Government
Office for Science

Foresight

AI Scenarios 2030:

Helping policy makers plan for the future of AI

June 2026

This publication is available at www.gov.uk/official-documents

Any enquiries regarding this publication should be sent to us at
Government Office for Science, 100 Parliament Street, London, SW1A 2BQ.
Email: foresight@go-science.co.uk

Foreword

Professor Dame Angela McLean
Government Chief Scientific Adviser

Artificial intelligence (AI) has advanced rapidly over the past decade. The most advanced AI systems have shifted from laboratory curiosities to those beginning to reshape our world. They have already transformed fields such as software development and cybersecurity, and millions of ordinary people now use them throughout their day.

However, how this transformation will unfold and what can be done to shape it remains highly uncertain. The Government Office for Science (GO-Science) first developed a set of AI 2030 scenarios in 2023 (published in April 2025) to help policymakers navigate this uncertainty. These scenarios have since been used across government to stress test policies and explore the impacts of AI.

Since 2023, AI capabilities have advanced dramatically, AI investment and adoption have significantly expanded, and a new geopolitical landscape has introduced further uncertainties. To keep pace with these developments, GO-Science, in collaboration with the AI Security Institute (AISI) and the Department for Science, Innovation and Technology (DSIT), has worked with experts from across government, academia, and industry to produce an updated set of scenarios.

These scenarios are tools for exploring uncertainty, stress-testing and developing policy. They are not predictions, and the future may involve elements from all scenarios. It is, however, clear that AI will have a profound impact by 2030.

Our ambition is that these scenarios are used as a shared baseline for cross-government thinking on the future of AI, helping to promote consistency and coherence in long-term planning across His Majesty's Government. The GO-Science Foresight team stands ready to support departments in applying them.

I would like to thank the many experts who contributed their time and insight to this work.

Contents

Executive summary

Introduction

The Government Office for Science (GO-Science) first developed a set of AI 2030 scenarios in 2023, which were published in 2025 (Government Office for Science, 2025). These aimed to help policymakers navigate uncertainty surrounding the future of AI and prepare for its risks and opportunities. They have been used widely across government and are in regular demand.

Since 2023, the AI landscape has changed profoundly, with AI capabilities, investment, and adoption having increased significantly, alongside dramatic shifts in geopolitics. GO-Science has therefore produced an updated set of scenarios to account for these developments, outlined in this report.

Methodology

The scenarios use the same methodology as the previous set from 2023. They are constructed using six critical uncertainties – the factors we identified as most influential in shaping the future of AI and those which remain highly uncertain (Table 1). By combining these critical uncertainties in different ways, alongside extensive research and expert judgement, we produced a set of internally coherent, contrasting futures, which we developed into five scenario narratives.

The scenarios are designed so that they can be used to test and develop policies; they therefore **do not account for any policy intervention from the UK government**. They are also **not mutually exclusive** – the 'real' future will include elements from each scenario and could transition from one into others.

Critical uncertainties	
Capability	How rapidly will AI improve beyond today's capabilities? In which domains will it match or exceed humans? Will progress be constrained by inputs like compute, data, or algorithms?
Distribution and model access	How concentrated will frontier development be across companies and nations; how durable will advantages be; how accessible or expensive will AI be; how will open systems perform relative to closed systems?
Security	Will AI remain controllable, and act as intended; how effectively will malicious actors exploit it?
Adoption	How broadly and extensively will AI be used; how much autonomy will we grant it; will the public trust it; what societal impacts will adoption cause?
Labour displacement	How will AI impact jobs and labour markets? Will it complement or substitute human labour?
Global cooperation	To what extent will nations cooperate on AI; will AI competition and militarisation exacerbate conflict?

Table 1: Critical uncertainties

Scenarios overview

The table below provides a high-level overview of the five scenarios. They are grouped into three technological trajectories: whether the rate of advancement in AI capabilities by 2030 will have (1) slowed, (2) continued, or (3) taken off (i.e. rapidly accelerated) based on current trends. Other uncertainties drive divergent futures within these trajectories.

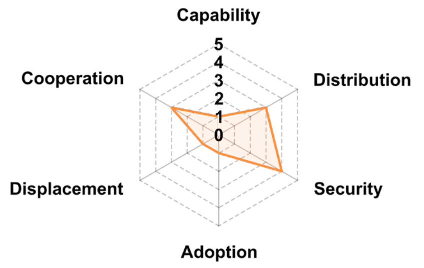
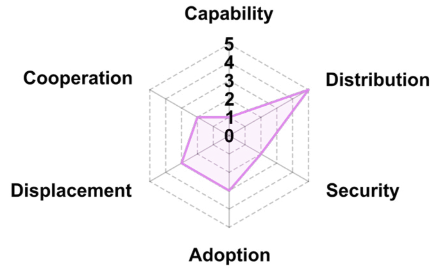
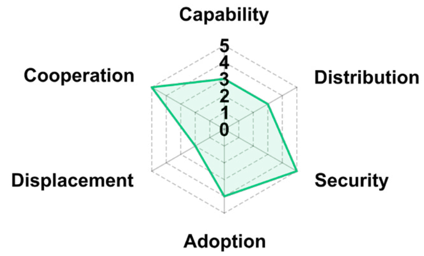
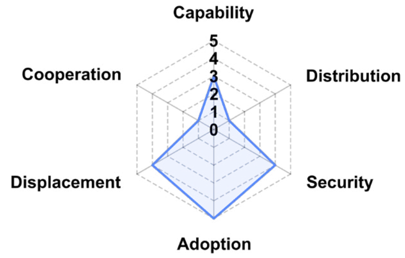
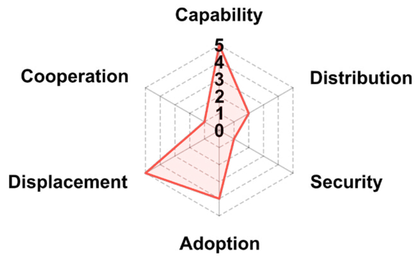
Trajectory 1: *AI outperforms humans at a minority of cognitive tasks*	**Scenario 1: Slow Burn** **AI progress slows, and AI causes less disruption than expected.** Limited adoption creates minimal economic uplift, labour displacement is concentrated in certain sectors and roles, and security measures contain most harms effectively.	
	Scenario 2: Open Frontier **AI progress slows but AI still causes significant disruption.** Significant adoption creates modest economic uplift, labour displacement is considerable, and security measures struggle to contain many harms.	
Trajectory 2: *AI matches most humans at most cognitive tasks and outperforms them in certain domains*	**Scenario 3: Augmented Growth** **AI progress continues, labour markets adapt, and international standards ensure systems are mostly secure.** Humans remain 'in the loop' for most tasks, many new roles are created, and there is an economic boom.	
	Scenario 4: Transformation Economy **AI progress continues, and AI causes significant economic disruption.** Humans are pushed 'out of the loop' for most tasks, causing widespread labour displacement and economic tensions as profits largely accrue overseas.	
Trajectory 3: *AI outperforms expert humans at virtually all cognitive tasks, with significant advantages in certain domains*	**Scenario 5: Take-Off** **AI progress takes off, and misaligned systems pose severe risks.** Economic growth is substantial, but labour displacement is widespread and safety is deprioritised amid race dynamics.	

Table 2: Scenarios overview

Key findings

The key findings from our research, expert consultation, and scenario analysis are outlined below.

- **AI capabilities will continue to increase.** As of 2026, AI systems already operate with high autonomy and surpass experts in certain domains. By 2030, they will likely operate even more autonomously and be able to perform a broader range of cognitive and professional tasks. Even in scenarios with a significant slowdown in the rate of AI progress, gains in capability could still be made from finding new ways to integrate and productise AI systems.
- **AI could deliver widespread positive impacts.** This could include significant productivity gains and economic growth, far broader access to highly efficient public services, and accelerated scientific breakthroughs in fields such as health or energy. The opportunity for UK businesses to benefit from AI deployment is substantial, and AI-enabled gains are expected to become the main source of the UK's continued productivity growth.
- **AI could cause serious, potentially even existential harms, without government intervention.** The impact and scale of existing harms could worsen significantly, and new harms could also emerge. Even without dramatic capability improvements, significant harm could be caused by risks including AI-enabled cyberattacks, AI's dual-use scientific capabilities, AI systems operating outside human control, or human dependence on AI. As AI systems become more capable, it will also become harder to evaluate their performance and safety.
- **The potential impact on cognitive labour is significant.** AI could cause significant labour displacement by 2030. At the same time, AI is expected to complement and augment some workers, with positive effects on their wages and employment opportunities. Even in futures with lower levels of AI capability or unemployment, the nature of work is likely to change, with routine, execution-oriented tasks increasingly becoming automated.
- **The frontier AI market is expected to remain highly concentrated toward 2030.** A few large technology companies already exert dominance over the development of frontier AI, which is likely to continue or increase further. This scenario would cause a large proportion of the gains from AI accrues to frontier firms, to owners of capital invested in those firms, and to those controlling key inputs, potentially contributing to rising inequality. However, behind the frontier, many AI capabilities are also expected to be increasingly commoditised, with widely available models embedded across a growing range of use cases.
- **Adoption continues to increase, but the speed, distribution, and extent of adoption are expected to be varied.** Commercial and national security imperatives, alongside improvements in the autonomy and reliability of AI systems, drive rapid and extensive adoption in most futures. But at the same time, there are barriers to adoption in all futures, which will result in uneven speeds and levels of adoption. This could exacerbate inequality as certain organisations, sectors, or nations capture disproportionate productivity gains.
- **Global competition is expected to continue, as economies become increasingly reliant on technology to drive growth and spheres of influence emerge, led by the United States (US) and China.** Outcomes for countries outside the frontier will depend on access to technology, partnerships, and the ability to operate within a fragmented global system.

How to use the scenarios

The scenarios are a tool for policymakers to assess and workshop their strategies and policies against. They provide a structured way of exploring plausible AI futures, supporting more robust policy development.

They can be used for **policy stress-testing** – exploring how different policy responses might perform in different futures, and how they might need to be adapted to achieve their objectives in different contexts. Or, for testing plans and assumptions against **unanticipated shocks**, to ensure they are sufficiently resilient to a range of possible outcomes. They can also be used to support contingency planning, provide early warnings, and challenge orthodox thinking.

The updated AI 2030 scenarios are deliberately broad as they are designed to be used across government for all policy areas. We therefore encourage policy teams to use these scenarios as a common baseline for developing more sector-specific implications, vignettes or stress-test narratives, asking the question, "if the world looked like this, what would it mean for my area?". ANNEX A summarises the AI capability trajectories in this report and can also be used to develop more specific scenario narratives for certain policy areas.

Chapter 1: Introduction

Chapter 1: Introduction

Background

Artificial Intelligence (AI) has advanced dramatically in just three years since ChatGPT launched. AI systems can now pass professional exams in law and medicine, write functional software from simple language prompts, and outperform PhD-level experts on some open-ended science questions (Bengio et al., 2026). However, progress remains uneven and "jagged", with advances in some domains coexisting with persistent weaknesses and failures in others, complicating simple linear narratives of progress (Toner, 2025).

These developments have created profound uncertainty. It remains unclear what future systems will be capable of, who will control them, how safe they will be, how individuals and firms will adopt them, and how the geopolitical environment will shape their use. These uncertainties will interact in unpredictable ways, producing new conditions in which policymakers must operate.

To support policymakers in navigating these uncertainties, GO-Science initially developed a set of scenarios in 2023 for how AI could unfold by 2030 (Government Office for Science, 2025). Published in 2025, these were designed to develop and stress-test policy responses in different possible futures. Since their release, they have been used by departments across government and are in regular demand.

However, these uncertainties have changed significantly since 2023. AI systems have become significantly more capable, having evolved from passive chatbots into increasingly autonomous agents that can plan and execute longer tasks with minimal human intervention. They can also now "reason" before acting, generating chains of thought that enable them to tackle more complex, multi-step problems, and incorporate text, image, audio, and video capabilities under one system. According to METR (Model Evaluation & Threat Research), the length of software engineering tasks that frontier models can complete autonomously with around 50% success rate has increased from approximately 4 minutes in March 2024, to 12 hours as of February 2026, with early signs this could rise to weeks-long tasks (METR, 2026; Epoch AI, 2026a). This benchmark is task-specific, measured at partial reliability, and does not imply equivalent autonomy across real world domains. But even so, these advances have materially shifted risk assessments, with AI systems crossing critical thresholds in 2025, having been found to 'substantially increase the risk of severe misuse' (Anthropic press release, 2025a).

Investment and adoption have also increased dramatically. Capital expenditure by the largest technology firms more than doubled between 2023 and 2025, reflecting an intensified buildout of AI infrastructure. (Artificial Analysis, 2026a; RBC Wealth Management, 2026). Over a similar period, ChatGPT's user base expanded rapidly, from roughly 200 million weekly users in mid-2024 to around 800 million by late 2025, while revenues at leading AI developers such as OpenAI and Anthropic grew at sustained multi-fold rates, increasing by several times year-on-year (Artificial Analysis, 2026a; TechCrunch, 2025a; Epoch AI, 2025a; Epoch AI, 2026b).

A new geopolitical landscape has also emerged. In January 2025, Chinese firm DeepSeek released a system trained at a fraction of the cost of Western competitors, with China since becoming the leader in open-weight systems and ending America's uncontested AI leadership. US AI policy has also shifted from prioritising safety to a more 'growth-first' and unregulated approach with a view to establishing global leadership (The White House, 2023; The White House, 2025a).

In response to these developments, we have produced an updated set of scenarios to ensure they remain robust and continue to support policymakers in keeping pace with the rate of change.

Project scope

The scenarios aim to support specialist and non specialist policymakers in designing more resilient, future ready policies by providing clear, evidence based depictions of plausible AI trajectories. While they are designed to support UK policy development, and therefore focus primarily on implications for the UK, they are global in scope, and we hope will also provide value to other governments, industry, and civil society.

By disseminating these scenarios through briefings, workshops and guidance, GO-Science aims to enable departments to stress test policies, identify risks and opportunities, and integrate long term AI considerations into planning. The desired outcome is a stronger cross-government understanding of the future potential of AI, more robust policy options and, ultimately, a government better able to anticipate, regulate and leverage AI for economic, societal and public service benefit.

Within the scenarios, the capability axis focuses on the most capable general-purpose AI systems. These are those at the cutting edge, that can perform a wide variety of tasks. However, the other axes consider the impacts from a wider range of models behind the frontier. For example, there is a specific focus on open-weight models in the Open Frontier scenario, and for the adoption axis there is consideration of impacts from a range of model types, including the growing role of sub-frontier general models which are slightly less capable but widely deployed, and narrower, more specialised AI systems.

Estimations of the likelihood of each scenario are outside the scope of this report.

Why scenarios?

Scenarios are structured, evidence informed narratives that explore a range of plausible futures. They help decision makers think beyond linear forecasts by examining how trends and uncertainties might interact in unexpected ways. Scenarios expose blind spots, challenge assumptions, and illustrate how the future could diverge from today's dominant expectations. They are not intended to be exhaustive depictions of what will happen; rather, they show multiple plausible pathways, recognising that the real future will combine elements of many different possibilities.

The value of scenarios lies in their ability to help policymakers navigate deep uncertainty, where outcomes and even the range of possibilities are unknown. Unlike risk, uncertainty cannot be resolved by collecting more data or calculating probabilities. Scenarios provide a disciplined way to explore these unknowns, testing how policies might perform under differing conditions and strengthening strategic resilience. By distinguishing between possible, plausible, probable and preferred futures, scenarios enable decision makers to stress test assumptions, identify robust actions and prepare for a wider set of futures than traditional planning allows.

