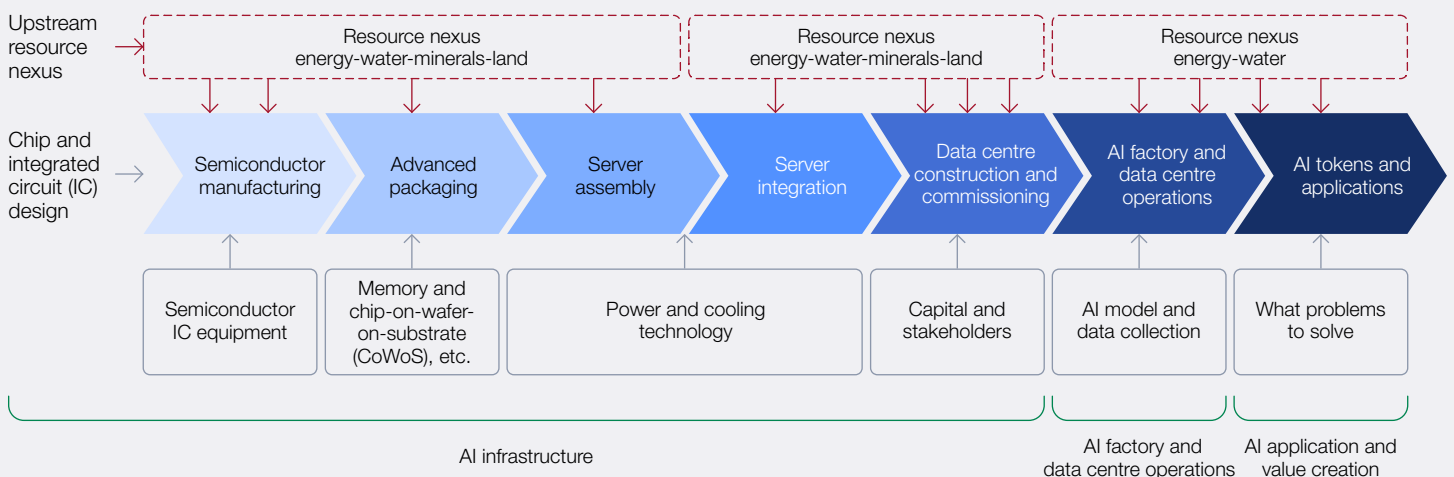
Since then, rapid gains in computational efficiency and scale have accelerated the evolution of AI from a collection of technical tools into waves of value creating applications. Innovation has accelerated across industries, unlocking new possibilities and addressing challenges once considered

out of reach. Adoption is rising quickly and the pace of progress continues to intensify, with new capabilities, use cases and business models emerging almost daily. Whether in the news, on social platforms, in boardrooms or at global conferences, AI now dominates the conversation.

To clearly understand the economic, social, political and environmental impact and significance of this shift, it is helpful to examine the AI value chain, which can be grouped into:

- AI infrastructure
- AI factory and data centre operations
- AI applications

FIGURE 1 The AI value chain





1.1 AI infrastructure

AI infrastructure – the compute, data, energy and resources and tangible systems that power AI – covers the entire upstream stack required to build, scale and operate modern data centres. It spans chip and integrated circuit design, semiconductor equipment and manufacturing, advanced packaging, server assembly and integration, and the construction and commissioning of data centre facilities. It also connects the broader resource ecosystem – power generation and distribution, cooling technologies and water systems, minerals and the land needed to host these facilities.

Three forces define the evolution of AI infrastructure: capability, capacity and efficiency. Architectural priorities are shifting from traditional cloud environments to specialized AI factories built around dense Graphics Processing Unit (GPU) or Application-Specific Integrated Circuit (ASIC) accelerator clusters, high bandwidth networking and next generation cooling. As these systems scale, the strategic importance of AI infrastructure is heightening geopolitical competition and regulatory scrutiny, including export controls on advanced AI chips tied to national security concerns. As this insight report illustrates, it is also heightening awareness of and pressures on the AI-energy-water-minerals-land nexus.

Over the coming years, AI infrastructure will be shaped by explosive growth in high performance compute, widespread adoption of advanced cooling, tighter integration of data pipelines with training environments, and increasingly autonomous operations driven by AI native orchestration. Purpose built AI factories – data centres designed as integrated production systems for training, inference and large scale data processing – are emerging as the next phase of digital infrastructure.

Because AI infrastructure spans such a broad value chain, multiple bottlenecks can arise as demand surges. These include constraints on electricity and water, limited semiconductor and high bandwidth memory capacity, shortages of critical materials, slow innovation in cooling, scarcity of suitable land, inadequate access to capital, local community opposition and security driven regulatory controls. At the same time, sustainability pressures are pushing the sector towards low carbon energy, energy aligned and net water positive designs, and more circular approaches to materials when building AI infrastructure.

Building a resilient future for AI infrastructure will require coordinated action across industry, government and communities to relieve these bottlenecks and enable AI factories that are more powerful, efficient and self optimizing.

1.2 AI factory and data centre operations

AI factory operations have evolved from manual, hardware centric environments into highly automated, software defined ecosystems. Virtualization, cloud computing and containerization laid the groundwork; AI is now pushing operations towards GPU or ASIC optimized clusters, high bandwidth interconnects, energy efficient cooling and real time telemetry. Automation, predictive maintenance and autonomous workload scheduling are now embedded into the operational core.

AI models and data centre operations are becoming deeply interdependent. Large scale foundation models require massive compute, fast data movement and continuous access to accelerators, which directly influence facility design, power architecture and cooling strategy. In parallel, AI is increasingly used to operate these environments through predictive maintenance, energy optimization, workload scheduling, anomaly detection and autonomous control loops.

This creates a reinforcing cycle: more advanced models demand more advanced data centres, which in turn rely on AI to operate at scale.

Data collection now sits at the heart of AI factory operations. High resolution telemetry on power, thermal performance, workloads and hardware health feeds models that optimize behaviour, detect anomalies and forecast resource needs. As AI workloads grow more complex, the quality, granularity and governance of this data become critical to automation effectiveness, operational safety and regulatory compliance.

Critical capabilities for modern data centre or AI factory operations:

- **Operational excellence and reliability:** High uptime, resilient architectures and rigorous change management.
- **Scalable, high performance infrastructure:** Dense accelerators, high throughput (token per second), cost-efficient compute, low latency networking and optimized storage pipelines.
- **Energy and water efficiency and sustainability:** Strong power usage effectiveness (PUE) and water usage effectiveness (WUE) performance, renewable energy integration and advanced cooling technologies.
- **Automation and intelligent orchestration:** AI driven monitoring, self healing systems and automated provisioning to reduce human error and accelerate deployment cycles.
- **Talent and cross functional collaboration:** Skilled operators, data engineers and AI specialists working together to manage increasingly complex environments.

The key challenges for data centre and AI factory operations include:

- **Power and cooling constraints for dense AI clusters:** As GPU and ASIC racks grow hotter and more power intensive, facilities struggle to secure enough electricity and deploy advanced cooling systems fast enough to keep pace with demand.
- **Specialized supply chain constraints:** Chips, optics and power components face long lead times and limited sources.
- **Workload complexity:** Modern AI workloads vary widely in size, duration and resource intensity, requiring sophisticated orchestration to optimize throughput, minimize idle time and maintain reliability.
- **Escalating security and compliance demands:** Protecting data, models and operations grows harder under expanding regulations.
- **Energy-water-land-cost trade-offs:** Hyperscale operators must optimize all four simultaneously as models grow.

In the next few years, AI factory and data centre operations will continue to evolve towards hyper-scaled campuses, greater autonomy, advanced liquid cooling innovation, expansion of edge AI sites, diversification of accelerator hardware, and stronger sustainability mandates such as carbon aware scheduling, greener data centres and circular hardware life cycles.

1.3 AI applications

AI applications are the point where technological capability can translate into tangible economic and strategic value. AI infrastructure and models form the backbone, but real impact emerges when organizations embed AI into workflows, products, decision-making and physical systems. These applications span predictive analytics, intelligent automation, generative content creation, autonomous systems and physical AI – unlocking new levels of productivity, efficiency, personalization and innovation across industries.

There are several evolutionary steps from pure creation to autonomous action and physical embodiment: Predictive AI analyses historical data to estimate outcomes; Generative AI creates content – text, images, code – based on prompts; Agentic AI plans, uses tools and executes multi step

workflows autonomously; and Physical AI extends these capabilities into robots and machines that interact with and act in the real world, with further advancements continuing along this trajectory.

In many enterprises and industries, AI is already delivering measurable gains through operational efficiency and improved decision quality. Organizations are automating routine tasks, optimizing supply chains, forecasting demand, detecting anomalies and augmenting human expertise. Generative AI accelerates content creation, software development and customer engagement, shortening cycle times, reducing costs and enabling employees to focus on higher value work. Early adopters are reporting significant improvements in productivity and revenue as AI becomes embedded in core business processes.

Beyond efficiency, AI is also enabling new business models and revenue streams. Industries such as healthcare, finance, manufacturing and retail are using AI to deliver hyper personalized services, predictive maintenance, real time risk assessment and intelligent product recommendations. Fields like biotechnology and materials science are compressing R&D cycles and achieving breakthroughs that were previously out of reach.

Ultimately, the value of AI applications lies in what they make possible. Organizations that embed AI into their strategy unlock compounding advantages – faster innovation, richer customer value creation, and more adaptive operating models. These gains accelerate over time, widening the distance between leaders and laggards and positioning AI driven value creation as a defining competitive force in the decades ahead.

Current economics of the AI ecosystem – and its potential










Across nearly every economic indicator, the speed and scale of AI’s rise is unmistakable and unprecedented. In this section, the authors highlight the expanding scale and market dominance of leading AI driven companies, the long term transformative potential of AI as a mega platform shift, the implications of surging infrastructure demand and the historically high levels of capital investment fueling this trajectory.

To illustrate the economic scale of the post 2022 AI wave, the authors analysed market capitalization

and stock price performance as proxies for value creation. While market valuations are not the same as actual economic output, they provide a clear signal of investor conviction in the expected impact of AI. By the end of 2025, as shown in Table 1, the world’s nine largest publicly traded companies were all meaningfully engaged in AI related businesses.

While AI participation extends far beyond these leaders, the “Top 9” serve as a practical proxy for AI driven enterprises, with the broader S&P 500 representing overall market performance.

TABLE 1 **Comparative analysis of “Top 9” market cap companies vs. S&P 500** (year-end 2022 to year-end 2025)

Ranking	Company	Market Cap 2025 (\$B)	Market Cap 2022 (\$B)	Δ(\$B)	Stock 2025 (\$)	Stock 2022 (\$)	Δ%
1	 Nvidia	4638	364	4274	186.5	14.6	1177 %
2	 Apple	4057	2066	1991	271.86	128.0	112 %
3	 Alphabet	3802	1145	2657	313.8	88.07	256 %
4	 Microsoft	3625	1787	1838	483.62	234.22	106 %
5	 Amazon	2485	857	1628	230.82	84.0	175 %
6	 Meta	1671	320	1351	660.09	119.51	452 %
7	 Broadcom	1669	234	1435	346.1	53.46	547 %
8	 Tesla	1580	389	1191	449.72	123.18	265 %
9	 TSMC	1570	386	1184	303.89	71.09	327 %
S&P 500 (est.)		58440	32130	26310	6845.5	3839.5	78 %

The top nine companies' market capitalization (market cap) increased by \$17.5 trillion, or 233%, from year end 2022 to year end 2025 – representing roughly two thirds of the S&P 500's total market cap increase of \$26.3 trillion (an 82% gain) over the same period. In other words, the AI leading companies grew at three times the rate of the overall S&P 500.⁵

Excluding the top eight AI leaders (TSMC is not included in S&P500), the remaining S&P 500 would have grown less than 40%. The growth rate ratio between the AI proxy and the rest of the index is therefore approximately 6:1.⁶ This divergence underscores a simple reality: AI is advancing at unprecedented speed, driving breakthroughs in productivity, automation and decision making across industries while reshaping economies and societies.

The absolute scale is equally striking. The top nine companies' market cap increase alone exceeds the entire market capitalization of any country except

the United States (US) – for example: China (\$11.2 trillion), Japan (\$6.6 trillion), the United Kingdom (\$4.4 trillion)⁷ – and surpasses the combined market cap changes of the top 10 countries excluding the US. This reflects the extraordinary economic impact of AI, its accelerating trajectory and its global significance.

The magnitude of this appreciation has prompted some observers to question whether AI is entering a bubble or approaching its peak. It is also important to note that this economic trajectory is uneven across regions. A stark disconnect exists in emerging hubs like India, which generates nearly one-fifth of global data but currently holds only ~1.2% of global data centre capacity.⁸ Therefore, it is essential to ground the discussion in data, historical patterns, economic growth cycles and the structural dynamics of platform shifts, safeguarding digital sovereignty while managing resource pressures.

1.4 AI transformation: A platform shift

Major technology transitions, such as the one from mainframe to PC to the internet to mobile, typically unfold over 10-15 years before reaching full industry penetration. AI is following a similar trajectory, but with far broader scope because it spans compute infrastructure, software, consumer devices and enterprise workflows.

Gartner projects global AI spending to reach nearly \$1.5 trillion in 2025,⁹ reflecting early stage but rapidly accelerating adoption across sectors.

Generative AI alone is expected to reach \$644 billion in 2025,¹⁰ growing 76.4% year-on-year, signalling a market still in the steep ascent phase – not maturity.

These long cycle investment patterns indicate that AI is not peaking. It is in the early innings of a multi-year platform transition with significant runway ahead.

1.5 Early stage AI application growth amid infrastructure constraints

Today's ecosystem is constrained not by demand, but by insufficient infrastructure capacity.

AI optimized servers are projected to grow 49% in 2026,¹¹ making them one of the fastest expanding segments of global IT spend.

Hyperscalers have increased capital expenditures by more than 60% year-on-year since 2023, reaching \$700 billion in 2026, largely to build out AI factories and data centres. The 2026 capex projections of five of these companies are:¹²

- Amazon: ~\$200 billion
- Alphabet: \$175-185 billion

- Microsoft: >\$140 billion
- Meta: \$115-130 billion
- Oracle: ~\$50 billion

On the application side, generative AI is still early in its adoption curve. Gartner, for example, places AI in the "trough of disillusionment,"¹³ meaning mainstream productivity gains and mass market applications are still ahead.

This combination – capacity constraints plus early stage application development – strongly suggests that AI value creation is far from saturated.

1.6 AI factories and access to capital

Global AI investment continues to expand at extraordinary speed. Worldwide AI spending is projected to exceed \$2.5 trillion in 2026¹⁴ – up sharply from nearly \$1.5 trillion in 2025 – with AI infrastructure representing more than half of that total as organizations accelerate infrastructure buildout and demand for advanced compute continues to surge.

Generative AI consumer spending is expected to grow from \$225 billion in 2023 to ~\$700 billion by 2030,¹⁵ reflecting rapid adoption of AI enabled devices and software.

Industry analysts also estimate that meeting AI compute demand by 2030 will require \$2 trillion in new annual revenue to sustain AI infrastructure buildout,¹⁶ underscoring the scale required to support next generation AI models.

In summary, across all indicators, the authors observed no evidence of a peak. Concerns about an “AI bubble” or “peak AI” are premature when viewed against the underlying signals:

- Spending is accelerating.
- Infrastructure demand exceeds supply.
- Enterprise and consumer adoption curves remain early.
- Hyperscaler capital investment is expanding to historic levels.

Taken together, these trends point to a sustained, multi-year global investment cycle. AI is not nearing its peak – it is entering the most significant expansion phase yet, positioning itself as one of the defining economic forces of the coming decade.

Key system stressors driven by exponential AI growth

The rapid expansion of AI is exerting unprecedented pressure on foundational systems never designed for this scale or speed of computational demand. Industrial scale drivers – ranging from the race to build AI factories with far greater computational capacity and density, to accelerating chip innovation for speed and efficiency, to the deployment of advanced cooling technologies for performance and uptime, to the efforts to meet surging demand for critical minerals – are reshaping the AI-energy-water-minerals-land nexus. Their convergence is transforming infrastructure expectations, business models and governance norms, creating system level stressors that challenge grid and water reliability, resource availability and long term sustainability.

A surge in electricity demand and grid requirements

To meet soaring AI demand, companies are rapidly building AI factories and next generation data centres with far greater capacity and density. Facilities that once operated as modest digital infrastructure now resemble heavy industrial plants. Rack densities – the amount of power used by equipment in a single server rack – that historically averaged 5-10 kilowatt (kW) have risen to 30-50 kW, with leading AI systems expected to exceed 100 kW per rack.¹⁷

This escalation is driving steep increases in electricity demand and straining generation, transmission and distribution systems. The need for reliable, high quality, low carbon power is outpacing regional supply. As AI facilities compete with households, manufacturing, EV charging and heating for limited capacity, they risk crowding out other users and increasing reliance on fossil generation in regions without sufficient clean power – creating tension between AI expansion and decarbonization goals.

Grid stress is already visible. Northern Virginia in the US and Dublin in Ireland, for example, face capacity bottlenecks, connection delays and reliability concerns. In some markets, rising AI demand threatens to delay coal and gas retirements and in some cases is leading to the reactivation of legacy plants. The clustering of multi gigawatt AI campuses also creates concentrated points of failure, heightening vulnerability to storms, heatwaves, cyber incidents and equipment outages.