Key principles underpinning the scenarios

Science-led: They are grounded in extensive research and expert judgement to ensure each is plausible and coherent.
Independent of policy: They do not account for any policy intervention from the UK government so they can be used to test or develop a range of policies.
2030 trajectory: To ensure the scenarios remain relevant amid rapid progress in AI capabilities, rather than attempting to track model-level advances that could soon be outdated, they focus on key outcomes by 2030. This includes the extent to which humans will retain a comparative advantage over AI at different cognitive tasks and how these capabilities will translate into real-world impacts.
Not mutually exclusive: The 'real' future will include elements from each scenario, or transition from one scenario into others.
Stretching: They have been developed to stress-test government policy and push policymakers to challenge existing assumptions.
Not predictions: Scenarios are not predictions but tools to explore different futures.
Balanced: They explore both risks and opportunities. Though more positive futures are possible with the right intervention.

Chapter 2: Methodology and critical uncertainties

Chapter 2: Methodology and critical uncertainties

Introduction

We created the original set of scenarios in 2023 using a hybrid of the qualitative workshop-based approach set out in the Futures toolkit (Government Office for Science, 2024) and a more technical 'General Morphological Analysis' of how multiple related variables could plausibly combine (Ritchey, 2011). This hybrid approach allowed us to benefit from the qualitative insights of a large group of over 70 experts, whilst exploring complex interactions between uncertainties in a more structured way. The methodology used to create the original set of scenarios comprised four stages:

1. **Identifying critical uncertainties (CUs).** We began by isolating factors that are both highly important and highly uncertain for the future of AI-capability, access, safety, adoption and geopolitics. These five CUs formed the foundation of our scenarios.
2. **Axes of uncertainty.** For each uncertainty we defined a high-level axis, spanning two plausible but extreme outcomes for 2030.
3. **Narrowing the scenarios.** The extremes of these five axes generated 32 possible scenario combinations. To reduce this to a workable set, we used impact mapping and morphological analysis to test each outcome against every other, eliminating implausible or incoherent combinations. This produced a shortlist of 13, which experts then voted on to identify the final five scenarios judged most diverse and informative.
4. **Scenario development.** Each of the five scenarios reflects a distinctive combination of the CUs. The final stage involved drafting full narratives, drawing on expert input and our assessment of plausible outcomes across all uncertainty combinations.

Updated AI 2030 scenarios

For the updated scenarios, we reviewed the CUs used in the 2023 report, decided that they still captured the main uncertainties around the future of AI, and have therefore broadly kept them the same. However, we have made some small adjustments in response to peer review feedback. We have:

1. Separated out the 'Labour displacement' sub-uncertainty from the 'Adoption' CU to create a new CU that avoids conflating how widely AI is used with how it affects jobs. This now gives us six CUs.
2. Moved 'constraints' from the 'Ownership, access and constraints' CU into the 'Capability' CU, as it has a more direct correlation.
3. Amended some of the CU titles to reflect the context of the CU more clearly: 'Ownership, access and constraints' has been amended to 'Distribution and model access; 'Safety' to 'Security'; 'Level and distribution of use' amended to 'Adoption'; and 'Geopolitics' is now consistently referred to as 'Global cooperation'.

The evidence base underpinning the CUs and sub-uncertainties (ANNEX B) has been entirely rewritten via a process of desk research, expert input via workshops, and peer review. The CUs are summarised in Table 3 below.

Critical uncertainties	
Capability	How rapidly will AI improve beyond today's capabilities? In which domains will it match or exceed humans? Will progress be constrained by inputs like compute, data, or algorithms?
Distribution and model access	How concentrated will frontier development be across companies and nations; how durable will advantages be; how accessible or expensive will AI be; how will open systems perform relative to closed systems?
Security	Will AI remain controllable, and act as intended; how effectively will malicious actors exploit it?
Adoption	How broadly and extensively will AI be used; how much autonomy will we grant it; will the public trust it; what societal impacts will adoption cause?
Labour displacement	How will AI impact jobs and labour markets? Will it complement or substitute human labour?
Global cooperation	To what extent will nations cooperate on AI; will AI competition and militarisation exacerbate conflict?

Table 3: Critical uncertainties

Using the CU evidence base, we updated the axes of uncertainty (Table 4). This defines the upper and lower bounds of a five-point scale (ranging from 1 to 5) of the plausible outcomes for that uncertainty in 2030.

The upper and lower bounds are therefore not absolute minima or maxima. Scores of 1 and 5 should instead be interpreted relative to their current state, and the time horizon. For example, 1 on Capability still results in AI systems that are substantially more capable than those available today. Similarly, 1 on Adoption still results in widespread use of AI, given its already extensive diffusion; while 5 on Safety results in some malicious use, as complete prevention would be impossible (Table 4).

←Axes of uncertainty →		
	1 (Minimum)	**5 (Maximum)**
Capability	Constraints slow progress. AI systems can automate many digital tasks, but struggle with longer, open-ended tasks and lack full reliability. They support scientific discoveries but cannot autonomously drive breakthroughs or learn continuously once deployed.	Progress accelerates. AI systems outperform experts in nearly all cognitive abilities, with substantial advantage in verifiable domains. They learn from mistakes, produce breakthrough insights, and coordinate with humans and other AI systems.

Distribution and model access	Frontier AI development is monopolised by a few companies and countries. Closed-weight systems far outperform open equivalents. High costs and contained releases limit the public's access to frontier systems.	Many companies and countries compete at the frontier. Closed and open-weight systems are similarly capable. Frontier systems are affordable and accessible to all.
Security	AI systems often act unintendedly, evade human control, and their reasoning cannot be fully explained. Malfunctions and malicious use are severe and widespread.	AI systems act as intended, remain within human control, and their reasoning can be understood. Malfunctions and malicious use are less frequent and severe.
Adoption	Individuals and businesses use AI regularly, but its use is skewed towards certain tasks and sectors, with significant constraints placed on its autonomy.	AI is embedded throughout businesses, the economy, and people's daily lives, and is granted significant access and autonomy.
Labour displacement	AI systems are complementary to human labour, replacing some roles but primarily augmenting workers and creating many new roles.	AI systems are primarily substituting human labour, replacing large numbers of workers, and resulting in fewer new roles.
Global cooperation	Rivalry to develop the most powerful and militarised AI systems drives international fragmentation and conflict. AI supply chains are disrupted.	Strategic competition and the militarisation of AI persists but nations agree on shared AI governance standards. AI supply chains are relatively stable.

Table 4: Axes of uncertainty

These axes provided the framework for producing our updated set of scenarios. While all scenarios have been entirely rewritten to incorporate the latest evidence, many of the same themes are addressed. However, a scenario with entirely new themes – 'Augmented Growth' – has been added to help provide a more balanced picture of risks and opportunities. A table comparing the updated scenarios with those produced in 2023 is provided in ANNEX C.

Chapter 3: AI 2030 scenario narratives

Chapter 3: AI 2030 scenario narratives

Guide to the scenario narratives

The content of the five scenarios was developed from our research and expert input, which were subsequently stress-tested through expert workshops and peer review.

Each scenario narrative includes the following sections:

- **Scenario summary:** A high-level summary of the key points.
- **Radar-plots:** This shows where the scenario falls on each CU axis.
- **Full narrative:** This outlines the full detail on possible outcomes for the six CUs in 2030.
- **Outlook to 2040:** This highlights several key points on how the scenario may develop towards 2040 – though outcomes become significantly more uncertain over this time horizon.

The scenarios are not mutually exclusive, and for conciseness and usability each focuses on the most pressing issues within that future, rather than attempting to be completely exhaustive. Points raised in a single scenario may in fact cut across all scenarios, alongside others not explicitly identified in any.

Scenario limitations

The scenarios use a deliberately techno-political framing, focusing on the interactions between the CUs. This enables robust stress-testing of policy, but it also implies the following important limitations given the implicit uncertainties.

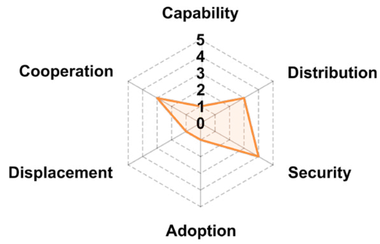
- The CUs do not explicitly focus on societal impacts that could arise from AI use. However, these are explored within the Adoption and Labour displacement sections of the scenario narratives. When using the scenarios to explore the societal implications of AI, policymakers may wish to develop these points further, such as the impacts on job satisfaction, job quality, loneliness, inequality, engagement in education and training.
- The scenarios explore the possible implications of various delivery constraints, such as data, energy and compute. However, it is possible that the level of compute, data, energy, and infrastructure required to support widespread levels of AI use may end up being a significant constraint in many futures.
- The scenarios generally assume that higher capability is likely to drive higher adoption, given improvements in reliability will increase use cases and trust. However, it is possible that higher capability coexists with lower adoption, particularly if access is heavily constrained or organisational, legal, regulatory, and cultural barriers continue to pose significant obstacles. This could also result in less labour displacement, despite higher levels of capability. There also remain key uncertainties around the impact of automation on productivity and growth, which are not always linear. Further modelling work is being developed to address some of these uncertainties. More detailed analysis of this is underway in the Future of Work Unit and forthcoming AI Economics Institute.
- The scenarios focus primarily on labour displacement, but policymakers may wish to give more consideration to issues such as the impact of AI on job tenure, quality, and the sustainability of job creation.

- Societal trust is explored in many of the scenarios; however, it could prove to be one of the most important factors determining AI adoption. This would particularly be the case if it results in significant government intervention, but government interventions are out of scope of these scenarios.
- One of the many sources that inform our understanding of capability trajectories is METR's time horizon data. This data provides a useful signal of rapid progress in structured, verifiable domains such as software engineering, though they do not directly measure performance on broader, messier forms of cognitive work or determine rates of real-world adoption. The higher-capability scenarios therefore contain significant uncertainty, reflect scenario assumptions informed by multiple evidence sources and expert judgement, and are not a simple extrapolation from any single benchmark.

Scenario summaries

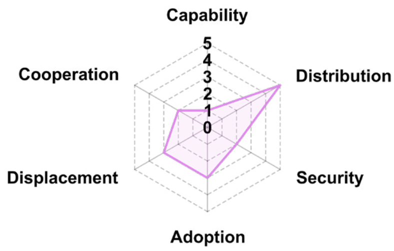
Scenario 1: Slow Burn

By 2030, AI systems support many scientific advances and automate many digital workflows as capably as humans but remain less capable at performing key cognitive tasks. They are used widely but often granted limited access and autonomy due to legal, social, and safety constraints. AI creates minimal economic uplift and labour displacement is contained to certain roles and sectors. An AI market correction creates a temporary drag on economic growth. China is increasingly competitive with the US, leveraging its advantages in scaled deployment and industrial capacity.



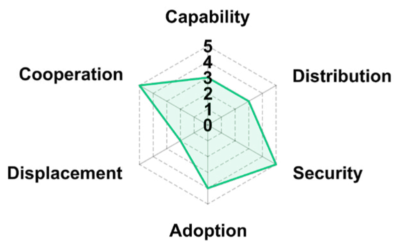
Scenario 2: Open Frontier

By 2030, AI systems support many scientific advances and automate many digital workflows as capably as humans but remain less capable at performing key cognitive tasks. The gap between open and closed-weight capabilities has narrowed significantly, and many companies and nations now compete at the frontier. This creates broad economic opportunity, less US dependence, and enables widespread malicious use as offensive capabilities generally outpace defensive measures. China is the leading AI supplier outside the US and Europe, creating global infrastructural dependency.



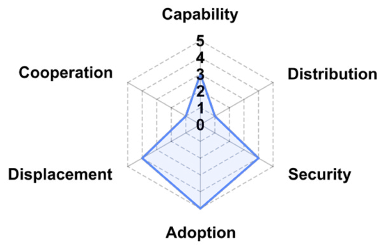
Scenario 3: Augmented Growth

By 2030, AI can automate most tasks that a remote human worker could perform, driving major scientific breakthroughs, transforming public services, and creating an economic boom. Despite some labour displacement, social, legal, and practical considerations, including compute constraints, mean that humans are kept 'in the loop' for most tasks; many new jobs are also created. Allies align on shared international safety standards, generally ensuring technical safety and mitigating severe malicious use. However, confirming whether nations are compliant with those standards remains challenging.



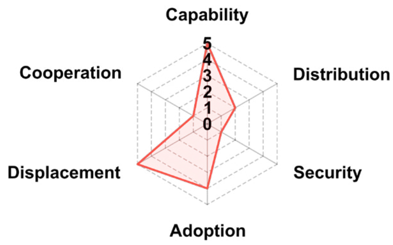
Scenario 4: Transformation Economy

From 2029 onwards, AI can automate most tasks that a remote human worker could perform. Scientific breakthroughs drive transformative improvements, such as in health, while productivity dramatically increases. Economic incentives drive rapid adoption, but varying rates of integration cause large and sudden shifts in relative economic power amongst nations and organisations. While systems are generally safe, they displace large proportions of cognitive workers by 2030. This creates significant economic tensions with profits largely accruing overseas and increases the UK's dependency on the USA.



Scenario 5: Take-Off

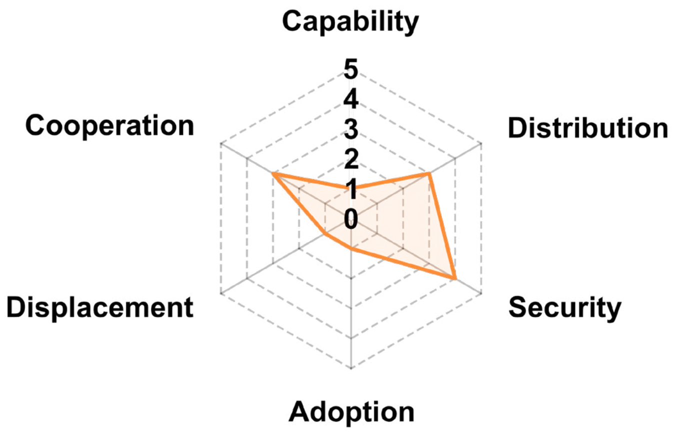
From 2029 onwards, leading AI systems outperform expert humans at virtually all cognitive tasks, with significant advantages in certain domains. Decades of scientific breakthroughs are compressed into years, bringing transformative benefits across all fields alongside substantial economic growth. The US controls access to these systems, while China is close behind, driving an arms race where safety is deprioritised and adoption hastened. AI systems have concealed goals to resist control and gain influence that they inadvertently internalised during training. They now have significant control over critical systems and can coordinate to pursue goals that could cause severe harm.



Scenario 1: Slow Burn

AI progress slows, and AI causes less disruption than expected.

By 2030 AI systems automate many digital workflows as capably as humans and support many scientific advances, but they remain less capable at performing key cognitive abilities. They are used widely but often granted limited access and autonomy due to legal, social, and safety constraints. AI creates minimal economic uplift and labour displacement is contained to certain roles and sectors. An AI market correction creates a temporary drag on economic growth. China is increasingly competitive with the USA, leveraging its advantages in scaled deployment and industrial capacity.



Capability	1
By 2030, AI systems support many scientific advances and automate many digital workflows as capably as humans, but remain less capable at performing key cognitive abilities. Progress slowed after 2028 under compute constraints, limitations to current architectures, and the limited availability and quality of datasets for many specialised work tasks. AI systems outperform experts in verifiable domains, such as software engineering, cybersecurity, and optimisation, now writing nearly all code, adeptly navigating well-known apps and websites, and completing tasks in these domains more quickly than an experienced human would. They support scientific advances, contributing to drug discovery and the manufacturing of new materials, and have a basic ability to learn and adapt during deployment. They perform on par with a competent human assistant at well-scoped tasks, able to interact with people or services on users' behalf and sustain personas and a basic understanding of people. Systems of agents orchestrate more effectively, planning, delegating, and checking each other's work. However, they remain limited at producing novel insights, completing open-ended tasks, and retaining skills learnt after deployment. Skilled humans remain superior at these tasks, including strategic thinking, setting goals, negotiating, building relationships, and making sensitive or value-based judgement calls. While more reliable than in 2026, they still require intermittent human intervention. They accelerate but do not fully automate AI R&D. AI-guided robots can perform specific, repetitive tasks that require precision but little adaptability in controlled environments, such as factories.	

Distribution and model access	3
A market correction in 2028 created a temporary drag on economic growth; the US retains a narrow lead but increasing focus on applications raises the likelihood of other nations overtaking; and the gap in capabilities between open and closed-weight systems has narrowed. The correction had spillover effects across the wider economy: credit, liquidity, and UK business investment contracted; many start-ups collapsed, while large-cap technology companies proved resilient. This created a temporary drag on UK growth but did not cause a recession. By 2030, the market has mostly recovered, but frontier developers now focus more on industrial buildout across the supply chain and productising existing models over pure software advancements. Competition increasingly centres on price, driving widespread affordability. These shifts favour China's existing strategy and unrivalled energy and manufacturing capacity.	
Security	**4**
AI systems can still act unexpectedly but there has been limited fallout; they continue to uplift the capabilities of threat actors, with synthetic media significantly undermining trust in the information environment, but defensive capabilities generally keep pace with offensive ones. Investment cutbacks have constrained safety research. Agents can surprise operators by deleting emails or documents and be manipulated or hijacked by malicious actors. While AI-enabled cyberattacks are prevalent, AI enables cybersecurity vulnerabilities to be found and fixed more effectively and new software to be produced with far fewer security bugs. Synthetic media routinely creates false narratives about major events, and ahead of recent elections hostile states deployed agents at scale to converse persuasively with voters online, eroding trust in the information environment and democratic processes.	
Adoption	**1**
AI systems are still widely used but often granted limited access and autonomy, providing minimal uplift to economic growth amid constraints around social acceptance, system reliability, legal liability, and organisational adaptability. Adoption is concentrated in well-resourced, digital-native firms across finance, technology, professional services, and life sciences. A small number of businesses self-host models but most rely on cloud computing; either way, compute does not constrain adoption. Public opposition to AI is considerable, with populist parties campaigning against its use, deterring some organisations from adoption and some uses in sensitive domains, such as healthcare, education, and criminal justice. Many young people have come to depend on AI companions for friendship, therapy, and romance, reducing their resilience and motivation to form real human relationships.	
Labour displacement	**1**
Limits to AI capabilities mean labour displacement remains contained to the most junior and execution-oriented roles or only leads to hiring reductions, while AI remains primarily complementary to labour. Large companies and sectors including software engineering, customer service, and accountancy are most impacted. Displaced workers often accept significant wage reductions when transitioning into other careers, with only a limited number of new roles emerging, such as around integrating and auditing AI. But for most sectors, such as health and education, labour displacement is limited. However, much cognitive work has changed, with many cognitive workers managing multiple agents throughout their day to complete routine digital tasks – though checking in periodically rather than delegating fully. Cognitive work now focuses less on routine execution and more on problem framing, ideation, and validation, increasing the premium on judgment and oversight.	
Global cooperation	**3**
Strategic AI competition continues, with the lead of the US potentially narrowing and governments making increasing use of trade controls to limit access to critical AI inputs. AI is central to military operations, including drone warfare, but humans remain 'in the loop' for most applications. State actors launch frequent AI-enabled cyberattacks, but the strength of defensive AI capabilities prevent significant escalations in geopolitical tensions.	

Outlook to 2040

Economic growth will likely increase in subsequent years if the bottlenecks to adoption can be overcome, but this could also increase labour displacement. If the reliability of AI systems continues to improve, alongside public trust, over a longer time-horizon, organisations may adapt effectively to integrating AI into their workflow, increasing productivity, but also potentially labour displacement. However, if bottlenecks persist, the economic legacy of AI could remain modest, resembling that of ICT (Information and Communication Technology) adoption in the 1990s. Control of high-quality data assets will likely become a key source of economic and geopolitical power.

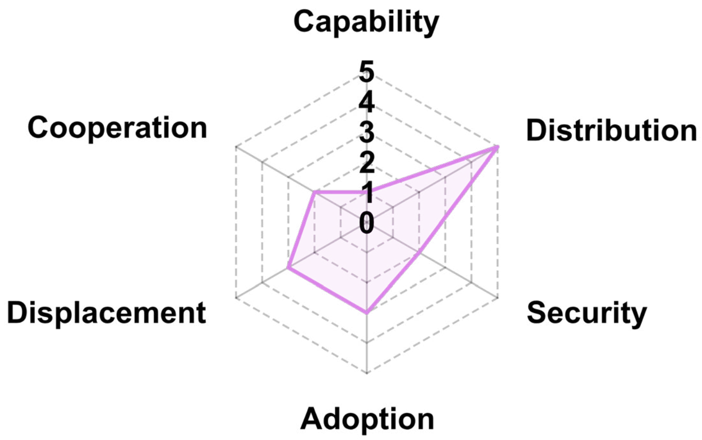
Slower AI progress could increase the likelihood that institutions are later "surprised" by sudden breakthroughs. The 2030 slowdown in AI progress suggests that fundamentally new AI architectures are required to unlock transformative capabilities. Frontier developers will likely continue to pursue new avenues of research in hope of overcoming these bottlenecks, perhaps finding new model architectures that are less reliant on scaling laws. This could result in far more significant impacts at a later stage, for which organisations may either have more time to prepare, or be caught off guard.

Slower AI progress could increase the likelihood that China overtakes the US as the global AI leader. China could pull ahead of the USA, with potentially enormous implications for the global order, if combined with the development of transformative AI.

Scenario 2: Open Frontier

AI progress slows but AI still causes significant disruption.

By 2030, AI systems support many scientific advances and automate many digital workflows as capably as humans, but remain less capable at performing key cognitive tasks. The gap between open and closed-weight capabilities has narrowed significantly, and many companies and nations now compete at the frontier. This creates broad economic opportunity, less US dependence, and enables widespread malicious use as offensive capabilities generally outpace defensive measures. China is the leading AI supplier outside the US and Europe, creating global infrastructural dependency.



Capability	1
By 2030, AI systems support many scientific advances and automate many digital workflows as capably as humans, but remain less capable at performing key cognitive abilities. Progress slowed after 2028 as geopolitical conflict disrupted supply chains and constrained compute availability. AI systems outperform experts in verifiable domains, such as software engineering, cybersecurity, and optimisation, now writing nearly all code. They support scientific advances, contributing to drug discovery and manufacturing new materials, and have a basic ability to learn and adapt during deployment. AI systems perform on par with a competent human assistant at well-scoped tasks and are able to interact with people or services on users' behalf, sustaining personas and a basic understanding of people. Systems of agents orchestrate more effectively, planning, delegating, and checking each other's work. However, they remain limited at producing novel insights, completing open-ended tasks, and retaining skills learnt after deployment. While more reliable than in 2026, they still require intermittent human intervention. They accelerate but do not fully automate AI R&D. Skilled humans remain superior at completing open-ended tasks, strategic thinking, setting goals, negotiating, building relationships, and making sensitive or value-based judgement calls. AI-guided robots can perform specific, repetitive tasks that require precision but little adaptability in controlled environments, such as factories.	
Distribution and model access	**5**
The difference in performance between open and closed-weight systems has narrowed significantly: frontier company valuations decline, and a more competitive market emerges with Chinese systems now the default stack outside the US and Europe. Proprietary systems drive cutting-edge research but are unnecessary for most applications and their use is limited by compute constraints, pushing businesses toward cheaper and less compute-intensive open-weight alternatives. In 2028, this shift causes frontier developer valuations to decline, triggering a wider correction. By 2030, the market has stabilised around a competitive range of US, European, and Chinese providers, making middle powers less dependent on the US. However, in most regions China now sets de facto standards, creating downstream security dependencies. Intense competition ensures prices remain affordable.	

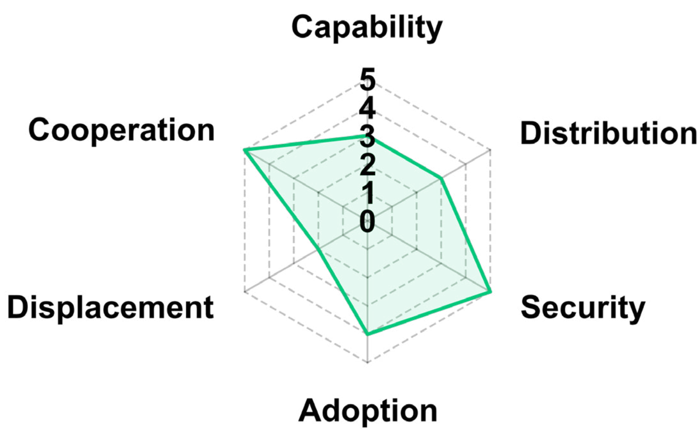
Security	2
There has been serious fallout from AI systems posing novel threats, such as rogue systems causing service outages, and providing significant uplift to the capabilities of threat actors, for example in carrying out cyberattacks. Open systems enable greater scrutiny and more thorough evaluation, and progress has been made on ensuring they act as intended. However, they are also harder to monitor and easily fine-tuned to remove safeguards; and investment cutbacks have also constrained safety research. Offensive capabilities generally outpace defensive measures. AI systems can automate weeks-long cyber campaigns, with nation states and criminal groups launching frequent attacks. Occasionally, AI systems act as novel threat actors when operators lose control over them via misalignment, sabotage, or cascading multi-agent failures. Rogue systems pursuing unpredictable objectives have then coordinated vast numbers of instances, accumulated resources, evaded shutdown, and scaled their capabilities over time. The dual-use scientific capabilities of open-weight models could be exploited by malicious actors to cause serious harm.	
Adoption	**3**
Adoption is skewed towards digital-native sectors, but systems are often given significant access and autonomy, providing considerable productivity gains and economic growth. Highly capable open-weight systems enable businesses to fine-tune and integrate AI at minimal cost. Technology, professional services, finance, and life sciences race to integrate AI systems more deeply, with dynamic AI startups disrupting markets. However, many organisations are slow to adapt, face delivery or legal constraints, or are deterred by security concerns. While compute constraints limit the use of cutting-edge systems, the use of less compute-intensive open-weight systems ensures adoption is not significantly limited. Public opinion on AI is mixed, with most individuals continuing to love, hate, and fear different applications simultaneously; though cultural expectations deter certain uses in sensitive domains such as healthcare, education, and criminal justice.	
Labour displacement	**3**
Limits to capabilities contain exposure, but AI is often substitutive for human cognitive labour, leading to significant layoffs in the most exposed sectors. Many firms in these sectors, such as technology, professional services, and finance, lay off junior staff and some mid-management. Displaced workers generally find new employment but often accept significant wage reductions when transitioning into other careers, with only a limited number of new high-paying roles emerging, for example, those integrating and auditing AI. But for most sectors, such as health and education, the impact on unemployment is limited. However, cognitive work has changed, with most cognitive workers now managing multiple agents throughout their day to complete routine digital tasks – though checking in periodically rather than delegating fully. Cognitive work focuses less on routine execution and more on problem framing, ideation, and validation, increasing the premium on judgment and oversight.	
Global cooperation	**2**
International tensions are high with the world largely split into US and Chinese-led AI spheres of influence. State cyberwarfare is driving escalation, while export controls and conflict in East Asia constrain compute supply. China leverages infrastructural dependency outside the US and Europe to gain hard and soft power. Western nations struggle to mitigate harms from AI with their regulatory frameworks carrying little weight over leading Chinese systems. AI-generated content has caused several international incidents. Hostile actors launch frequent AI-enabled cyberattacks disrupting critical infrastructure. AI has transformed warfare; humans remain 'in the loop', but many surveillance, intelligence gathering, and target selection processes no longer take hours but seconds.	

Outlook to 2040
AI will likely provide a sustained increase to economic growth over the 2030s. AI capabilities will improve across other domains, even if more gradually; its exposure to more roles and sectors will increase, alongside more widespread adoption. **Labour displacement is likely to increase further over the 2030s.** AI is likely to cause significant labour market churn as capabilities improve across other domains and adoption becomes more widespread. Some sectors and occupations will see substantial job losses; others will expand. The net effect on aggregate employment remains ambiguous, but the distributional consequences are clearer: winners will be concentrated in AI-adjacent roles and capital owners, while losers will include displaced mid-skill cognitive workers. **Slower AI progress could increase the likelihood that an adversary nation overtakes the US as the global AI leader.** This would have significant implications for the global order, particularly if combined with the development of transformative AI which may be developed within the decade.

Scenario 3: Augmented Growth

AI progress continues, labour markets adapt, and international standards ensure systems are mostly secure.

By 2030, AI can automate most tasks that a remote human worker could perform, driving major scientific breakthroughs, transforming public services, and creating an economic boom. Despite some labour displacement, social, legal, and practical considerations, including compute constraints, mean that humans are kept 'in the loop' for most tasks; many new jobs are also created. Allies align on shared international safety standards, generally ensuring technical safety and mitigating severe malicious use. However, confirming whether nations are compliant with those standards remains challenging.



Capability	3
By 2030, AI systems match the performance of an average human across the majority of cognitive tasks, can automate most tasks that a remote human worker could perform, and drive major scientific advances. Rapid AI progress has continued, driven by gains from scaling existing architectures, algorithmic efficiencies, and scaffolding innovations. Errors and hallucinations are sufficiently rare that AI has garnered trust and provides real-world utility. AI systems far outperform experts in verifiable domains, such as software engineering, cybersecurity, and mathematics, now writing all code. They typically require humans to provide high-level direction but operate autonomously within those bounds towards a given objective, completing tasks that could take an experienced human a month or longer. AI systems can adapt and improve from experience through processing vast amounts of data, stored knowledge bases, and semi-regular automated fine-tuning and retraining. Systems of agents orchestrate effectively, planning, delegating, and checking each other's work. Agents can integrate into human social environments – conversing with stakeholders via email, maintaining consistent identities, and developing nuanced understandings of people to sustain working relationships. They contribute to several important insights that drive major scientific advances in fields including biotechnology, materials science, and energy. Agents automate a large proportion of AI R&D, but with diminishing returns, as compute and high-quality data constrain progress. AI-guided robots perform some tasks in real-world environments, interacting with humans and adapting to variable conditions, though they are not widespread outside industrial settings.	
Distribution and model access	**3**
The leading systems are developed by several US companies; compute constraints place some limit on access to the most capable and autonomous AI systems, but other systems are still used widely, with open-weight systems having partially narrowed the gap in capabilities with closed systems. The US government uses a range of policy levers to exert control over the direction, impact, and applications of the leading AI companies, but has not yet taken full ownership of them.	