Compounding the challenge, grid planners often learn of massive AI loads only months before they are required, leaving little time for coordinated, sustainable expansion. A single AI campus can draw as much power as a mid sized city, reshaping regional energy systems almost overnight. Without integrated planning across utilities, regulators, developers and policy-makers, the pace of AI growth risks outstripping the resilience of the energy systems it depends on.

AI chip innovation to boost speed and efficiency is creating new system pressures

As AI demand accelerates, operators must expand computational capacity while advancing the chips that power modern AI workloads. The industry is pushing towards faster, more efficient architectures – driven by breakthroughs in GPUs, specialized accelerators, memory bandwidth, interconnects and packaging. This rapid innovation cycle is essential to keep pace with exponential growth in model size, training complexity and real time inference.

The urgency is heightened by persistent demand-supply imbalances for high-performance compute. Even as new AI campuses come online, access to cutting-edge chips remains constrained. Higher performance chips can deliver more compute per watt, per litre of water and per square metre of land, but they also increase power draw and heat generation, requiring more advanced cooling, more resilient power delivery and more sophisticated thermal management.

The pace of innovation also disrupts traditional upgrade cycles. With new architectures emerging every 12-18 months, operators face pressure to refresh infrastructure more frequently, intensifying tensions between efficiency gains, circularity goals and capital return expectations. Faster obsolescence increases e-waste, strains mineral supply chains and complicates sustainability commitments. Meanwhile, the capital intensity of continuous upgrades can outpace depreciation schedules, forcing a rethink of financial models and long-term infrastructure strategy.

Accelerating chip innovation is both a solution and a stressor – critical for performance and efficiency, yet amplifying operational, environmental and economic pressures. Managing this duality will be a defining challenge for the next decade of AI growth.

Escalating cooling requirements: Water stress and thermal constraints

As AI factories scale, their cooling requirements rise just as quickly. Higher density racks, higher computational GPUs/ASICs, advanced accelerators and increasingly complex model architectures generate far more heat than traditional IT systems. This thermal intensity is pushing operators towards

more aggressive facility-level thermal management and server cooling strategies – liquid cooling, immersion systems and hybrid approaches that can manage extreme heat loads while maintaining uptime and performance. These technologies are becoming essential infrastructure.

The shift towards advanced cooling has profound implications for water systems and for the pace of liquid thermal technology innovation. Many high performance cooling solutions rely on significant volumes of water for heat rejection, evaporation or recirculation. In regions already facing water scarcity, drought cycles or competing municipal and agricultural needs, large scale AI campuses introduce new stress on local water availability. Even in water abundant regions, the sheer scale of AI cooling demand can strain treatment capacity, wastewater systems and watershed resilience. Water is becoming a critical constraint on where and how AI infrastructure can grow.

Cooling has become a strategic bottleneck in the AI infrastructure ecosystem. Managing water stress and thermal constraints will require integrated planning across the energy, water, land and technology domains – ensuring that AI growth does not outpace the natural and engineered systems that support it.

The AI infrastructure boom is sharply increasing the need for critical minerals

AI hardware and supporting infrastructure rely on a narrow set of specialized minerals. Copper underpins power delivery across chips, data centres and grid infrastructure. Lithium and cobalt are essential for backup power systems and energy storage. Rare earth elements and specialty metals enable motors, magnets, advanced packaging and precision manufacturing. As AI accelerates, demand for these materials is rising sharply, intensifying pressure on already strained supply chains.

Rising mineral demand places AI growth squarely within broader debates about resource security, environmental stewardship and geopolitical influence. Without diversified supply chains, responsible sourcing and circular economy strategies, mineral requirements risk becoming a structural vulnerability – shaping not only the pace of AI expansion but also its social and environmental footprint.



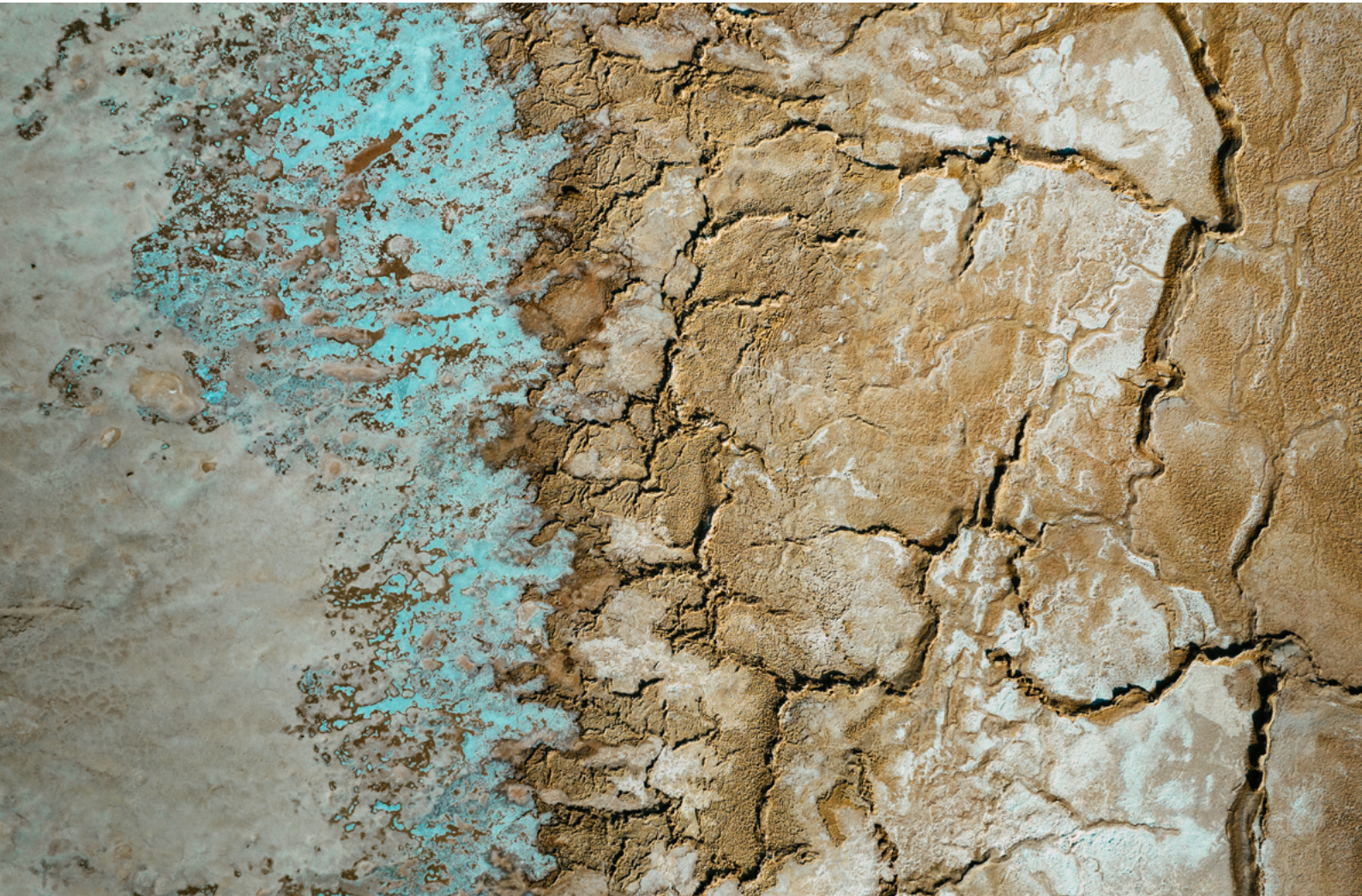
TABLE 2 | Key system stressors driven by AI growth