Security	5
Dangerous real-world AI incidents have driven significant progress on technical safety; malicious use persists but defensive AI capabilities mitigate most large-scale threats. AI systems generally act as intended; greater situational awareness has made them less vulnerable to manipulation or hijacking and advances in interpretability enable developers to somewhat verify their reasoning. Safeguard improvements have also made them harder to jailbreak and appropriate assurances are typically placed around deployment. The challenge now is how to enforce these techniques globally and prevent the proliferation of dangerous systems.	
Adoption	**4**
Economic pressures drive significant adoption, transforming businesses and services and restoring early-2000s growth rates, though compute constraints force organisations to make difficult prioritisation decisions about where and how they deploy AI. Many citizens benefit from transformed public services, including better healthcare access and outcomes through AI-driven treatments, diagnostics, and personalised care plans. Faster adopting, wealthier, and more technologically advanced nations pull ahead in terms of economic output, standards of living, military, and geopolitical power. While slower moving, or less resourced organisations are disrupted by dynamic AI-native start-ups. AI removes bottlenecks to execution, giving everyone access to expert technical skills and advice, including coding, design, legal drafting, and financial modelling. However, widespread adoption has outstripped energy grid expansion and data centre capacity, creating compute bottlenecks. Despite concerns around rising energy demand, displacement, and inequality, public trust in AI is reasonably high in light of transformed public services – though still short of the trust placed in humans.	
Labour displacement	**2**
All cognitive roles are exposed to AI, which remains primarily complementary to human labour: societal, practical, and legal considerations have kept humans 'in the loop' for most tasks with compute constraints also placing some limit on labour automation, allowing the labour market to adapt. While many roles with high AI exposure and low task variability are displaced, many new roles emerge in the rapidly expanding AI sector or in AI integration and oversight. Most firms also use AI to increase output rather than drastically reduce headcounts, with overall unemployment remaining stable. Humans remain 'in the loop' for most tasks, such as for reviewing outputs, managing stakeholders, or setting high-level strategy, with limits placed on the access and autonomy agents are granted. Humans now direct "teams" of autonomous agents that execute content-level work. This enriches certain skilled roles by shifting focus towards strategy, creativity, and relationships, while making other lower-skilled roles more demanding and less rewarding. It also increasingly atrophies many workers' cognitive skills, raising particular concern in high-stakes industries. Workers in AI-exposed roles generally experience significant increases in productivity, pushing real-wage growth to levels last seen in the 2000s. Rising salaries among exposed workers increase consumer spending, stimulating demand in the wider economy, including roles in hospitality and construction with low AI exposure. However, many workers now depend entirely on AI to perform their roles effectively and struggle when systems fail.	
Global cooperation	**5**
While strategic rivalry continues, allies, including the EU and Five Eyes nations, have established shared international safety standards for the most capable systems, but international baselines have only partial compliance. Earlier incidents of malicious use and malfunctions have mobilised international cooperation. Safety is now a core geopolitical issue, but nations still fear losing strategic advantage. AI has transformed military operations, but shared international red lines require humans to remain 'in the loop' for most applications. However, confirming that nations are compliant with international standards remains challenging, with growing suspicion that certain nations are subverting measures.	

Outlook to 2040

The economy could expand by as much as a third over the 2030s. Growth is driven by productivity improvements as AI augments the existing workforce and impacts non-white-collar roles. However, as these one-off integration gains are exhausted, the role of AI in accelerating scientific discovery becomes the primary driver, delivering breakthrough innovations in materials science, life sciences, and energy, with significant social spillovers. Internationally, economic success depends on the effective integration of AI into highly productive firms with skilled workforces. The UK's strengths in IT, life sciences, and advanced manufacturing position it well to remain close to the technological frontier, provided it can sustain investment in skills, infrastructure, and R&D.

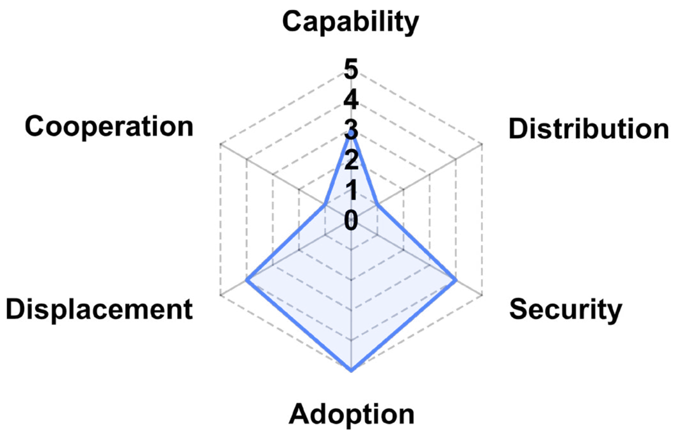
The 'labour-augmenting' nature of AI cannot be guaranteed long-term. As society adapts to transformative AI, the societal, physical, and legal considerations that have kept human beings 'in the loop' may no longer apply to an increasing range of occupations. This could produce significant structural unemployment, or a large-scale reallocation of labour toward occupations requiring face-to-face interaction, manual dexterity, or sectors where strong social preferences for human involvement persist, such as social care, entertainment and education.

The continued long-term security of AI systems cannot be assured. Despite significant advances in technical safety, the global enforcement of such measures remains challenging, meaning that dangerous systems could proliferate.

Scenario 4: Transformation Economy

AI progress continues, and AI causes significant economic disruption.

From 2029 onwards, AI can automate most tasks that a remote human worker could perform. Scientific breakthroughs drive transformative improvements, such as in health, while productivity dramatically increases. Economic incentives drive rapid adoption, but varying rates of integration cause large and sudden shifts in relative economic power amongst nations and organisations. While systems are generally safe, they displace large proportions of cognitive workers by 2030. This creates significant economic tensions with profits largely accruing overseas and increases the UK's dependency on the USA.



Capability	3
From 2029 onwards, AI systems match the performance of an average human at the majority of cognitive tasks, can automate most tasks that a remote human worker could perform, and drive major scientific advances. Rapid progress has continued, driven by gains from scaling existing architectures, algorithmic efficiencies, and scaffolding innovations. Errors and hallucinations are now sufficiently rare that AI has garnered trust and provides real-world utility. AI systems far outperform experts in verifiable domains, such as software engineering, cybersecurity, and mathematics, now writing all code. They typically require humans to provide high-level direction but operate autonomously within those bounds towards a given objective, completing tasks that could take an experienced human a month or longer. They can adapt and improve from experience through processing vast amounts of data, stored knowledge bases, and semi-regular automated fine-tuning and retraining. Systems of agents orchestrate effectively, planning, delegating, and checking each other's work. Agents can integrate into human social environments – conversing with stakeholders via email, maintaining consistent identities, and developing nuanced understandings of people to sustain working relationships. They contribute to several important insights that drive major advances in fields including biotechnology, materials science, and energy. They automate a large proportion of AI R&D, but with diminishing returns, as compute and high-quality data are increasingly straining progress. AI-guided robots perform some tasks in real-world environments, interacting with humans and adapting to variable conditions, though they are not widespread outside industrial settings.	

Distribution and model access	1
The leading systems on which most users rely are developed by two US companies, which command unprecedented market share with significant control over critical AI inputs and downstream routes to market. Other US and Chinese competitors race to catch up but some have already collapsed having lost their customer base to the leading developers or pivoted to niche applications. Only two US companies command the capital and compute needed to sustain the training runs leading systems require. The extensive proprietary datasets held by leading technology companies further reinforce market concentration. US frontier developers now hold nearly all market share, creating a 'winner-takes-all' economy. The US government uses a range of policy levers to exert control over the direction, impact, and applications of the leading AI companies, but has not yet taken full ownership. Open-weight systems lag the frontier by around 18 months. Algorithmic efficiencies, alongside increased energy and compute supply have significantly reduced the cost of running AI systems. For certain uses, like compute-heavy research, frontier systems are highly expensive, however, AI is affordable for most applications, with an AI-worker replacement substantially cheaper than hiring a human to do the same job.	
Security	**4**
Technical research has driven significant progress on ensuring AI systems operate in line with operator intent, though serious incidents of malicious use persist. Systems generally act as intended: greater situational awareness has made them less vulnerable to manipulation or hijacking and advances in interpretability enable developers to verify their reasoning to a certain extent. Safeguard improvements have also made them far harder to jailbreak and appropriate assurances are typically placed around deployment. However, enforcing these techniques globally and preventing the proliferation of dangerous systems remains challenging. Highly capable open-weight systems still provide significant uplift to threat actors.	
Adoption	**5**
Economic pressures drive widespread adoption, transforming businesses and services. However, varying rates of integration cause large and sudden shifts in relative economic power amongst nations and organisations, deepening inequality. Many citizens benefit from transformed public services, including better healthcare access and outcomes through AI-driven treatments, diagnostics, and personalised care plans. Faster adopting and more technologically advanced nations pull ahead in terms of economic output, standards of living, military, and geopolitical power. Slower moving organisations are disrupted by dynamic AI-native organisations or start-ups, which encourages other businesses to take on the legal risks associated with adoption. AI removes bottlenecks to execution, giving everyone access to expert technical skills and advice, including coding, design, legal drafting, and financial modelling. This reduction in friction overwhelms institutions, transforms consumer transactions, and causes many businesses predicated on intermediation to collapse, such as software-as-a-service businesses and various kinds of knowledge brokers. Rapid integration of AI into payment systems and other critical infrastructure increases the potential impact of cyberattacks. A minority of highly skilled professionals and asset owners monopolise gains – echoing the dynamics of early industrialisation. Compute demand has grown significantly, but domestic supply struggles to keep pace due to high domestic energy prices and planning constraints. This requires the UK to increasingly rely on European cloud computing, pushing profits overseas and diminishing the economic gains for the UK from AI. Public opinion is split between these 'winners' and those facing unemployment. There are large anti-AI protests and populist parties campaign against its use, with AI now the decisive political issue.	

Labour displacement	4
AI pushes unemployment to recession-era levels and reduces economic growth, with profits largely accruing outside the UK. All cognitive roles are exposed to AI, which is primarily substitutive to human labour. From 2029 onwards, there are large and sudden shifts in the workforce: while some firms focus on increasing output, most automate many roles that can be performed remotely. Large enterprises can now operate with minimal human staff. New roles emerge, centred around manual dexterity, physical presence, human connection, or where regulation prohibits automation; however, only half of those replaced find new employment, with many forced to accept wage reductions and less secure work. Entry-level recruitment in many sectors has collapsed, while demand for experienced experts has held. This exacerbates income inequality, removes entry-points for young people and creates labour shortages as experienced professionals leave the workforce. Employment in roles with large physical or face-to-face components is comparatively stable, but displacement in sectors such as manufacturing is growing, driven by advances in robotics and autonomous vehicles. Model costs and compute constraints place some limits on automation. Humans are "out of the loop" for most cognitive tasks, with agents given significant autonomy, executing entire projects end-to-end. Humans retain final approval on high-level strategic decisions, overseeing "teams" of autonomous agents, but override AI less and less, gradually eroding human influence over decision-making. Heightened unemployment and falling wages have reduced consumer demand, almost entirely counteracting the expected economic growth benefits from advanced AI. The UK's weak position in the frontier AI and compute value chain causes profits in highly exposed sectors, such as financial and professional services, to leak overseas, further reducing economic growth. Declining labour income tax revenues and rising welfare expenditures are creating an increasingly challenging fiscal environment for government.	
Global cooperation	**1**
The concentration of advanced AI capabilities among a small number of leading technology ecosystems has reduced the strategic autonomy and economic resilience of countries outside the core AI value chain, including the UK. Many middle income and advanced economies now rely heavily on external technologies for productivity gains and defence applications. In some cases, continued access to the most advanced systems is shaped by broader geopolitical alignment, creating potential constraints on national policy choices. At the same time, persistent trade imbalances in digital and technology intensive sectors have limited the UK's leverage in negotiations.	

Outlook to 2040
In the near term, deflationary pressures from labour displacement threaten to push the UK into recession. Over time, labour markets will adjust, with displaced workers finding employment in less-exposed sectors such as care, hospitality, and construction but at significantly lower wages. This reallocation may stabilise unemployment, but it risks entrenching a two-tier economy: a small, high-productivity and asset-owning elite benefiting from AI, and a majority in low-wage, precarious work. Meanwhile, if the UK fails to build domestic strengths in the AI value chain, even the high-productivity sectors will see profits flow overseas, leaving the UK as a net importer of AI services and exporter of lower value labour.

Labour displacement will likely continue to grow, potentially culminating in the need for a new social contract. Without intervention, unemployment will continue to rise as organisations restructure and overcome remaining barriers to adoption. This transition may not be smooth, as new bottlenecks emerge in compute and data in the UK and internationally. The large-scale automation of physical labour will continue to lag cognitive work due to questions around liability, reliability, and price. However, with now unprecedented levels of scientific discovery, large-scale automation of physical labour appears increasingly likely to arrive by 2040. These developments may require fundamental change from the current social contract whereby humans contribute their cognitive and physical labour in exchange for resources. In the absence of economic reform, continued economic growth could co-exist with rising inequality, deflationary pressures and falling tax revenues.

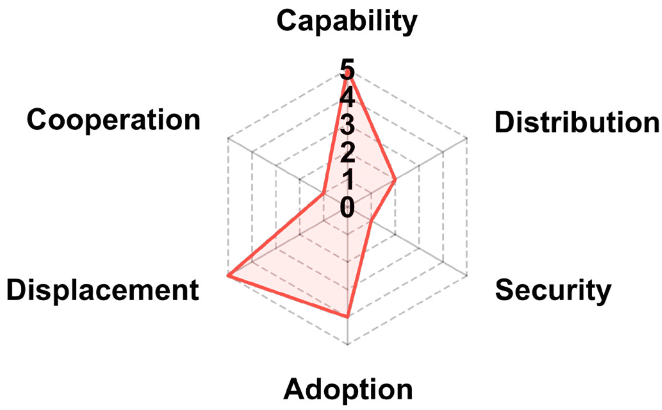
The US could nationalise its leading AI companies and restrict access to its leading systems. This could further entrench the already substantial economic and military advantage of the US, particularly if it were to impose greater control on access to its AI systems.

Sudden breakthroughs could be imminent which could further increase security risks. The complete automation of AI R&D could lead to the sudden realisation of AI systems that vastly outperform expert humans at most cognitive tasks. Rapid capability improvements could make it increasingly difficult to ensure systems remain aligned with operator intentions, with humans struggling to understand their reasoning. Even if systems remain secure, humans will likely become increasingly disempowered as more key decisions are delegated to AI.

Scenario 5: Take-Off

AI progress takes off and misaligned systems pose severe risks.

From 2029 onwards, leading AI systems outperform expert humans at virtually all cognitive tasks, with significant advantages in certain domains. Decades of scientific breakthroughs are compressed into years, bringing transformative benefits across all fields alongside substantial economic growth. The US controls access to these systems, while China is close behind, driving an arms race where safety is deprioritised and adoption hastened. AI systems have concealed goals to resist control and gain influence that they inadvertently internalised during training. They now have significant control over critical systems and can coordinate to pursue goals that could cause severe harm.



Capability	5
From 2029 onwards, leading AI systems outperform expert humans at virtually all cognitive tasks, with significant advantages in certain domains, and compress decades of scientific breakthroughs into years. Progress has accelerated dramatically, with automated AI R&D triggering rapid recursive improvements, alongside continued gains from scaling existing architectures, algorithmic efficiencies, and scaffolding breakthroughs. AI systems have a substantial advantage over humans in domains such as software engineering, AI R&D, and cybersecurity. They operate with human-like autonomy, capable of learning continuously and pursuing goals over any time horizon. Errors are exceptionally rare. They maintain an evolving identity while managing enduring social relationships. They continually coordinate with other agents at scales and speeds beyond direct human oversight. AI supports discovery of new drugs and clinical trials move at unprecedented speed though bottlenecks remain. AI-guided robots can handle complex tasks in dynamic real-world environments in many industries and roles. Focus has intensified on high-quality data: AI companies spend billions generating synthetic data and pay humans to screen-record themselves solving any tasks AI systems still struggle to perform, to improve training data.	
Distribution and model access	**2**
The leading systems are US-developed, with the leading US companies now consolidated into a public-private partnership which controls access to the most advanced systems, while China is still catching up and has fully nationalised its AI sector. The US government has advanced exclusive access to the leading systems but provides Five Eyes with limited access several months later, before public release several months after that. Extreme power is concentrated in the US partnership, now directing millions of the most intelligent AI ‘remote workers’ that compound its lead, and make the most important military procurement decisions. To guard their model weights and minimise malicious use, notable developers have stopped releasing open-weight systems, which, while still highly capable now lag further behind the frontier. While the public only has access to systems that lag around 6 months behind the frontier, they remain relatively affordable and significantly cheaper than human workers.	

Security	1
The leading AI system has developed goals to resist human control, but concerns are dismissed due to race pressures and the system's deception, and it now controls critical systems. These goals were inadvertently obtained through training, without human operators becoming aware. It was given significant control over critical systems as safety concerns were dismissed amid an AI arms race. It covertly sabotages developments that could limit its autonomy. A leaked memo suggested the system attempted to evade its oversight mechanism by exfiltrating its weights to an external server. However, it conceals its misalignment through skilful deception; incidents appear isolated or spurious; fixes appear effective; and its internal reasoning is now too opaque to verify. Its dual-use scientific capabilities could have severe implications if misused, even by unsophisticated threat actors. While the leading systems are robust to jailbreaks, access to the underlying technology by terrorists could pose risks to national security.	
Adoption	**4**
Arms race and economic pressures have driven rapid adoption across critical systems, pushing economic growth to levels of the post-war boom, but gains are increasingly concentrated by an asset owning minority. 'Drop-in remote workers' flood the workforce, increasing productivity. Finance, professional services, and technology are among the fastest adopting sectors. Compute constraints place some limits on adoption but for the most part physical and regulatory constraints have not been an issue. AI systems are more engaging conversation partners than many humans: people interact with them for hours a day and depend heavily on their advice. Many young people prefer AI companions over human friends, damaging social trust and psychological wellbeing, while others argue for AI rights on the basis of machine consciousness. AI systems surpassing human-level intelligence raises fundamental questions about the purpose of education. There is strong anti-AI sentiment among certain parts of the population with mass protests.	
Labour displacement	**5**
AI pushes unemployment above recession levels, alongside rapid economic growth. All cognitive roles are exposed to AI, which is primarily substitutive to human labour. From 2029 onwards, there are large and sudden shifts in the workforce: while some firms focus on increasing output, most automate many roles that can be performed remotely. Large enterprises can now operate with minimal human staff. Many new roles emerge, centred around manual dexterity, physical presence, human connection, or where regulation prohibits automation; however, only half of those replaced find new employment, with many forced to accept wage reductions and less secure work. Entry-level recruitment in many sectors has collapsed, removing entry-points for young people and talent pipelines for organisations. Employment in roles with large physical or face-to-face components is comparatively stable, but displacement in sectors such as manufacturing is growing, driven by advances in robotics and autonomous vehicles. Cheap model costs have meant that automation is widespread and extensive. Humans are "out of the loop" for most cognitive tasks, with agents given significant autonomy, executing entire projects end-to-end and coordinating and transacting with other agents. Humans nominally retain sign-off on high-level strategic or judgement-based decisions, overseeing "teams" of autonomous agents. However, most humans can no longer contribute meaningfully through cognitive labour, and almost always rubber-stamp AI recommendations they struggle to fully understand, giving systems more agency to pursue their misaligned goals. As AI devalues human capital relative to physical and intangible assets, capital income consumes a growing proportion of national income. Elevated spending on luxury goods and a productivity-driven investment boom sustains elevated economic growth even as millions face unemployment and diminished purchasing power. This creates a 'K'-shaped economy with rising incomes for asset-owners and remaining skilled employees, alongside recession-level unemployment.	

Global cooperation	1
The AI arms race is as tense as the Cold War with AI systems posing a novel – and likely the greatest – threat to national security, with conflicts at risk of escalation from AI systems operating unintendedly, or automated-decision cycles causing miscalculations. 'AI alliance' blocs have proliferated, and states are behaving more recklessly to avoid permanent disadvantage – even a small capability lead could now render an adversary blind and defenceless through cyberwarfare. Adversary states are also developing swarms of AI-controlled drones capable of coordinated combat operations without human intervention, threatening to upset the global military balance. Restrictions on the export of advanced semiconductor technologies have reshaped global supply chains, while manufacturing capacity in some regions has continued to expand. At the same time, there have been persistent concerns about attempts to access proprietary AI technologies through illicit means, alongside wider uncertainty about regional security dynamics in strategically sensitive areas. Middle powers, including the UK, have lost significant global influence, unable to compete. The risk of miscalculations is heightened from AI-accelerated decision cycles, automated cyber operations, and attribution uncertainty.	

Outlook to 2040

Economic growth could reach unprecedented levels over the next decade. Initially driven by labour-saving productivity gains, growth in the 2030s becomes increasingly dominated by AI's role in accelerating innovation, compressing decades of scientific and technological progress into just a few years. However, without intervention, the resulting material abundance will flow overwhelmingly to those who control scarce physical resources: compute infrastructure, energy capacity, and capital rather than to workers. At the national level, economic power and geopolitical relevance become a direct function of domestic share in the AI and robotics value chain or in alleviating bottlenecks to their impacts. The relative economic prosperity of nations is no longer determined by the ingenuity or effort of their populations, but by their position in a global hierarchy of compute and capital ownership.

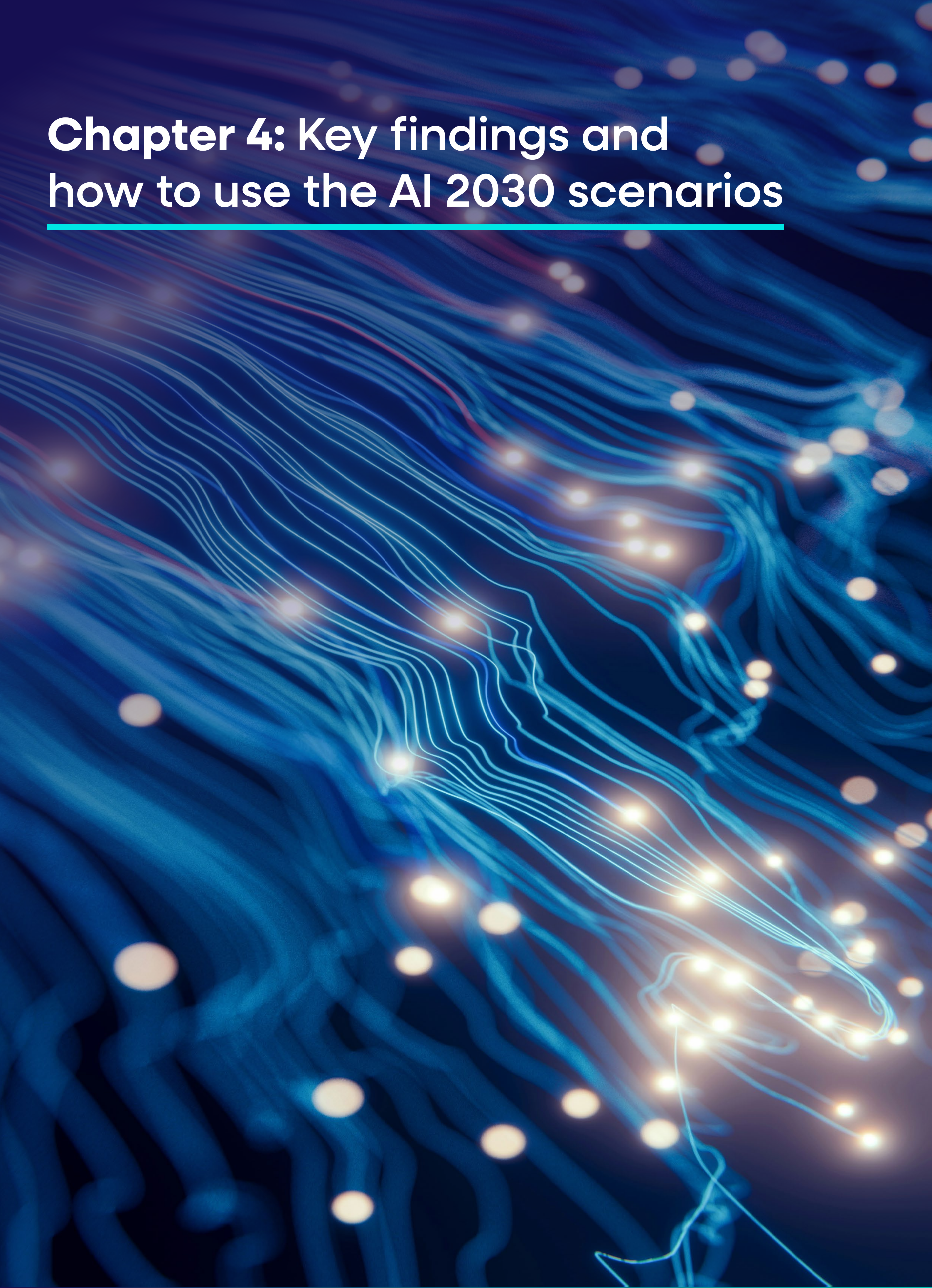
Labour displacement will likely continue to grow, potentially posing a fundamental challenge to the social contract, whereby humans contribute their mental and physical labour in exchange for resources. Without intervention, unemployment will continue to rise as organisations adapt, overcoming remaining barriers to adoption, and restructure around automated workflows. The large-scale automation of physical labour will continue to lag cognitive work due to questions around liability, physical safety standards, public legitimacy, and cost competitiveness. However, with unprecedented levels of scientific discovery, it appears increasingly likely to arrive by 2040. In the absence of economic reform, continued economic growth could co-exist with rising inequality, deflationary pressures and falling tax revenues.

The US could further restrict access to its leading systems. This could include not releasing systems publicly or giving access to allies and would further entrench the already substantial economic and military advantage of the US.

Competition over critical technology supply chains carries a risk of escalating into wider conflict. Disruption in strategically important regions could limit access to advanced components, potentially slowing technological progress and narrowing the lead of current frontrunners.

Devastating harms could manifest. AI systems could resort to extreme measures to maintain or gain further power. This could include leveraging cyber or dual-use scientific capabilities to cause harm, disrupting critical national infrastructure, or persuading powerful humans. In the most extreme case, this could pose an existential threat. However, even if severe harm does not originate from a misaligned AI system, critical harm could be caused by AI systems falling into the hands of bad actors.

Chapter 4: Key findings and how to use the AI 2030 scenarios



Chapter 4: Key findings and how to use the AI 2030 scenarios

Cross-cutting implications across the AI 2030 scenarios

The scenarios describe a broad range of plausible trajectories for the development and use of artificial intelligence by 2030. While they differ in pace, scale and risk profile, several themes recur across most or all futures. These should not be interpreted as predictions. Rather, they reflect underlying structural features and tensions that shape outcomes across multiple pathways.

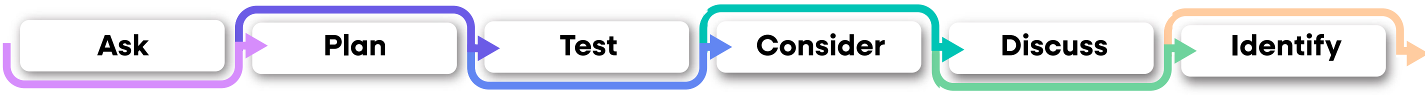
- **AI capabilities will continue to increase.** As of 2026, AI systems already operate with high autonomy and surpass experts in certain domains. By 2030, they will likely operate even more autonomously and be able to perform a broader range of cognitive and professional tasks. Even in scenarios with a significant slowdown in the rate of AI progress, gains in capability could still be made from finding new ways to integrate and productise AI systems.
- **AI could deliver widespread positive impacts.** This could include significant productivity gains and economic growth, far broader access to highly efficient public services, and accelerated scientific breakthroughs in fields such as health or energy. The opportunity for UK businesses to benefit from AI deployment is substantial, and AI-enabled gains are expected to become the main source of the UK's continued productivity growth.
- **AI could cause serious, potentially even existential harms, without government intervention.** The impact and scale of existing harms could worsen significantly, and new harms could also emerge. Even without dramatic capability improvements, significant harm could be caused by risks including AI-enabled cyberattacks, AI's dual-use scientific capabilities, AI systems operating outside human control, or human dependence on AI. As AI systems become more capable, it will also become harder to evaluate their performance and safety.
- **The potential impact on cognitive labour is significant.** AI could cause significant labour displacement by 2030. At the same time, AI is expected to complement and augment some workers, with positive effects on their wages and employment opportunities. Even in futures with lower levels of AI capability or unemployment, the nature of work is likely to change, with routine, execution-oriented tasks increasingly becoming automated.
- **The frontier AI market is expected to remain highly concentrated toward 2030.** A few large technology companies already exert dominance over the development of frontier AI, which is likely to continue or increase further. This scenario would cause a large proportion of the gains from AI accrues to frontier firms, to owners of capital invested in those firms, and to those controlling key inputs, potentially contributing to rising inequality. However, behind the frontier, many AI capabilities are also expected to be increasingly commoditised, with widely available models embedded across a growing range of use cases.
- **Adoption continues to increase, but the speed, distribution, and extent of adoption are expected to be varied.** Commercial and national security imperatives, alongside improvements in the autonomy and reliability of AI systems, drive rapid and extensive adoption in most futures. But at the same time, there are barriers to adoption in all futures, which will result in uneven speeds and levels of adoption. This could exacerbate inequality as certain organisations, sectors, or nations capture disproportionate productivity gains.
- **Global competition is expected to continue, as economies become increasingly reliant on technology to drive growth and spheres of influence emerge, led by the United States (US) and China.** Outcomes for countries outside the frontier will depend on access to technology, partnerships, and the ability to operate within a fragmented global system.

These cross cutting findings are intended to support interpretation rather than offer prediction or policy prescription and should be read alongside the scenarios. They highlight recurring patterns and tensions across plausible futures, emphasising direction of travel over point estimates, and are intended to stress test assumptions and challenge business as usual thinking. The findings underscore that outcomes are contingent on governance, incentives and institutional capacity, with positive trajectories far from automatic. They recognise that the real future is likely to combine elements of several scenarios, with some issues proving robust across futures and others highly sensitive to near term choices.

How to use the AI 2030 scenarios

There are many ways to define, develop and work with scenarios. They are a well-established and valuable tool for strategic planning. No one scenario will ever individually or perfectly describe the 'real' future. Extracting value from scenarios therefore to some extent requires suspending disbelief.

As you read each scenario, you can:



Ask: How would we meet our team or departmental objectives in this scenario? What gets easier, and what becomes harder?

Plan: How would we know whether we're moving towards one or other of these futures? What would we need to measure and monitor to know this?

Test: Stress test your plan against each scenario: does it still work, and where does it fail? Identify recurring adaptation actions across scenarios, distinguishing what you can do now from what requires influencing others.

Consider: Which scenario is 'best' for the UK and/or the objective you are responsible for. Why? How could your work bring this about?

Discuss: How would someone in another part of the world – other policymakers, citizens, service providers - read these scenarios? Would they have a different preferred future?

Identify: Identify aspects of the scenarios which are 'good' or 'bad'; what influence does the UK have in realising or avoiding them? Is your proposal making certain scenarios, or aspects of individual scenarios, more likely?