 Industrial driver	 System stressor	 Proof points	 Potential actions
Race to build AI factories with far greater computational capacity and density to meet surging demand	<ul style="list-style-type: none"> Higher electricity demand strains generation and grids Sizeable needs of reliable, high-quality, low-carbon energy Competition with other users Emissions rise if fossil energy fills gaps 	<ul style="list-style-type: none"> AI could drive ~20% of global demand growth by 2030; ~40% of that from fossil at today's mix DC power use projected to reach ~945 TWh by 2030 (equivalent to Germany and France's current total electricity usage) 	<ul style="list-style-type: none"> Drive PUE down: target <1.2 (from ~1.57). Cut compute per job: use efficient chips, trim models, sleep idle gear, improve thermal management. Site near clean power and pair with 24/7 clean energy and storage to enhance efficiency and cut real emissions
Accelerating AI chip innovation to boost speed, efficiency and responsiveness	<ul style="list-style-type: none"> Supply-demand imbalance for high-performance chips amid increasingly regulated global environments Thermal challenges – more power, more transistors, higher density and more heat Legacy system upgrade cycles and tensions among efficiency, circularity goals and capital return expectations 	<ul style="list-style-type: none"> NVIDIA's introduction of H100, B100 and GB200 NVL72 (successive generations of data centre AI computing infrastructure), and advance of Rubin CPX Performance race by leaders – NVIDIA GPUs vs. Google TPUs (Tensor Processing Units) vs. AMD vs. Broadcom Rubin GPUs expected to consume up to 3.6kW per chip, compared to Blackwell chips at 1.2kW to 1.4kW. 	<ul style="list-style-type: none"> Enhance supply chain capacity, build up strategic partnerships, diversify value chains and drive policy alignment Advance cooling solutions, server architectures and material innovation Balance return on investment, pursue circularity initiatives and explore policy-driven incentives
Data centre cooling as escalating compute generates more heat	<ul style="list-style-type: none"> Rising water withdrawal/ consumption for cooling, especially in stressed basins Competition with households, farms, industry Social licence to operate and stranded-asset risk 	<ul style="list-style-type: none"> DC water use forecast >50% above 2022 by 2030 Two in three of new US sites planned in high-stress regions By 2050s, ~45% of DCs face high water risk 	<ul style="list-style-type: none"> Pick the right basin using current and future-climate scenarios Design better to transform WUE and take on water reclamation projects to achieve net-positive water impacts Measure and give back: Track withdrawals by basin; fund in-basin replenishment
Demand for critical minerals for AI hardware; extraction to meet exponential needs (lithium, copper, cobalt, rare earths; fossil fuels for power)	<ul style="list-style-type: none"> Mining is water/energy-intensive with risks of contamination, tailings failure, habitat loss Land-use conflict with farms, homes, biodiversity Hotspot sourcing (e.g. cobalt from the Democratic Republic of Congo) raises human-rights and biodiversity risks 	<ul style="list-style-type: none"> Transition-mineral demand may triple by 2030 One in three mines in key biodiversity areas; Chile/Peru supply ~40% of copper ~43% of transition-related mines on/near Indigenous/ community lands 	<ul style="list-style-type: none"> People and communities: Secure consent, agree local jobs/benefits, provide grievance and remedy Nature : Ensure no-go in critical habitats; avoid, reduce, restore; independent checks. Water and tailings: Recycle/ reinject water, use clean power and adopt dry-stack/ in-pit tailings

2

A call to action

The AI-energy-water-minerals-land nexus, and why it is critical.



The rapid expansion of AI is accelerating the buildout of data centres and AI factories, sharply increasing demand for electricity, water for cooling, the minerals required for chips, batteries and power systems, and the land required to host these facilities and their supporting infrastructure.

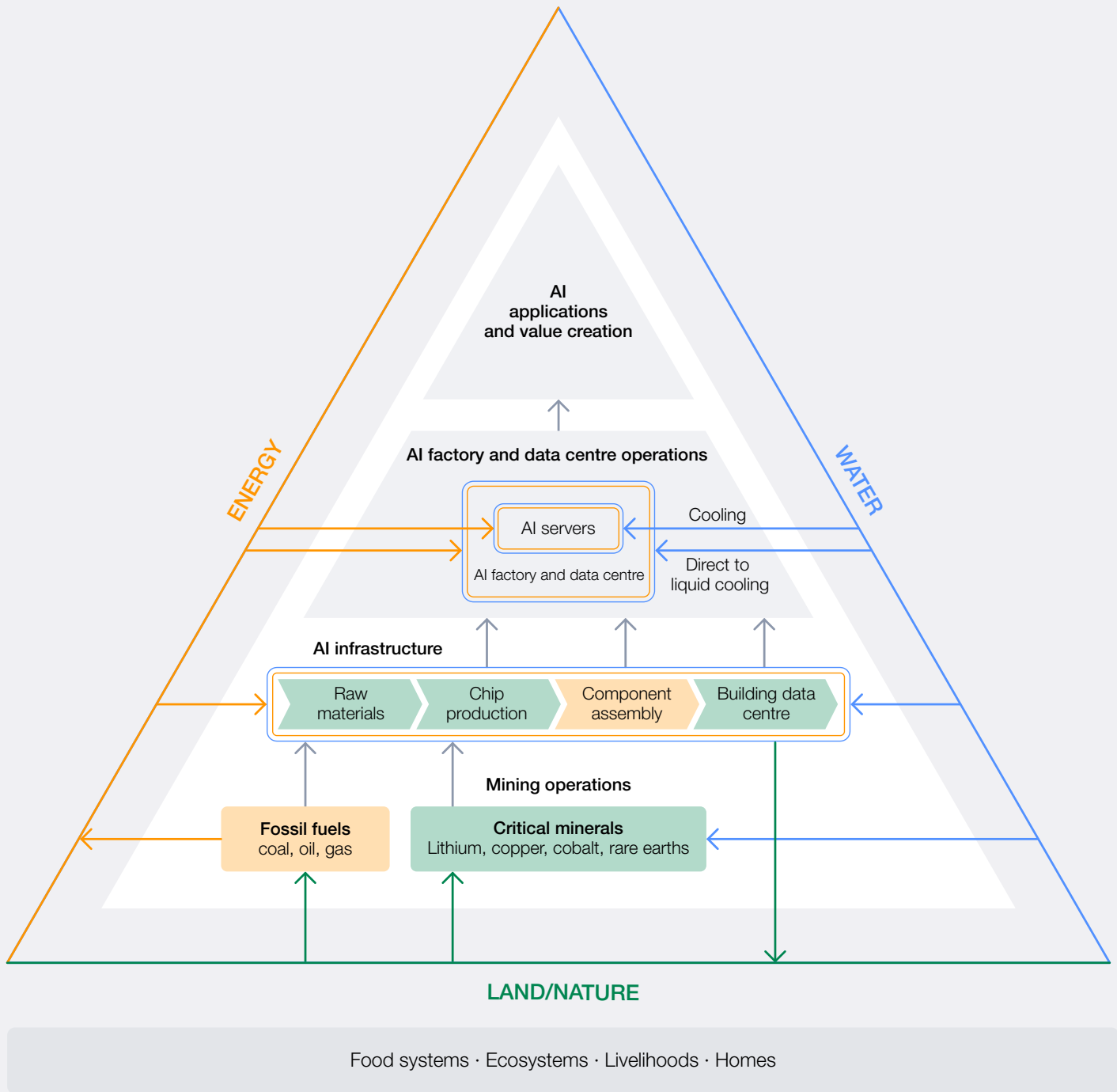
This surge places AI in direct competition with agriculture, ecosystems and local communities. Because these resource systems are tightly interconnected – higher energy use raises emissions, water diverted for cooling strains food production, and mineral extraction disrupts communities and biodiversity – stress in any one domain quickly cascades across the others.

The AI-energy-water-minerals-land nexus captures this interdependence. As AI models scale and infrastructure expands, they draw simultaneously

on electricity, water, minerals and land – each with its own physical limits, environmental impacts and geopolitical risks. None of these systems operate independently; strain in one area quickly propagates into others, shaping the pace, cost and sustainability of AI deployment.

The nexus framework highlights that optimizing one resource in isolation often shifts burdens elsewhere. More efficient chips may reduce energy per computation but increase heat density and water demand. Expanding renewable energy lowers emissions but requires land and minerals. New data centres can strain local water systems or displace other land uses. These interdependencies require integrated planning that accounts for cumulative impacts across the full resource system.

FIGURE 2 | The interplay of the AI value chain with energy, water, minerals and land



2.1 Energy: A highly visible choke point

Energy is the most visible pressure point, as high density compute drives steep growth in electricity demand and grid congestion, reshaping power systems by changing both the scale and profile of load. Large scale AI compute requires continuous, high quality, 24/7 power with almost no tolerance for interruption – far more rigid than traditional commercial or industrial demand. This shift intensifies pressure on the energy trilemma of security, affordability and decarbonization.

Global data centre electricity use is projected to rise from 415 terrawatt-hour (TWh) in 2024 to 945 TWh by 2030,¹⁸ exceeding the current consumption of Germany and France combined. AI may drive more than 20% of total electricity demand growth until 2030,¹⁹ with fossil fuels still supplying nearly 40% of new generation. Integrating renewables is essential to manage continuity, cost and emissions, but renewables alone cannot meet the round the clock power profile of AI.

While solar and wind are essential to decarbonization, their intermittency conflicts with AI's need for constant, high-load baseload power. As a result, AI is simultaneously:

- Accelerating renewable deployment (through long-term power-purchase agreements).
- Driving new investment in firm generation, including gas, nuclear and extended fossil capacity, particularly where storage and grid flexibility lag.

This tension is already reshaping energy infrastructure decisions. Hyperscalers are moving beyond renewable certificates to direct ownership of generation assets. For example, Microsoft's support for restarting the Three Mile Island nuclear facility signals renewed interest in large scale dispatchable power. Similar moves into gas backed generation, small modular reactors and private

transmission reflect a broader reality: AI growth depends on firm power.

Where clean baseload is limited, new gas plants risk fossil lock in, as 20-30 year financing horizons embed emissions well past mid-century. At the same time, the geographic clustering of AI loads creates local grid saturation, driving costly upgrades and heightening exposure to outages, extreme weather and cyber risk. Furthermore, these bottlenecks are often institutional as much as physical. Tremendous regulatory friction from siloed permitting processes and unsustainable cross-subsidy surcharges compounds project risk and financing costs, even when clean power technology is available.

AI is not simply increasing electricity demand – it is reconfiguring power systems, and exposing structural gaps in grid flexibility, storage, transmission and clean baseload supply. How utilities, regulators and operators respond will determine whether AI becomes a catalyst for clean energy acceleration or a driver of extended fossil fuel dependence.

Energy is inseparable from water, which is essential for cooling and many forms of power generation; from minerals, which underpin chip manufacturing and renewable energy systems; and from land, which determines where AI campuses, substations, transmission lines and clean energy projects can be sited. Each resource faces its own bottlenecks – water scarcity, mineral supply concentration, land use conflicts and grid constraints – and AI growth amplifies all of them simultaneously.

Rising demand across these systems fuels inflation, disrupts communities, and heightens environmental and social risks. These interdependencies underscore the need for integrated resource strategies, as highlighted in the World Economic Forum's "Artificial Intelligence's Energy Paradox."²⁰

2.2 Water: The overlooked AI growth limiter

Water is becoming one of the most binding – and least visible – constraints on AI expansion. While electricity demand receives most attention, water is embedded across the AI value chain, from cooling systems in data centres to electricity generation and semiconductor manufacturing. Because many AI facilities rely on freshwater or potable quality water, their growth can create direct trade offs with households, agriculture and ecosystems.

AI data centres require large volumes of water for cooling: a one-megawatt facility uses up to 25.5 million litres annually,²¹ illustrating how even

relatively small facilities can place meaningful demands on local water systems. At scale, the cumulative impact is significant. Global data centre water consumption is projected to reach 450 million gallons per day by 2030 – up from 292 million gallons per day in 2022,²² equivalent to the daily water needs of roughly 5 million people.

Water demand is driven primarily by thermal management. Training and running large AI models generates intense heat, and many hyperscale facilities rely on evaporative cooling, which consumes freshwater rather than simply circulating it.

Research suggests that training a single frontier AI model can require hundreds of thousands of liters of freshwater,²³ depending on location, cooling design and the electricity mix.

At the application level, AI inference also carries a meaningful footprint. Estimates suggest roughly 500 ml of water per 10-50 responses for models such as GPT 3.²⁴ While the water use per interaction is small, the scale of global AI usage means these impacts accumulate rapidly.

Water demand is also highly uneven geographically. Two-thirds of US data centres are located in high-water-stress regions,²⁵ while 43% of global facilities operate in areas already experiencing water scarcity.²⁶ Many of the fastest-growing AI infrastructure hubs are located in regions where water availability is already under pressure.

In the US, Phoenix and Maricopa County, Arizona, have become major data centre development zones despite long-term groundwater depletion and restrictions on new housing tied to water availability. Northern Virginia in the US, the world's largest concentration of data centres, faces growing concerns about the cumulative impacts of hyperscale facilities on regional water and energy systems. In Dalles, Oregon, tensions have emerged over water withdrawals used to cool large data centre campuses drawing from the Columbia river basin.

These dynamics are not limited to the US. In Europe, major data centre expansions are occurring

in Aragón and Catalonia in Spain, regions that have experienced prolonged drought in recent years. Similar tensions are emerging in Chile and India, where water scarcity and industrial demand are already creating competition between sectors.

Large data centres can withdraw millions of gallons of water per day, competing directly with agriculture, households and ecosystems. Excessive withdrawals can degrade rivers and aquifers, reduce streamflow and harm biodiversity. Climate change compounds the challenge by increasing the frequency and severity of heatwaves – which raises cooling demand – while making water supplies less predictable.

Public resistance is rising where communities perceive scarce water being prioritized for AI infrastructure. In Arizona, several municipalities have restricted or banned the use of potable water for data centre cooling following public backlash. Similar debates are emerging in Spain, Chile and India, where policy-makers are increasingly scrutinizing water allocations for digital infrastructure.

Water, once a secondary consideration in digital infrastructure planning, is rapidly becoming a critical siting and social-licence constraint for AI expansion. Without reliable water access – or widespread adoption of low-water or water-free cooling technologies – AI infrastructure faces growing operational risk.



2.3 Minerals: Supply chain and geopolitical vulnerability

AI infrastructure depends on a wide range of materials whose extraction is energy- and water-intensive and often environmentally damaging. Many of these minerals are scarce or concentrated in ecologically fragile or geopolitically sensitive regions, creating structural risks to scalability and resilience.

These supply chains are fragmented across tiers and geographies²⁷ – and highly concentrated. Roughly 70% of global cobalt production comes from the Democratic Republic of Congo.²⁸ China controls 80-90% of rare earth processing.²⁹ More than half of the world's lithium is extracted from the Andean salt flats; and Indonesia dominates nickel production, increasingly in biodiverse rainforest regions. With AI hardware demand expected to double or triple by 2030, these concentrations amplify ecological, social and geopolitical exposure.

Mining expansion increasingly intersects with biodiversity hotspots, water-stressed regions and Indigenous lands. As production ramps up, so do the pressures that drive deforestation, water contamination and community conflict – often in jurisdictions with limited oversight. Supply disruptions, export controls or shortages of niche materials such as gallium, germanium or high-purity silicon can halt entire production lines. Local constraints quickly cascade into global bottlenecks, slowing AI deployment and increasing operational and financial risk.

The rapid physical expansion of AI is amplifying pressure on mineral supply chains already strained by electrification, clean energy deployment and defence demand. Copper is the most acute example: indispensable across the AI value chain – from chips and GPUs to servers, data centre power systems and the grid infrastructure that delivers electricity. Global copper demand could rise 30%

by 2040,³⁰ but supply growth is lagging, as new mines require 10-15 years to permit and develop and ore grades continue to decline, creating a potential structural deficit by the early 2030s.

Other minerals face similar constraints:³¹

- **Lithium** demand is projected to increase five-fold by 2030,³² driven by batteries for energy storage, backup power and AI-enabled devices.
- **Cobalt and nickel** demand continues to rise, with supply highly concentrated in politically and environmentally fragile regions.
- **Rare earth elements**, critical for motors, power electronics and cooling systems, remain heavily dependent on Chinese processing capacity, exposing AI supply chains to geopolitical risk.

These pressures carry real operational consequences. S&P Global finds that 71%³³ of transition mineral mining projects overlap with areas of high biodiversity value, increasing the likelihood of permitting delays, legal challenges and reputational risk. In parallel, 43%³⁴ of transition mineral mines are located on or near Indigenous or local community lands, contributing to nearly 800 recorded community-related conflict incidents globally – many of which have disrupted operations or delayed projects.

As AI hardware demand accelerates, these constraints raise the risk of material bottlenecks, price volatility and stranded assets, particularly where supply chains lack social licence or fail to adapt to climate-driven disruptions. Without faster permitting, expanded recycling, material substitution and circular economy solutions, mineral availability could become a hard ceiling on AI infrastructure growth.



2.4 Land: Spatial pressure on communities and ecosystems

AI's physical footprint is reshaping land use faster than many regions can manage. Large campuses require extensive space for server halls, substations, backup generation, cooling systems, transmission corridors and security buffers. Individual sites can span 100-400 acres, rivalling major logistics hubs or industrial parks.

As AI infrastructure gets concentrated in clusters, land pressure becomes highly localized. The most attractive sites – near fibre routes, power infrastructure and urban demand centres – are also where land is already scarce and contested. Communities are increasingly pushing back over farmland loss, visual and noise impacts, rising housing costs and limited local benefit.