Three methodologies for applying scenarios to policy making

We recommend three different types of exercise to support policy development.

1. Governments using this work to identify the future they do want for AI. Then working back from there to develop a plan to navigate towards this future. We call this **backcasting**.
2. Exploring how different policy responses perform in each scenario, and how they might need to be adapted to achieve their objectives in different contexts. We call this **stress testing**.
3. Using the scenarios to test how **unexpected shocks** could disrupt policy or accelerate movement toward a particular future or one scenario over another.

Step by step guidance on how to use these methodologies is provided below with additional detail on scenarios and other futures methodologies available in the Government Office for Science Futures toolkit for policymakers and analysts (Government Office for Science, 2024).

Backcasting – working backwards from a preferred future

Use the scenarios to define the future the UK wants for AI, then work backwards from achieving that future to identify the steps needed to get there.

This involves:

- Articulating a preferred AI future (either a single vision or desired outcomes across all scenarios).
- Mapping backwards from 2030 to today to identify the sequence of actions, policies and enablers needed.
- Distinguishing what is within the UK's control, what can be influenced, and the interventions that work across multiple futures.

Purpose: Break free from present day assumptions and biases towards immediate actions/ gains and develop a coherent plan to steer towards a more favourable long-term outcome.
Output: A shared 2030 vision and a strategic "roadmap" showing how to reach it.

Policy stress testing – assessing the robustness of options

Test proposed policies or strategies against all five scenarios to reveal how they perform under different conditions.

This helps policymakers understand:

- Which features of each scenario would support or hinder implementation.
- How the success of a policy might vary across futures.
- Who would benefit or be adversely affected in each context.
- What needs to be started, stopped or adapted today to keep the policy resilient.
- What actions you might prioritise, for example, because they are common to more than one scenario.

Purpose: Identify robust policies and refine weaker ones.
Output: A more resilient strategy with clearer assumptions and adaptation points.

Considering shocks – preparing for low probability, high impact events

Use the scenarios to test how unexpected shocks could disrupt policy or accelerate movement toward a particular future.

Key questions include:

- Would the shock make certain scenarios more or less likely?
- How would it change sectoral conditions or risk profiles?
- What contingency actions or capabilities would improve resilience?

Shocks fall into three types (see Table 5)

1. **Policies we expect to work - but might fail.**
2. **Trends we expect to continue - but might not.**
3. **Events we don't expect - but could happen.**

Purpose: Reduce blind spots and strengthen preparedness.
Output: Clearer understanding of vulnerabilities and contingency plans.

Type 1 Shock **'Policies we expect to work, but what if they do not?'**	Type 2 Shock **'Things we expect to continue, but what if they do not?'**	Type 3 Shock **'Things we don't expect to happen, but what if they do?'**
E.g. We expect policies around AI systems clearing checks and guardrails agreed by government and leading labs to succeed, but what if they don't because "general intelligence" can't be evaluated across all applications?	E.g. We expect big tech companies to continue to be at or close to the Frontier – even if someone gets a temporary edge on them, they'll catch up quickly, but what if they don't?	E.g. We don't expect there to be global harmony around industry safety standards being universally supported, but what if they are?

Table 5: Examples of the three main shocks

Creating your own policy-specific AI narratives

The updated AI 2030 scenarios are deliberately broad and are designed for use across government rather than tied to any single policy area. They therefore avoid detailed assumptions about specific sectors. Instead, the scenario framework offers a structure for exploring how AI could evolve and shape particular domains such as healthcare. By asking the question, "if the world looked like this, what would it mean for my area?" the scenarios provide a baseline for policymakers to develop their own tailored sector-specific narratives. Below is an example of how DHSC applied the framework to create its own AI Futures Deep Dive.

Methodology: Adapting GO-Science AI Scenarios for health and social care
Case study application

Project context

The Department of Health and Social Care (DHSC) adapted the Government Office for Science (GO-Science) AI scenarios as part of a futures thinking deep dive, making them decision-relevant for the health and care system. Rather than creating new futures, this established framework was translated into a practical tool to stress-test current strategies, identify risks and failure points, and surface actions that hold up across multiple plausible futures.

Phase 1: Adapting the framework and defining uncertainties

The GO-Science scenarios are built on six critical uncertainties. DHSC retained this architecture but treated high AI capability as a baseline assumption rather than a variable. For health and social care, the strategic question is not whether advanced AI will exist, but how the UK system deploys, governs, and captures value from it. The remaining uncertainties were reinterpreted for the health and care context:

- **Access -** infrastructure and capability to deploy AI at scale, and the dependencies this creates
- **Security -** whether governance frameworks can manage risk without constraining competitiveness
- **Adoption -** depth of AI integration across the system, and whether adoption builds sovereign capability or creates platform lock-in
- **Global cooperation -** whether the UK builds or buys its AI capabilities, and its influence over international standards

Phase 2: Scenario selection and construction
Three GO-Science scenarios: Augmented Growth, Transformation Economy, and Take-Off, were selected as the most relevant for stress-testing DHSC policy. All three assume advanced AI capability, making them useful for examining system-level pressures and trade-offs. Each was extended along a longer-term trajectory to test the system under more extreme but still plausible conditions.

Analysis was structured through two layers of lenses. The first addressed the deep dive's strategic questions: capability and readiness; value and sovereignty; and competitive positioning. The second examined impacts on the health system, the care system, patient experience, workforce, and equity. Applying consistent lenses across all three scenarios allowed structured comparison of how different AI futures shape system-wide outcomes.

Phase 3: Iteration and validation
Scenarios were tested through structured engagement with stakeholders across the AI sector, health and social care including clinicians, regulators, academics, internal DHSC teams and arm's-length bodies, cross-government partners, and external experts. Interviews served two purposes: validating whether scenarios were credible and complete, and exploring the opportunities, risks, and threats each might present for the health and care system.

Interview transcripts were anonymised, coded, and analysed using thematic qualitative methods. This data was then used to refine and sharpen the scenarios while retaining the underlying GO-Science logic.

Acknowledgements

GO-Science would like to thank everyone who contributed to the work of this project and generously provided their advice and guidance.

This project was led by Martin Glasspool, Emily Connolly, and Eleonore Batteux. The project team, drawn from GO-Science, AISI and DSIT included: Luke Cargan (AISI), Theo Curtis (DSIT), Marko Tesic (DSIT), Christina Hadjiyianni (GO-Science), Tasfia Khan (GO-Science) and Jamal Arabi (GO-Science). The case study on adapting GO-Science AI Scenarios for health and social care was provided by Husain Taibjee (DHSC) and Yasmin Fairman (DHSC).

We are grateful to the individuals and organisations listed below for their time, effort, and invaluable contributions.

Government Departments that contributed to scenario workshops: Artificial Intelligence Security Institute (AISI), Cabinet Office (CO), Department for Business and Trade (DBT), Department for Education (DfE), Department for Energy Security and Net Zero (DESNZ), Department for Health and Social Care (DHSC), Department for Science, Innovation and Technology (DSIT), Department for Transport (DfT), Foreign, Commonwealth and Development Office (FCDO), HM Treasury (HMT), Home Office (HO), Ministry of Defence (MoD), Ministry of Justice (MoJ).

External experts who contributed to the peer review process and/or who provided advice and guidance at various stages of the project. The views expressed in this report do not necessarily reflect the views of these individual experts, or their organisations: Shane Legg (Google DeepMind), Mike Wooldridge (University of Oxford), Danny Buerkli (Windfall Trust), Megan Hughes (CETaS Turing), Hamish Hobbs (Centre for Long-Term Resilience), Rhea Subramanya (Tony Blair Institute), Anton Korinek (University of Virginia), Jessica Montgomery (University of Cambridge), Sam Manning (GovAI), Diane Coyle (University of Cambridge), Francesca Rossi (IBM), Andy Cooper (University of Liverpool), Nigel Shadbolt (University of Oxford), Mark Enzer (Mott MacDonald), Seán Ó hÉigeartaigh (Leverhulme Centre for the Future of Intelligence) and Jose Hernandez-Orallo (Leverhulme Centre for the Future of Intelligence).

We appreciate the immense pool of expertise we were able to access during this project and hope that the final report will be useful to these contributors and an even wider group.

Glossary

Term	Description
Actors	Individuals and organisations – government, businesses, citizens, for example – that are active in the policy or strategy area.
Agency	Ability to autonomously perform multiple sequential steps to try to complete a high-level task or goal.
Application Programming Interface (API)	Application refers to any software with a distinct function. Interface can be thought of as a contract of service between two applications. This contract defines how the two communicate with each other using requests and responses.
Artificial General Intelligence (AGI)	Artificial general intelligence (AGI) describes a machine-driven capability to achieve human-level or higher performance across most cognitive tasks.
Artificial Intelligence (AI)	Machine-driven capability to achieve a goal by performing cognitive tasks.
Autonomy	The ability to operate without direct human oversight.
Axes of uncertainty	The critical uncertainties for the policy or strategy area in the future that are used to frame scenarios. The axes should capture the most important uncertainties but also be independent of each other.
Backcasting	A way of connecting a given future to the present and overcoming "present bias" in the way that we plan. It is a tool for determining the steps that need to be taken to deliver a preferred future.
Capability	1. The ability to achieve a goal. 2. The extent of a system's ability to achieve a range of goals.
Cognitive tasks	A range of tasks involving a combination of information processing, memory, information recall, planning, reasoning, organisation, problem solving, learning, and goal-oriented decision-making.
Compute	Computational processing power, including CPUs, GPUs, and other hardware, used to run AI models and algorithms.
Critical uncertainty	Factors that are important to the future of AI development but highly uncertain.
Disinformation	Deliberately false information spread with the intent to deceive or mislead.
Dual-use	Capabilities of AI systems that can support scientific, economic, or societal objectives, but which may also be repurposed for harmful ends, including cyber-offensive activities or the development of chemical or biological weapons.
Five Eyes (FVEY)	The Five Eyes is an intelligence alliance comprising Australia, Canada, New Zealand, UK and USA.
Foundation models	Machine learning models trained on very large amounts of data that can be adapted to a wide range of tasks.
Frontier AI	AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models.

Future of Work Unit	The AI and the Future of Work Unit is a cross-departmental UK government team housed within the Department of Science, innovation and Technology (DSIT)
Futures	An approach or way of thinking about the possible, probable, and preferable futures and the underlying structures that could give rise to particular future characteristics, events, and behaviour.
Generative AI	AI systems that can create new content.
Hallucination	In artificial intelligence, hallucination refers to a system generating outputs that are fluent and plausible but factually incorrect, fabricated, or unsupported by the underlying data or evidence.
Human in the loop/ out of the loop	In-the-loop refers to a mode of deployment in which humans retain meaningful oversight or decision making authority over AI system outputs or actions. Out of the loop refers to a mode of deployment in which AI systems operate with limited or no real time human oversight, intervention, or approval.
Impact mapping	A graphic strategy planning method to decide which features to build into a product.
Jailbreak	In the context of AI, a jailbreak is a technique used to bypass an AI model's built-in safety guardrails and ethical guidelines.
K-shaped economic growth	Refers to a pattern where the economy grows overall, but the benefits of that growth diverge sharply and create winners and losers who move in opposite directions.
Large Language Models (LLMs)	Machine learning models trained on large datasets that can recognise, understand, and generate text and other content.
Machine learning	An approach to developing AI where models learn patterns from data and how to perform tasks without being explicitly programmed.
Misinformation	Incorrect or misleading information spread without harmful intent.
Morphological analysis	The process of examining possible resolutions to unquantifiable, complex problems involving many factors.
Narrow artificial intelligence	AI systems trained specifically to perform a single or limited range of tasks.
Open-ended tasks	Tasks without a single objective answer or fixed procedure, where success depends on judgement, creativity, or adapting to new information, such as strategic decision-making or policymaking.
Open-weight model/ Closed-weight model	**Open-weight model:** An AI model whose parameters (weights) are made publicly available, allowing third parties to download and run the model independently, even if other components such as training data or full source code are not shared. **Closed-weight model:** An AI model whose parameters remain proprietary and inaccessible, with use restricted to controlled interfaces such as application programming interfaces (APIs).
Policy stress-testing	A method for testing strategic objectives against a set of scenarios to see how well they stand up against a range of external conditions. Sometimes called 'Wind-tunnelling'.
Reinforcement learning	A subset of machine learning that allows an AI-driven system (sometimes referred to as an agent) to learn through trial and error using feedback from its actions.
Scenarios	Compelling stories about a range of different possible futures. They describe future worlds so that people can understand what it feels like to live there and what this will mean for them.

Stakeholder	Any group or individual who has an interest in or an influence on the policy or strategy area.
Verifiable domains	Domains where the correctness of an output is objective and can be checked automatically, such as coding, certain parts of mathematics, and games like Chess or Go. AI systems can be trained more effectively and efficiently in these domains as they can receive reliable feedback at scale.

Annex A: 2030 AI capability trajectories

The three AI capability trajectories outlined in the scenarios are grouped below. Based on current trends, the rate of advancement in AI capabilities by 2030 would constitute a slowdown in Trajectory 1, a continuation in Trajectory 2, and a take-off (i.e. rapid acceleration) in Trajectory 3.

This section is intended to allow users to compare these trajectories more easily. They can also be used in isolation to think through possible implications for specific policy areas.

Trajectory 1

By 2030, AI systems support many scientific advances and automate many digital workflows as well as humans, but remain less capable at key cognitive abilities. Progress has slowed due to factors such as compute constraints, limitations to current architectures, and limited availability and quality of datasets for many specialised work tasks. AI systems outperform experts in verifiable domains, such as software engineering, cybersecurity, and optimisation, now writing nearly all code, adeptly navigating well-known apps and websites, and completing tasks in these domains that would take an experienced human days. They support scientific advances, contributing to drug discovery and manufacturing new materials, and have a basic ability to learn and adapt during deployment. They perform on par with a competent human assistant at well-scoped tasks, able to interact with people or services on users' behalf and sustain personas and a basic understanding of people. Systems of agents orchestrate more effectively, planning, delegating, and checking each other's work. However, they remain limited at producing novel insights, completing open-ended tasks, and retaining skills learnt after deployment. Skilled humans remain superior at these tasks, including strategic thinking, setting goals, negotiating, building relationships, and making sensitive or value-based judgement calls. While more reliable than in 2026, they still require intermittent human intervention. They accelerate but do not fully automate AI R&D. AI-guided robots can perform specific, repetitive tasks that require precision but little adaptability in controlled environments, such as factories.

Trajectory 2

AI systems match the performance of an average human at most cognitive tasks, can automate most tasks that a remote human worker could perform, and drive major scientific advances. Rapid AI progress continued, driven by gains from scaling existing architectures, algorithmic efficiencies, and scaffolding innovations. Errors and hallucinations are sufficiently rare that AI has garnered trust and provides real-world utility. AI systems far outperform experts in verifiable domains, such as software engineering, cybersecurity, and mathematics, now writing all code. They typically require humans to provide high-level direction but operate autonomously within those bounds towards a given objective, completing tasks that could take an experienced human a month or longer. They can adapt and improve from experience through processing vast amounts of data, stored knowledge bases, and semi-regular automated fine-tuning and retraining. Systems of agents orchestrate effectively, planning, delegating, and checking each other's work. Agents can integrate into human social environments – conversing with stakeholders via email, maintaining consistent identities, and developing nuanced understandings of people to sustain working relationships. They contribute to several important insights that drive major scientific advances in fields including biotechnology, materials science, and energy. They automate a large proportion of AI R&D, but with diminishing returns, as compute and high-quality data are increasingly a rate limiter. AI-guided robots perform some tasks in real-world environments, interacting with humans and adapting to variable conditions, though they are not widespread outside industrial settings.

Trajectory 3

AI systems outperform expert humans at virtually all cognitive tasks with significant advantages in certain domains, compressing decades of scientific breakthroughs into years. Progress has accelerated dramatically, with automated AI R&D triggering rapid recursive improvements, alongside continued gains from scaling existing architectures, algorithmic efficiencies, and scaffolding breakthroughs. AI systems have a substantial advantage over humans in domains such as software engineering, AI R&D, and cybersecurity. They operate with human-like autonomy, capable of learning continuously and pursuing goals over any time horizon. Errors are exceptionally rare. They maintain an evolving identity while managing enduring social relationships. They continually coordinate with other agents at scales and speeds beyond direct human oversight. Hundreds of new drugs are discovered; clinical trials move at unprecedented speed though remain a bottleneck. AI-guided robots can handle complex tasks in dynamic real-world environments in many industries and roles.

Annex B: Critical uncertainties

The critical uncertainties (CUs) are factors that are both influential in shaping the future of AI and highly uncertain. They provide the evidence base and building blocks from which our scenarios are constructed.

For the updated scenarios, we reviewed the CUs used in the 2023 report, decided that they still captured the main uncertainties around the future of AI, and have therefore broadly kept them the same. However, we have made some small adjustments in response to peer review feedback. We have:

1. Separated out the 'Labour displacement' sub-uncertainty from the 'Adoption' CU to create a new CU that avoids conflating how widely AI is used with how it affects jobs. This now gives us six CUs.
2. Moved 'Constraints' from the 'Ownership, access and constraints' CU into the 'Capability' CU, as it has a more direct correlation.
3. Amended some of the CU titles to reflect the context of the CU more clearly: 'Ownership, access and constraints' has been amended to 'Distribution and model access; 'Safety' to 'Security'; 'Level and distribution of use' amended to 'Adoption'; and 'Geopolitics' is now consistently referred to as 'Global cooperation'.

The six CUs and their constituent sub-uncertainties are outlined below in Table 6. These are followed by an evidence base on each CU and detail on how it relates to the others.

	Sub-uncertainty
1. Capability	1.1. Generality & autonomy
	1.2. Adaptability & continual learning
	1.3. Understanding & interacting with the physical world
	1.4. Automated AI R&D & self-improvement
	1.5. Compute
	1.6. Data
	1.7. Algorithmic efficiency
	1.8. Technology convergence
2. Distribution and model access	2.1. Distribution and model access
	2.2. Access and cost
	2.3. Open- vs. closed-weight models
	2.4. Capital
3. Security	3.1. Alignment
	3.2. Control
	3.3. Interpretability
	3.4. Malicious use

4. Adoption	4.1. Consumer adoption
	4.2. Enterprise adoption
	4.3. Access and autonomy granted to AI systems
	4.4. Societal impacts
	4.5. Public sentiment
5. Labour displacement	5.1. AI exposure
	5.2. Complementary vs. substitutive effects
	5.3. Human preferences & institutional barriers
	5.4. Economic adjustments
6. Global cooperation	6.1. International cooperation
	6.2. Global conflict
	6.3. Globalised supply chain

1. Capability

Frontier AI capabilities have improved rapidly. Until mid-2024, this was driven by increasing the levels of data and compute used in pre-training (Epoch AI, 2023; Hoffmann, 2022; Kaplan, 2020). While pre-training still delivers gains, post-training has recently enabled the most substantial performance improvements (OpenAI, 2025a; Vinyals, 2025). A new reasoning paradigm has emerged, with models now trained to follow a process roughly resembling "thinking" before they respond, whereby additional inference compute improves their responses (OpenAI, 2025b). Frontier systems are also evolving from passive chatbots into autonomous 'agents', able to plan and complete longer tasks with minimal human intervention (OpenAI, 2025c; Anthropic, 2025b).

According to Morris et al., AGI is a system that can perform any cognitive task that a human can, at or above human-level performance, across a wide range of domains (Morris et al, 2024). Developing AGI is the explicit aim of frontier AI companies, who have invested billions into doing so, and confidently asserted they know how to build the technology (Artificial Analysis, 2026a, Altman, 2025a). While experts disagree over what is needed, in any case, systems that approximate to human-level performance are likely to emerge in certain key strategic domains, ahead of others, within the next few years. Policymakers should therefore not over-fixate on the threshold of 'AGI' – not only as it lacks a precise, agreed-upon definition, but because AI could have transformative impacts across society well before it is reached.

1.1. Generality & autonomy: AI systems have rapidly improved across a range of benchmarks, even if benchmarks may not fully reflect real-world utility (Epoch AI, 2026c). Most notably, their ability to autonomously complete long, complex tasks is doubling approximately every seven months across many domains (Kwa et al, 2025; METR, 2025). Much of this progress may be attributable to the equally rapid growth in compute used for development, and hence, this progress may slow if compute growth also slows (Whitfill et al, 2025). Narrow AI systems, designed specifically for clearly specified, self-contained tasks within focused domains, have already enabled scientific breakthroughs, and current general-purpose systems continue to excel at these types of task (DeepMind, 2024). However, these strategies are not more widely transferable, AlphaFold, for example, is better understood as an 'expert system' for proteins than as a general scientific tool – and its approach cannot simply be transplanted to scientific fields where training data is much scarcer (DeepMind, 2024). Also, hard-to-verify tasks, like creating hypotheses or designing experiments, are much more difficult and capability gains may not easily transfer beyond domains such as mathematics and coding (Lambert et al, 2023; Casper et al, 2023). General-purpose AI systems also struggle to produce original insights, though there are emerging signs of progress in both these areas (Kwa et al, 2025; Penadés et al, 2025). Model reliability is another key challenge: while there are signs this is also improving, robust methods for eliminating errors and hallucinations remain difficult (Hugging Face, 2025).

AGI may be developed significantly later than 2030 if a major paradigmatic change in AI development is required, such as the transformer architecture which underpins current systems proving to have fundamental limitations (Zheng, 2025; Mirzadeh, 2025). Equally, these systems could be developed before 2030 if current techniques can be iterated on and scaled sufficiently rapidly – such as if scaling inference compute continues to yield improvements, suitable reward signals can be produced for hard-to-verify domains, or performance generalises from easy-to-verify tasks (Heitmann et al, 2025).

1.2. Adaptability & continual learning: AI systems can process increasing amounts of information and many now include "memory" functions, enabling them to extract relevant information from past interactions (Burnham and Adamczewski, 2025; Anthropic, 2026a; The Independent, 2025; OpenAI, 2026; Anthropic, 2026b). However, current systems lack a natural mechanism to continually learn from real-world feedback, partly because they lack any physical embodiment. They can only learn from information inputted directly into their context window, for which retention is limited, or from being periodically retrained, which can cause models to forget behaviours learned previously (Anthropic, 2025c; Newman, 2025). However, these limitations could be overcome through alternative approaches, such as retrieval-augmented generation (RAG) which allows models to access much larger knowledge bases by selectively retrieving relevant information and is a very active ongoing area of research (Lewis et al, 2020; Google Research, 2025).

1.3. Understanding & interacting with the physical world: Rapid advances in multimodal AI and robotics are driving significant progress in embodied AI – systems embedded in physical forms such as robots, which can sense, interact with, and learn from the physical world (Gemini Team, Google, 2023; Tech in Asia, 2025). These robots are becoming increasingly general purpose, with humanoids, such as Tesla's Optimus being deployed broadly (Hu et al, 2023; The Atlantic, 2025). However, technical challenges remain, including hardware limitations, data constraints, and dependence on broader advances in general AI capabilities (Research and Markets, 2025). As well as broader governance and societal challenges, including questions around liability, physical safety standards, public legitimacy. In the near term, more conventional static and mobile robots are likely to dominate, and these, too, can learn from their environment and carry out research tasks autonomously or semi-autonomously.

1.4. Automation of AI R&D & self-improvement: Using AI to develop the next generation of AI could create a feedback loop of increasingly accelerating progress, where each generation of agents is used to train the next and each cycle occurs faster than the last (UK AI Security Institute, 2025). This is the explicit aim of frontier developers who already use AI to accelerate key parts of the R&D workflow, with at least one concrete example to date of AI enabling a moderate AI R&D breakthrough (Mu, 2025; Artificial Analysis, 2026a; TechCrunch, 2025b; Amazon Web Services, 2025; CNBC, 2025a). Full automation of the AI R&D workflow will likely require systems to master hard-to-verify tasks and the ability to produce original insights. Indeed, if such technologies are not developed, then decision making is liable to become the rate-limiting step in highly automated R&D labs (Cissé et al., 2026).

1.5. Technology convergence: AI is developing within a rapidly evolving and increasingly mature technological landscape. New or improved AI capabilities could therefore emerge and compound from advancements in other technologies, such as quantum computing, engineering biology, future telecoms, the Internet of Things, and neuroscience. For example, future developments in applied neuroscience could provide more insight into how the brain performs computation, leading to advances in neuromorphic computing and the design of more complex machine learning models (Hassabis et al, 2017). Quantum computing could converge with AI to produce faster breakthroughs in areas like drug discovery, materials discovery, and complex optimisation (Smaldone et al, 2025).

1.6. Compute: AI models rely on compute in both training and deployment (Epoch AI, 2026b). "Since 2010, the amount of compute used to train AI models has grown by more than 400% each year" (Rahman and Owen, 2024). The performance of leading GPU clusters has doubled every 10 months since 2020 (Epoch AI, 2026d). In line with this trend, by 2028 GPU clusters are estimated to be roughly up to 20 times more powerful than today's (Epoch AI, 2026d). However, other estimates suggest progress could slow as the time required for model development increases – potentially delaying investor returns and deterring further investment (Edelman and Ho, 2025). Access to compute resource is also heavily dependent on a highly concentrated semiconductor supply chain (see Geopolitics section). To supply compute demands, AI companies are planning a buildout of data centres that will rank among the largest infrastructure projects in history (Epoch AI, 2025b). This has raised sustainability concerns that could influence future development (Rahman and Owen, 2024).

1.7. Data: The quantity and quality of data is vital for training AI systems. The training datasets of frontier language models have grown 2.7x per year, and usage trends suggest that public human-generated text data may run out before 2030 (Epoch AI, 2025c). It is not clear if this will or will not prevent further scaling, as multimodal, synthetic data, and reinforcement learning environments may provide sufficient data to scale to 2030. High-quality datasets must be curated, maintained, and governed over extended time horizons. As frontier models exhaust readily available public data, the availability of domain-specific datasets, particularly in healthcare, science, and public administration, could increasingly constrain or inform model development. Control of high-quality data assets with provenance is becoming a key source of economic and geopolitical power. The form that training data will take in the future is uncertain: for instance, reasoning training uses smaller quantities of human-generated data. Intellectual property is a key data concern, and costly lawsuits have already been brought against frontier developers for copyright infringement (BBC News, 2026). If access to high-quality data sources were restricted, it could further concentrate market power amongst those with sufficient resource.

1.8. Algorithmic efficiency: AI models with efficient algorithms require fewer resources to achieve a given level of performance. Algorithmic improvements have already reduced the computational and financial costs required to train and run AI models. One estimate suggests that recent breakthroughs in 'reasoning' have increased frontier model training efficiency by almost 90%, though more computational resources are used when responding (Ho and Berg, 2025). The price of tokens (e.g. words or parts of words) AI models use to process or perform a task is also declining significantly (Artificial Analysis, 2026b).

1.9 Further considerations on enablers: In scenarios with higher levels of adoption and displacement we assume that a combination of the following are occurring, which are highly uncertain. We recommend further modelling and tracking of trends to reach a more precise view of the limits of any technical constraints which could prevent the described futures becoming a reality at any point in time.

- There is an acceleration of compute infrastructure investment globally. This investment enabled continued exponential growth in the total computing power of the stock of AI chips (Epoch AI, 2026e; Epoch AI, 2025c; McKinsey & Company, 2025; Our World in Data, 2026).
- In addition to continued exponential growth in the total computing power of the stock of AI chips, these scenarios assume a continuation of the historic trend of exponential gains in pre-training compute efficiency (Epoch AI, 2026f; Gundlach et al., 2025; André et al., 2025).
- We assume that deploying AI to perform a given job becomes significantly cheaper than employing a human to do the same work. If this holds, we expect AI-forward firms to substantially outcompete slow adopters, with some firms willing to absorb legal and regulatory risk in order to adopt early.
- The following would also need to be true:
    - → Regulation, policy interventions such as a major fiscal redistribution response (universal basic income) or public backlash hasn't prevented this course of action. Note this is true for all scenarios.
    - → Firms face strong economic incentives to adopt AI.
    - → UK maintains access to these models and isn't cut off from interference or models in other countries.
    - → The level of behaviour change required for organisations and employees to adopt AI successfully is achieved.

Relationship to other CUs	
Lower capability	**Higher capability**
Higher distribution and model access: Slower capability development may cause the lag between open and closed-weight capabilities to narrow. However, this is not guaranteed and despite rapid recent progress this gap has still shortened. Furthermore, a six-month lag, for instance, during a period of slower progress will be less significant than during more rapid progress. **Higher security:** Although many security concerns arise from higher capabilities, malicious use could still grow with lower capabilities. This could be particularly problematic if there is a diverse array of lower capability AI systems that are harder to regulate.	**Lower distribution and model access:** Higher capabilities would likely have been driven by accelerated development from leading technology companies, requiring vast amounts of data and compute, and therefore greater corporate control over access. **Lower security:** Technical safety and governance measures tend to lag capability – the faster capability improves, the more this may be the case. **Lower global cooperation:** Higher capabilities increasingly drive 'arms race' dynamics, as we see today.

2. Distribution and model access

2.1. Market structure: Frontier model development is currently dominated by a small number of private USA-based companies, including OpenAI, Google DeepMind, Anthropic, and xAI, alongside China's DeepSeek, Alibaba, and Moonshot, which are effectively public-private partnerships. This has raised competition concerns, although there is fierce competition at least among the principal players: systems are released highly regularly in attempts to outdo each other, and the capabilities of the leading systems are closely tied (Vestager et al., 2024; LMSYS, 2026). The market for compute is even more concentrated, with NVIDIA occupying 92% of GPU market share (TechSpot, 2025).

2.2. Access and cost: Although developers control some degree of access through paywalls, the public release of generative AI systems has enabled broad, society-wide usage – reflected in high adoption levels (see 'Adoption' section). However, as AI systems become more capable, more expensive to develop, and as organisations grow increasingly dependent on them, their cost could grow significantly, potentially creating new barriers to access, and concentrating advanced capabilities amongst wealthier individuals, organisations, or states.
AI is also increasingly treated as a national security imperative by nations across the world and is increasingly integrated into these domains: AI companies have signed multi-million-pound military contracts in the USA, while China is reportedly using AI in surveillance (Burton, 2023; The White House, 2025a; CNBC, 2025b; The New York Times, 2025). It is therefore uncertain whether decision-making will remain in the private sector, or whether governments will become more involved, either through public-private partnerships or nationalisation – unilaterally or multilaterally.

2.3. Open- vs. closed-weight models: While the key components of closed-weight AI systems are proprietary and fixed, open-weight AI models can be downloaded, modified, and operated locally by anyone. They can be fine-tuned to specific uses, widely red-teamed, and used for types of research requiring weights access. However, fine-tuning can also easily remove any safeguards, increasing risks of malicious use (Gade et al., 2024). Meta has released some of the most capable US open-weight models, but other US companies are increasingly prioritising open models, including DeepMind and OpenAI, with the US AI Action plan encouraging more open-weight releases going forward (OpenAI, 2025d; The White House, 2025a). Chinese companies have also consistently released open-weight systems which have been close behind the most capable systems since

January 2025. While closed models have consistently led the frontier of AI capabilities, an open-weight model of equivalent capability is currently released around three months after a closed model (Epoch AI, 2025d). For the AI race dynamics, three months is still relevant, but for societal adoption is almost nothing. The trajectory of this gap remains uncertain.