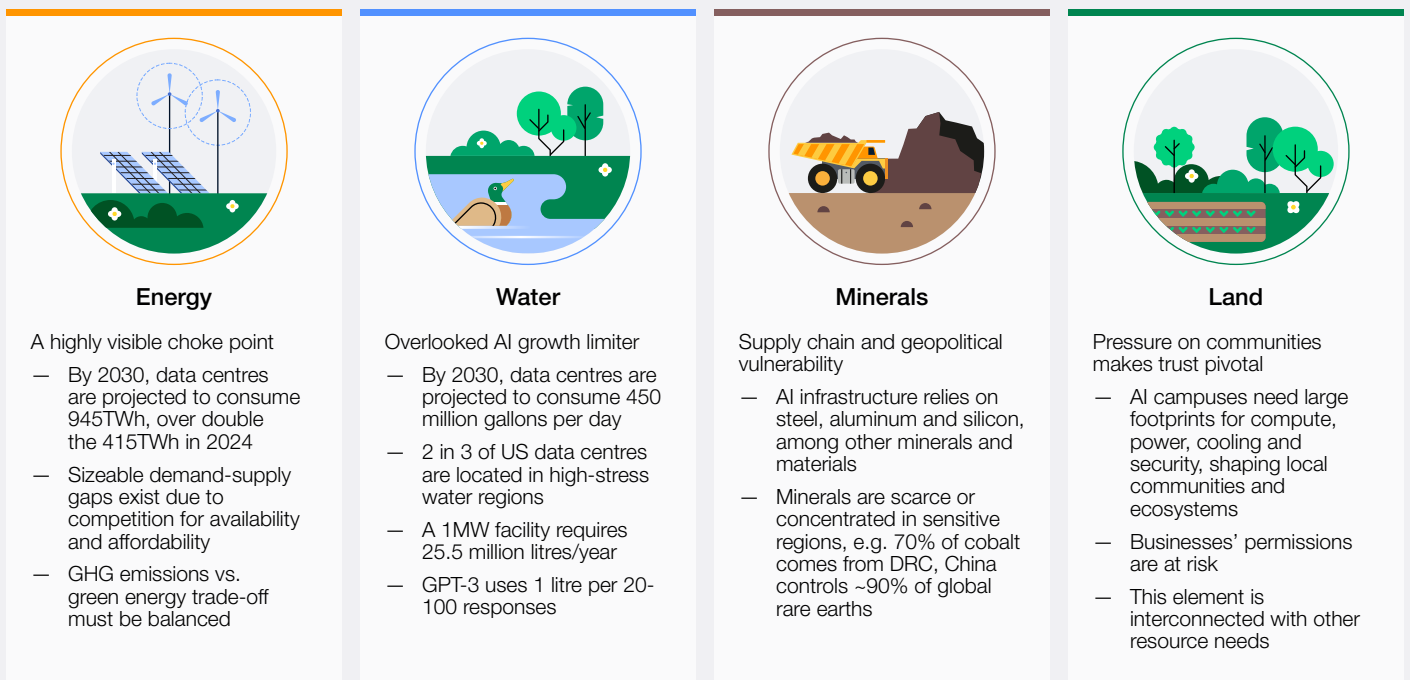
Across the US and Europe, land-use conflicts are delaying or derailing projects. Community opposition has stalled or cancelled tens of billions of dollars in proposed data centre developments over the past two years. In Northern Virginia in the US, Ireland and the Netherlands, land scarcity combined with grid and water constraints has forced governments to pause or tightly regulate new approvals.³⁵

Land pressure also intersects with biodiversity. Transmission lines, substations and renewable-energy facilities built to serve AI often overlap with sensitive habitats. Without integrated land-use planning, AI infrastructure risks fragmenting ecosystems and accelerating habitat loss – particularly when paired with upstream mining and energy development.

These dynamics underscore a broader shift: land is no longer a neutral input to digital growth. Where and how AI infrastructure occupies space increasingly determines its social acceptability, regulatory exposure and long-term viability.

Understanding the AI-energy-water-minerals-land nexus is essential for anticipating system level risks, identifying bottlenecks and designing resilient pathways for AI growth. It creates a shared framework for operators, utilities, policy-makers, investors and communities – ensuring AI expansion delivers economic and societal value without exceeding the limits of the natural and engineered systems that support it.

FIGURE 3 Core elements of the AI-energy-water-minerals-land nexus risks



Siloed approach vs. total solution – Why tackling of AI-and-energy nexus risks is crucial



Integrated action on the AI-energy-water-minerals-land nexus

As summarized before, the AI-energy-water-minerals-land systems are tightly interconnected. AI facilities make this visible: they consume significant amounts of electricity and water for cooling, while the power systems and semiconductor supply chains behind them depend on water, energy, minerals and land. A single water shortage can simultaneously limit cooling capacity, reduce power generation, constrain agriculture and damage ecosystems – turning a localized issue into a multi sector risk affecting grids, food systems, communities and AI operations.

Regulation, however, remains fragmented. Grid operators focus on capacity, water agencies on basin stress, land planners on zoning and AI policy-makers on model safety. This siloed approach leaves the convergence zone – where pressures accumulate – largely unmanaged. As a result, AI's resource dependencies and sustainability impacts remain underestimated and obscured within complex, geographically concentrated supply chains.

A nexus strategy treats these systems as an integrated whole. It aligns AI growth with realistic energy and water availability; plans data centres and renewable deployment with full visibility into mineral constraints; and ensures land use decisions balance ecological integrity, food security and industrial needs. Coordinated planning reduces unintended consequences such as water intensive clean energy projects or AI infrastructure that increases fossil based grid demand. When optimized together, these systems become more resilient, efficient, synergistic and equitable.

The urgency is already visible. Water stressed regions are struggling to support new data centres, mineral supply chains are lagging behind battery and renewable demand, and communities are resisting land intensive energy projects. These tensions slow deployment, raise costs and erode public trust. Integrated planning helps anticipate conflicts, map trade offs, identify synergies and direct investment towards the highest overall system value.

The AI-energy-water-minerals-land nexus is not a constraint but a strategic advantage for early movers. Cross sector coordination accelerates innovation, reduces resource and regulatory risk, and supports infrastructure that is both sustainable and future ready. Ensuring AI strengthens rather than destabilizes environmental and economic systems requires a shift from reactive, siloed management to proactive, systems level stewardship – ensuring circular resource flows, coordinated energy water planning, use of low impact materials and ecosystem aligned siting.

Integrated thinking is no longer optional. It is the foundation for unlocking and sustaining the future of AI – one that protects ecosystems, supports communities and ensures long-term viability as AI continues to scale.

3

What key stakeholders can do

From risk to opportunity: Harnessing AI productively, deliberately and safely.

The rapid expansion of AI brings undeniable risks – rising resource demands, opaque supply chains and growing environmental pressures – but it also opens the door to extraordinary opportunity. The task now is to shift from reactive concern to proactive design: building AI systems that deliver economic and social value while operating within sustainable limits. This requires moving beyond short term fixes towards long-term strategies that ensure the benefits outweigh the costs.

Turning constraints into competitive advantage:

The nexus constraints shaping the growth of AI are not inevitable; they stem from choices about siting, sourcing, technology design and supply chain management, and they now determine where AI can scale and at what cost. Left to uncoordinated market forces, scarce resources risk being consumed by low value uses while high impact applications lose out. But when treated as design parameters, these pressures become competitive advantages.

Aligning AI deployment with community and environmental priorities builds trust, reduces regulatory friction and strengthens long-term legitimacy. Integrated planning – co locating data centres with clean energy buildout, linking mineral strategy with domestic manufacturing and pairing compute loads with flexible demand – turns potential bottlenecks into engines of innovation and resilience. Sustainability becomes a core strategic asset, not a trade-off, for those defining the next era of AI leadership.

Several companies and countries are already responding to these challenges with forward-looking solutions. Princeton Digital Group, for instance, implemented India's first hourly carbon-free energy (CFE) matching scheme³⁶ at its 150MW data centre in Mumbai, and uses hydrotreated vegetable oil as a diesel substitute in Indonesia.³⁷ Similarly, Microsoft's "Community First AI Infrastructure" strategy emphasizes building AI infrastructure in a way that delivers local economic, environmental and social benefits, reflecting a broader shift towards more sustainable and inclusive approaches.³⁸

Strategic direction now determines who will thrive in the accelerating AI era, as countries and companies set deliberate pathways. Varying

approaches are evident, for example: the US is prioritizing competitive advantage through market driven innovation, semiconductor onshoring and limited regulatory oversight; China, facing restricted access to cutting edge GPUs, is pursuing long-term independence by scaling domestic compute and is competing for leadership in physical AI systems; Singapore and the United Arab Emirates (UAE) are investing in talent and sovereign compute to anchor safe, high impact deployment.

While these strategies reflect differing national priorities, they also highlight the risk of fragmented approaches to a deeply interconnected system. Countries that set strategic pathways – not just rapid ones – and take into account an integrated approach accounting for the local economic, environmental and social benefits, are best positioned to lead. In this context, new models of partnership and multi-stakeholder collaboration across industry, government and civil society will be critical to align incentives, manage shared resource constraints and unlock system-wide resilience and opportunity.

The same principle applies to industry: speed without a sustainability pathway creates fragility, while intentional growth builds resilience. Productive AI requires aligning innovation with responsible resource use – optimizing energy, water, minerals and land while managing AI's own footprint through clear performance expectations, sustainability metrics and transparent supply chains. Safety must extend beyond algorithms to the physical systems that support AI, requiring early risk anticipation, infrastructure safeguards and community protection. With deliberate planning, AI can accelerate clean energy transitions, circular material flows and climate resilience – but only when understood as part of a broader socio-ecological system.

Ultimately, the path forward is about balance – harnessing AI's transformative potential while ensuring its expansion remains safe, transparent and aligned with long-term societal goals. When risk is treated as a design challenge rather than an obstacle, governments and industries can turn AI into a driver of opportunity, resilience and shared prosperity.

Conclusion – The way forward

AI now sits at the centre of a broader economic, socio-ecological system spanning energy, water, minerals, land and community well being. It is a defining moment to unlock and sustain the future of AI. The window to guide the evolution of AI is now open and the deliberateness of stakeholders' actions will determine the path forward.

Roles and responsibilities of stakeholders and recommended actions across the value chain

Taking a reverse value chain perspective, the corresponding roles and responsibilities can be summarized as follows.

Building a net-positive AI ecosystem requires coordinated action across the entire value chain, because no single actor can address the intertwined challenges of energy demand, water use, mineral sourcing, land impacts or social equity alone. Beyond recognizing that the AI-energy-water-minerals-land nexus is a strategic advantage for early movers, stakeholders must shift from viewing

AI as a standalone technology to understanding it as part of a broader socio-ecological system.

Integrated thinking is now essential: each stakeholder group holds distinct technical, financial, regulatory and operational levers that shaped how the AI value chain is deployed and how its benefits are shared. Clarifying these roles is critical to align incentives, reduce friction, and accelerate sustainable, resilient and socially grounded AI growth.

The future trajectory of AI will be determined not only by technological breakthroughs but by the ability of stakeholders across the value chain to build synchronized, resilient and sustainable systems. This moment demands integrated thinking, coordinated action and shared accountability. Stakeholders who recognize and act on the AI-energy-water-minerals-land nexus early will shape both competitive advantage and societal outcomes. The choices made now will determine whether AI becomes a catalyst for long-term prosperity or a source of compounding systemic stress.

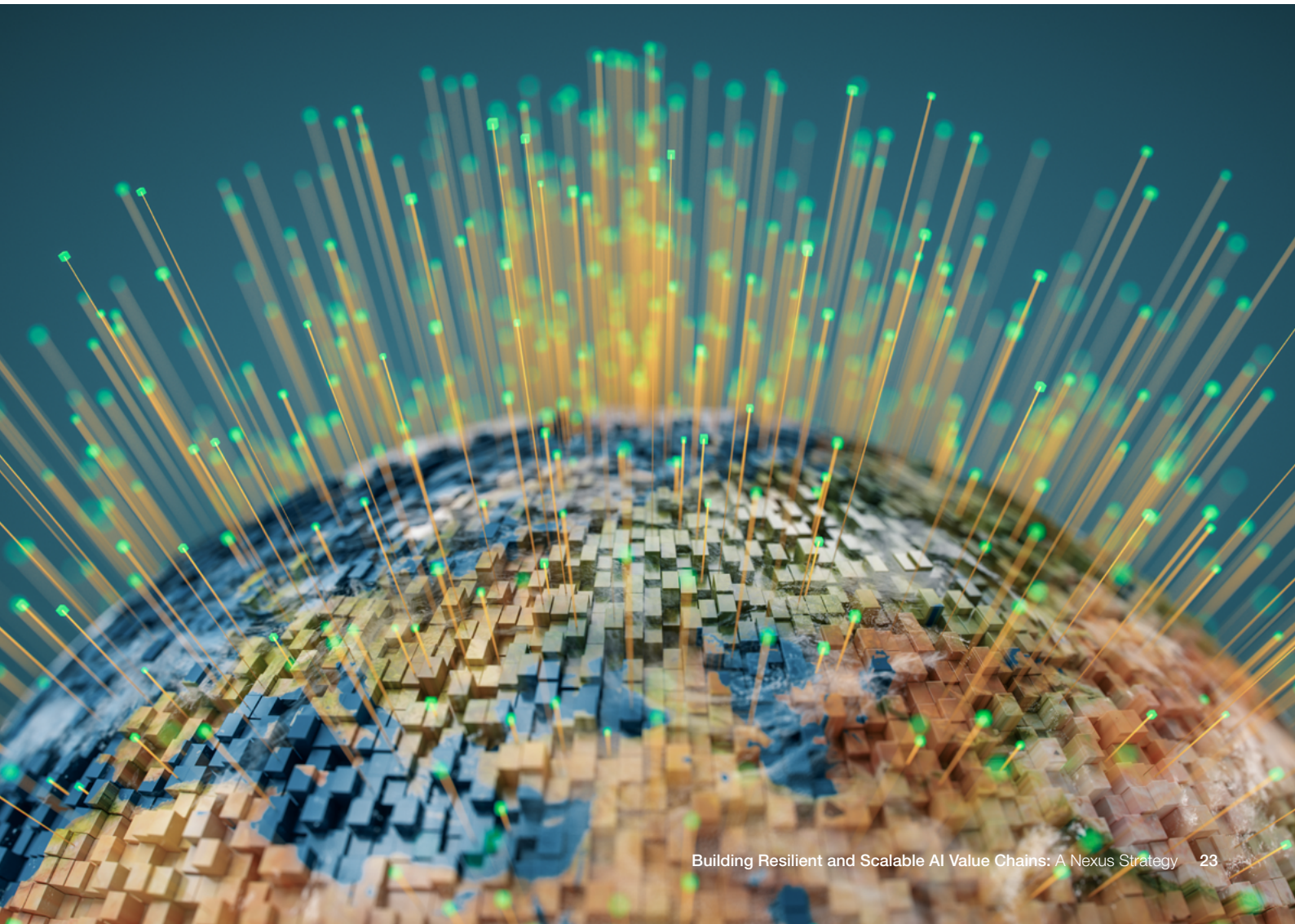


TABLE 3 | What stakeholders can do

Stakeholder group	Core responsibilities	Key recommended actions
AI developers; physical AI capabilities extended into robots and machines that interact with the physical world; solution companies	Create and commercialize value; optimize model architecture, training efficiency and data governance.	Use more efficient AI (sparsity/quantization): Design AI systems to do the same work using less computing power and energy; collaborate with value chains to build risk maps, coordinate mitigation actions; adopt carbon/water/sustainability goals; develop AI talent and upskill workforce.
AI factory and data centre operators	Site and operate AI factories; set uptime and performance goals; invest in clean energy; reduce water dependency, maintain resource transparency.	Locate and operate in alignment with energy, water, minerals and land constraints; target PUE < 1.2, WUE with net-positive impacts; co-locate with low-carbon energy; create flexible demand-response programmes; deploy closed-loop cooling, circular hardware strategies.
AI infrastructure companies	Innovate high-performance/low-impact hardware, sustainability, ethical mineral sourcing.	Collaborate to build risk maps, take mitigation actions, support greener data centres; design for modularity/recyclability; invest in water-recycling at fab plants; expand supplier verification.
Energy providers and grid operators	Scale clean generation; grid stability and affordability under rising AI demand.	Develop long-term AI load plans; expand transmission corridors; use data centres as flexible grid resources that can adjust power use in response to prices and demand.
Water utilities and environmental agencies	Protect stressed water basins; set industrial usage and discharge standards.	Require water-risk assessments; support innovation in cooling and water-efficient tech; incentivize reclaimed water; foster basin-wide collaboration; monitor biodiversity.
Mineral producers and partners	Ensure responsible extraction; build resilient, circular supply chains for critical minerals.	Adopt low-impact mining tech; scale recycling for lithium and cobalt; collaborate on closed-loop flows; take risk mitigating actions.
Governments and regulators	Set policy frameworks; protect ecosystems; advance sovereign AI and national security.	Facilitate public-private collaboration; establish clear siting guidelines aligned with energy, water, land and mineral constraints; require standardized reporting and disclosure of resource use and dependencies; and create incentives for low-impact, resource-efficient infrastructure. At the same time, develop and upskill AI talent, lead sovereign AI (a country can develop, control and run its own AI systems, data and infrastructure – rather than relying entirely on others), with strong safety and ethical standards; integrate nexus risk metrics into capital allocation.
Investors and financial institutions	Drive return on investments; prioritize long-term value; support AI's future.	Integrate nexus risk metrics into capital allocation; prioritize investment and incentives towards high-impact AI applications that deliver measurable social and environmental benefits; influence AI-nexus best practices; require disclosure of key dependencies; fund grid upgrades and water net-zero projects; and engage on labour standards and community impacts, including through public-private collaboration where relevant.
Communities and civil society	Align infrastructure with local needs; hold operators accountable for social licence.	Participate in environmental reviews; advocate for fair labour; partner on shared-benefit models (e.g. district heating).

Contributors

Lead authors

Sam Hsenghung Hsu

Member, Global Future Council, World Economic Forum; Board Member, Fastenal

Lauren Smart

Member, Global Future Council, World Economic Forum; Global Head of Sustainable Finance, Bloomberg

World Economic Forum

Ojasvee Arora

Programme Specialist, Centre for Energy and Materials, World Economic Forum

Olivia Zeydler

Council Manager, Global Future Council on Energy Nexus; Lead, Emerging Markets, Energy and Materials, World Economic Forum

Appreciation also goes to colleagues from the Forum's Centre for Energy and Materials, Centre for AI Excellence and Centre for Nature and Climate, who provided feedback to early versions of the report.

Acknowledgements

Sincere appreciation is extended to the members of the Global Future Council on Energy Nexus, who provided both critical feedback and use case contributions as well as the Co-Chairs and project team, listed below.

Ganesh Dileep

CEO's Chief of Staff, Council on Energy, Environment and Water (CEEW)

Arunabha Ghosh

Founding CEO, Council on Energy, Environment and Water (CEEW); Co-Chair, Global Future Council on Energy Nexus, World Economic Forum

Jane Nelson

Founding Director, Corporate Responsibility Initiative; Senior Research Fellow, Harvard Kennedy School of Government; Co-Chair, Global Future Council on Energy Nexus, World Economic Forum

Production

Bianca Gay-Fulconis

Designer, 1-Pact Edition

Tanya Korniichuk

Illustrator, 1-Pact Edition

Madhur Singh

Editor, World Economic Forum

Endnotes

1. Reuters. (2026, February 23). *Big Tech to invest about \$650 billion in AI in 2026, Bridgewater says*. <https://www.reuters.com/business/big-tech-invest-about-650-billion-ai-2026-bridgewater-says-2026-02-23/>
2. International Energy Agency. (2025a). *Energy demand from AI*. <https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai>
3. S&P Global. (2025a). *Beneath the surface: Water stress in data centres*. <https://www.spglobal.com/sustainable1/en/insights/special-editorial/beneath-the-surface-water-stress-in-data-centers>
4. International Energy Agency. (2025b). *Overview of outlook for key minerals*. <https://www.iea.org/reports/global-critical-minerals-outlook-2025/overview-of-outlook-for-key-minerals>
5. CompaniesMarketCap. (n.d.). *Countries ranked by market cap*. <https://companiesmarketcap.com/all-countries/>
6. Ibid
7. Ibid
8. CEEW. (2026). *Scaling India's data centre ecosystem*. <https://www.ceew.in/sites/default/files/ceew-data-centre-study-web-ready-final.pdf>
9. Gartner. (2025a). *Gartner Says Worldwide AI Spending Will Total \$1.5 Trillion in 2025*. <https://www.gartner.com/en/newsroom/press-releases/2025-09-17-gartner-says-worldwide-ai-spending-will-total-1-point-5-trillion-in-2025>
10. Gartner. (2025b). *Gartner Forecasts Worldwide GenAI Spending to Reach \$644 Billion in 2025*. <https://www.gartner.com/en/newsroom/press-releases/2025-03-31-gartner-forecasts-worldwide-genai-spending-to-reach-644-billion-in-2025>
11. Gartner. (2026a). *Gartner says worldwide AI spending will total \$2.5 trillion in 2026*. <https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026>
12. S&P Global. (2026). *Sector Review: U.S. Tech Earnings: Hyperscalers Again Are Hyperspending*. <https://www.spglobal.com/ratings/en/regulatory/article/sector-review-us-tech-earnings-hyperscalers-again-are-hyperspending-s101669934>
13. Gartner. (2025c). *Latest Hype Cycle For Artificial Intelligence Goes Beyond GenAI*. <https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence>
14. Gartner. (2026b). *Gartner says worldwide AI spending will total \$2.5 trillion in 2026*. <https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026>
15. Counterpoint Research. (2025). *Counterpoint Forecasts Global Generative AI Consumer Spending to Reach \$699 Billion by 2030*. <https://www.itiger.com/news/1158342842>
16. Bain & Company. (2025). *Bain & Company's 6th annual global technology report*. [https://www.bain.com/about/media-center/press-releases/20252/\\$2-trillion-in-new-revenue-needed-to-fund-ais-scaling-trend---bain-companys-6th-annual-global-technology-report/](https://www.bain.com/about/media-center/press-releases/20252/$2-trillion-in-new-revenue-needed-to-fund-ais-scaling-trend---bain-companys-6th-annual-global-technology-report/)
17. Deloitte Insights. (2024). *As generative AI asks for more power, data centers seek more reliable, cleaner energy solutions*. <https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/genai-power-consumption-creates-need-for-more-sustainable-data-centers.html>
18. International Energy Agency. (2025c). *Energy demand from AI*. <https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai>
19. Goldman Sachs. (2025). *AI to drive 165% increase in data center power demand by 2030*. <https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030>
20. World Economic Forum. (2025a). *Artificial intelligence's energy paradox: Balancing challenges and opportunities*. <https://www.weforum.org/publications/industries-in-the-intelligent-age-white-paper-series/energy/>
21. CloudTech. (2025). *Cloud's hidden cost: Data centre water consumption creates a global crisis*. <https://www.cloudcomputing-news.net/news/data-centre-water-consumption-crisis/>
22. Red Oak Consulting. (2025). *Data Centres and Water Consumption*. <https://www.redoakconsulting.co.uk/blog/data-centres-and-water-consumption/>
23. Gizmodo. (2023). *Thirsty AI: Training ChatGPT required enough water to fill a nuclear reactor's cooling tower, study finds*. <https://gizmodo.com/chatgpt-ai-water-185000-gallons-training-nuclear-1850324249>
24. Zhang, Q., et al. (2025). *Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models*. arXiv. <https://arxiv.org/pdf/2304.03271>
25. Bloomberg. (2025). *AI is draining water from areas that need it most*. <https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/>
26. S&P Global. (2025b). *Beneath the surface: Water stress in data centres*. <https://www.spglobal.com/sustainable1/en/insights/special-editorial/beneath-the-surface-water-stress-in-data-centers>
27. World Economic Forum. (2025b). *From Minerals to Megawatts: Building Resilience for EVs, Data Centres and Power Grids*. <https://www.weforum.org/publications/from-minerals-to-megawatts-building-resilience-for-evs-data-centers-and-power-grids/>

28. World Economic Forum. (2025c). *Making Mining Safe and Fair*. <https://www.weforum.org/publications/making-mining-safe-and-fair-artisanal-cobalt-extraction-in-the-democratic-republic-of-the-congo/>
29. German Marshall Fund of the United States. (2025). *Rare Earth Statecraft Phase Two*. <https://www.gmfus.org/news/rare-earth-statecraft-phase-two>
30. International Energy Agency. (2025d). *Overview of Outlook for Key Minerals*. <https://www.iea.org/reports/global-critical-minerals-outlook-2025/overview-of-outlook-for-key-minerals>
31. S&P Global. (2024). *Rocks and hard places*. <https://www.spglobal.com/sustainable1/en/insights/special-editorial/rocks-and-hard-places-the-ecosystem-risks-of-mining-for-energy-transition-minerals>
32. S&P Global. (2022). *Battery metal demand set to rise substantially by 2030*. <https://www.spglobal.com/energy/en/news-research/latest-news/energy-transition/052322-battery-metal-demand-set-to-rise-substantially-by-2030-iea>
33. S&P Global. (2025c). *Community matters: Mining, local engagement and the race for critical energy transition minerals*. <https://www.spglobal.com/sustainable1/en/insights/special-editorial/community-matters-mining-local-engagement-and-the-race-for-critical-energy-transition-minerals>
34. S&P Global. (2025d). *Community matters: Mining, local engagement and the race for critical energy transition minerals*. <https://www.spglobal.com/sustainable1/en/insights/special-editorial/community-matters-mining-local-engagement-and-the-race-for-critical-energy-transition-minerals>
35. Data Centre Watch. (2025). *Q2 2025 report*. <https://www.datacenterwatch.org/q22025>
36. Princeton Digital Group. (2025). *Princeton Digital Group partners with Flexidao and Tata Power Renewables to pioneer time-matched carbon-free energy at Mumbai data center campus*. <https://princetonodg.com/newsroom/princeton-digital-group-partners-with-flexidao-and-tata-power-renewables-to-pioneer-time-matched-carbon-free-energy-at-mumbai-data-center-campus/>
37. Princeton Digital Group. (2025, September 10). *Princeton Digital Group and Pertamina Patra Niaga forge landmark HVO partnership pioneering sustainable fuel use in Indonesia to advance climate and community well-being*. <https://princetonodg.com/newsroom/princeton-digital-group-and-pertamina-patra-niaga-forge-landmark-hvo-partnership-pioneering-sustainable-fuel-use-in-indonesia-to-advance-climate-and-community-well-being/>
38. Microsoft. (2026). *Building community-first AI infrastructure*. <https://blogs.microsoft.com/on-the-issues/2026/01/13/community-first-ai-infrastructure/>



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org