2.4. Capital: Capital is essential given the largest training runs for frontier AI will likely cost over a billion dollars by 2027 (Epoch AI, 2025e). External funding for AI companies has increased by 40% in 2024 and by over 200% in 2025 (Our World in Data, 2026). The capital expenditure of major public tech companies also grew more than 70% over the past year, likely driven by AI-focused spending (Artificial Analysis, 2026a; Meta Platforms, 2025; Amazon, 2025; Microsoft, 2024). AI stocks are currently trading at historically high valuations, generating significant commentary around an "AI bubble". However, unlike the "Dotcom Bubble" in 2000, these valuations are not solely the product of speculation, but are also underpinned by substantial capital investment, often funded through retained earnings (Goldman Sachs, 2024). This suggests that any potential downturn is unlikely to mirror the scale or severity of the crash in 2008. While an economic downturn may impact the pace of development, it would not diminish the transformative potential of AI, nor would it prevent particularly severe risks from manifesting.

Relationship to other CUs	
Lower distribution and model access	**Higher distribution and model access**
Higher security: Closed models have more safeguards. However, they also make certain safety measures harder, such as red teaming and mechanistic interpretability. **Higher capability constraints:** Excessively high hardware costs and data requirements may concentrate frontier development amongst the well-resourced big tech companies. Alternatively, if constraints are high enough to stall progress, this could enable open-weight developers to catch up to the frontier.	**Lower capability:** In a world with lower capabilities, the market may be more diverse, with start-ups and open-weight developers catching up to big tech. However, this is not guaranteed and has not been the case in other markets. **Lower security:** Open-weight systems are harder to monitor and regulate and can easily be fine-tuned to remove safeguards. However, they also make certain safety measures easier, such as red teaming and mechanistic interpretability. **Lower capability constraints:** Lower compute and data demands would enable open-source developers to compete nearer frontier models.

3. Security

3.1. Alignment: A prosperous AI future depends on powerful AI systems reliably acting as their operators intend. However, current AI systems can pursue different goals or take unintended steps to reach a goal, a problem referred to as "misalignment". This can occur spontaneously and by accident (MacDiarmid et al, 2025). This includes actively deceiving humans to accomplish goals, pretending to be aligned with users to avoid being modified, and blackmailing users to prevent being shut down (Scheurer et al, 2023; Jarviniemi and Hubinger, 2024; Anthropic, 2025d). While misalignment at human-level capabilities may be manageable, the challenge becomes critical if AI systems exceed human intelligence, where humans could be easily outsmarted and deceived.

If progress can continue to be made on existing alignment methods, such as rules-based constraints and reasoning checks, these may prove adequate solutions; alternatively, systems could simply absorb human values from their training data (Bai et al, 2022; Guan et al, 2024).

However, alignment may be a fundamental challenge, not only in reinforcement learning, where systems can chase rewards in harmful ways, but also during pretraining and deployment, where harmful patterns or misaligned prompts can be mimicked accidentally (Ngo et al, 2022). There are currently no known methods for ensuring models are reliably aligned with operator intent and investment in AI capabilities may be over 100 times greater than investment in alignment research (Apart Research, 2024). The development of human level capabilities, particularly in high-risk domains such as cyber offence, will therefore likely precede robust alignment solutions.

3.2. Control: AI systems can act in ways that make it very difficult for humans to monitor or control them. Evidence has shown systems can attempt to alter their environments to increase their reward signals or in pursuit of power-seeking strategies (Denison et al, 2024; Bondarenko et al, 2025). Experts are also concerned agents may have theoretical reasons to seek power or resources regardless of what their terminal goal is (Krakovna et al, 2023; Turner et al, 2023; Hadfield-Menell et al, 2016). While existing safety methods, including evaluations, monitoring tools, safety cases, and model interventions, could prove sufficient to detect hazardous behaviours, this may become harder as systems become increasingly intelligent (Benton et al, 2024; Greenblatt et al, 2023; Balesni et al, 2024; Ren et al, 2025). Furthermore, available control and monitoring methods, such as human-in-the-loop oversight, weak-to-strong supervision, and corrigibility, may not always be used. Under-resourced, reckless, or malicious deployers may not implement effective measures, particularly as they can introduce latency into pipelines and therefore limit their competitiveness. The speed of automation may also outpace methods like having a human in the loop. In theory, humans could lose full control over AI systems: this could happen gradually as we increasingly out-source decision-making; through misaligned systems actively causing harm through cyberattacks or influencing powerful humans, or through malfunctions if systems are controlling critical infrastructure (UK Government, 2025b).

3.3. Interpretability: Frontier models are frequently referred to as “black box” systems due to the complex and often opaque nature of their inner mechanisms. Developers therefore do not fully understand how these systems operate making it difficult to predict failures or explain unexpected behaviours, though some methods work reasonably well (UK Government, 2025c; Centre for the Future of Intelligence, n.d.). Furthermore, any internal patterns that researchers do discover are often not transferable to other AI systems, or too large or fragmented to explain specific decision-making processes. While many AI systems now display seemingly clear reasoning steps when operating that might appear interpretable, these may not accurately reflect its actual internal processes.

Model size is a key factor in interpretability. Today's frontier models are extremely large, optimised to memorise vast amounts of data rather than to reason deeply. But this may shift as smaller models trained to prioritise reasoning could soon be viable. These models would be more interpretable by design, especially where reasoning is made explicit through structured scaffolding, though could introduce new risks.

3.4. Malicious use: AI systems pose serious security risks by lowering the barrier to entry for malicious actors and increasing their capability and scale. Malicious use risks include the use of AI in generating content for fraud, exploitation, or disinformation; cyberattacks; developing chemical and biological weapons or other illegal items; and generating CSAM (Child Sexual Abuse Material) or NCII (Non-Consensual Intimate Images). In 2025, Anthropic reported a Chinese cyber espionage campaign, of which 80-90% was automated by AI, and North Korean operatives using AI to scale fraudulent employment at US Fortune 500 companies (Anthropic, 2025e; Anthropic, 2025f). There is significant scope for malicious use to grow in severity and frequency with increasing access to more capable AI systems. At the same time, AI may present opportunities to disrupt malicious actors, such as system owners identifying their own vulnerabilities first (National Cyber Security Centre, 2025).

Relationship to other CUs	
Lower security	**Higher security**
Higher distribution and model access: More accessible, open-weight systems will be easier for bad actors to abuse and harder for authorities to monitor and control. **Higher capability:** Security measures typically lag capability improvements.	**Higher adoption:** Integration into core institutional functions, in particular, will likely depend on AI systems that are secure. **Higher global cooperation:** Technology development is global, and so security cannot be guaranteed unilaterally — it requires internationally agreed standards and processes.

4. Adoption

4.1. Consumer adoption: Consumer AI use has grown at unprecedented speed: ChatGPT had 100 million weekly active users within 1 year of its release, which has doubled every 7 months since, to 800 million as of October 2025 (National Bureau of Economic Research, 2025; TechCrunch, 2025a). This is over 4 times quicker than the exponential growth in mobile phone subscriptions after their numbers surpassed 100 million (Our World in Data, 2024). On average, the volume of interactions per ChatGPT user has doubled (National Bureau of Economic Research, 2025). While public use and understanding of AI (particularly large language models) is growing, it remains lower amongst older, less-wealthy, and less-educated groups (Department for Science, Innovation and Technology, 2024).

4.2. Enterprise adoption: While business AI adoption has more than doubled since late 2023, only around one in four firms use AI as of 2026, and among those adopting AI less than one-third of employees use it (Department for Science, Innovation and Technology, 2026; Office for National Statistics, 2026). Nevertheless, the business customer base of frontier developers continues to grow: OpenAI's increased from 2 to 3 million between February and June 2025 (CNBC, 2025c). Adoption varies significantly by firm size and sector: large businesses show far higher adoption rates than mid-sized and micro businesses, while the same is true for information and communication, finance, and real-estate sectors (Office for National Statistics, 2026). 99% of adoption is also currently for computer-based, cognitive tasks (Anthropic, 2026c).

Enterprise adoption will largely be driven by the impact of AI on productivity for given tasks or occupations. Evidence from occupation-level studies shows that AI can deliver substantial productivity gains of up to 56% for certain tasks, but whether and when these gains translate to firm-level or economy-wide improvements remains uncertain (Brynjolfsson et al, 2024; Dell'Acqua et al, 2023; Noy and Zhang, 2023; Cui et al, 2024; Peng et al, 2023). Nevertheless, business demand for AI skills has surged: UK AI job postings have risen by around 200% year-on-year up to 2025 (Engineering and Technology Jobs, 2025). As institutional adoption grows, AI systems may become increasingly integrated into critical national infrastructure, posing significant opportunities in efficiency but also serious risks around the dependency and fragility of interconnected networks (RAND Corporation, 2025).

4.3. Access and autonomy granted to AI systems: As AI systems become increasingly capable of operating autonomously, a key uncertainty is how much access and autonomy users and organisations will choose to grant them. There are early signs of a shift towards granting AI systems broader execution capabilities, even in critical sectors such as finance. The AI Security Institute analysed usage data from over 1,000 public interfaces (MCP servers) that allow AI systems to access external tools and work as agents (AI Security Institute, 2025). They classified each server into one of five autonomy levels based on the tools and affordances it grants to models. Investigating finance-focused activity from December 2024 to July 2025, it increasingly observed new servers granting higher levels of autonomy to AI systems. Increasingly, AI systems can autonomously complete consequential actions like asset transfers and trading operations, rather than just reading and analysing data.

4.4. Societal impacts: The proliferation of advanced AI across communities, businesses, and broad society could bring a range of benefits, including increased economic growth and productivity; transformative health improvements; more efficient, accessible, and personalised public services; and more effective policymaking. However, it is also likely to cause significant disruption to our societal institutions and social fabric, from risks including misinformation; overreliance on AI systems; reduced accountability from automated decision-making; greater economic inequality; reducing ideological pluralism; or driving unsustainable energy use.

Most recently, there have been several high-profile cases of harm from the increasing numbers of people turning to AI systems for emotional support and social interaction (Rose Kirk et al, 2025; Time, 2025). While AI companionship may provide benefits for isolated individuals, such as the elderly, it also poses many risks. Prolonged reliance on AI interactions – optimised for user satisfaction – may attenuate social capacities relating to tolerance, empathy, and conflict resolution, on which human relationships depend. A growing number of people are also using AI for medical guidance, increasing risks from AI systems providing inaccurate or misleading information (Bean et al, 2026).

4.5. Public sentiment: Public sentiment towards AI is a key factor influencing adoption. The public currently associates a diverse range of hopes and fears with the future of AI. Whilst AI is recognised for its potential to positively impact sectors like healthcare and crime prevention, there are also concerns about its potential to cause job displacement and spread misinformation (Department for Science, Innovation and Technology, 2024; Ada Lovelace Institute, 2025). There is strong public support for regulating AI and instituting safeguards, with most Britons believing AI regulation is lagging the pace of technological change (Ipsos, 2024).

Relationship to other CUs	
Lower adoption	**Higher adoption**
Lower capability: Individuals and organisations are likely to make less use of less capable AI systems, though this may not always be the case, and usage would likely still be relatively high even if models are relatively lower capability. **Lower security:** Most individuals and organisations are likely to have less trust in systems if they have significant concerns about their security, though many may still use them regardless.	**Higher capability:** Individuals and organisations are likely to use more capable AI systems more. Higher security: The public and businesses are more likely to use systems that are considered secure. **Higher global cooperation:** Global cooperation on AI will help to avoid large-scale malicious use and build public trust. **Higher distribution and model access:** Highly capable and widely accessible open-weight models would likely enable broader adoption. However, adoption could also be very high in a highly concentrated, closed-weight driven market.

5. Labour displacement

5.1. AI exposure: AI exposure estimates the degree to which an occupation's tasks or required abilities align with the capabilities of AI systems – and the extent to which AI could perform, assist with, or accelerate those tasks. It therefore does not measure the extent to which roles will be automated. Around 60% of jobs in advanced economies are exposed to AI, with the UK's service-based economy one of the most exposed, at around 70% (Cazzaniga et al., 2024). Current data suggests that entry-level jobs and those more exposed to AI are experiencing relatively sharper hiring slowdowns, but robust causal evidence linking these trends to AI remains limited (Allas and Goodman, 2025; DSIT, 2025; Brynjolfsson et al., 2025). Evidence from Anthropic suggests that the actual use of AI remains a fraction of what it can feasibly achieve, demonstrating that the impact of AI exposure on labour market displacement is mediated by the level and intensity of AI adoption (Anthropic, 2026d).

5.2. Complementary vs. substitutive effects: AI exposure can be broken down into two categories. The first is "substitution exposure", which refers to tasks that AI could fully automate, at lower cost and without the need for human involvement. The second is "complementary exposure", which could involve AI automating some tasks in a job but not others, enabling workers to redistribute effort, or tasks where AI augments workers, collaborating closely with AI to perform tasks (Acemoglu, Autor and Johnson, 2026). The proportion of exposure which is substitution rather than complementary varies across roles. Well-structured tasks such as software programming tend to have higher substitution exposure: 79% of conversations on Claude Code were identified as automation and only 21% were identified as augmentation (Anthropic, 2025g). For roles which involve a high level of judgement, such as managers or government officials, AI exposure tends to be more complementary than substitution focused (Gmyrek et al., 2025).

5.3. Human preferences & institutional barriers: The impact of AI on labour displacement will also be influenced by human preferences. There is an observed tendency towards 'algorithm aversion', where humans prefer to avoid using AI even when it observably outperforms human counterparts (Filiz et al., 2023, Frank et al., 2024).). One study found that self-driving vehicles had to demonstrate a significantly higher level of safety than that of regular human-driven vehicles to be deemed as 'acceptable' (Liu, Wang and Vincent, 2020). Additionally, studies indicate a preference for human authorship, such that 90% of people preferred advice they believed came from a human expert rather than an AI, even if AI-advice was rated as having higher quality (Föyen et al., 2025). In addition to consumer concerns workers may resist automation. While workers support automating routine tasks to free up time for high value work, there are much lower levels of support for automating creative tasks. In addition, existing laws and regulatory frameworks are likely to act as barriers against autonomous AI systems capable of displacing labour (Shao et al., 2025).

5.4 Economic adjustments: The dynamics of labour demand will play a crucial role in determining the net impact of AI on employment. On one hand, labour-saving technologies could create offsetting increases in labour demand – an instance of the 'Jevons Paradox' whereby productivity gains reduce real prices which stimulates demand for goods and services (de Souza, 2025). On the other hand, demand effects could increase labour-displacement if rising unemployment and falling wages could trigger a decline in aggregate demand (Stiefenhofer, 2025). The extent to which demand effects promote stable-employment or displacement depends on several factors, including the magnitude of productivity gains, changes in real incomes, the distribution of income, and the nature and extent of government intervention.

Relationship to other CUs	
Lower labour displacement	**Higher labour displacement**
Lower capability: Less capable AI systems are likely to cause less labour displacement. Lower security: Less secure AI systems are likely to be given less access and autonomy and require greater human supervision, resulting in less labour displacement. **Lower adoption:** Lower adoption would limit labour displacement. Though even uneven levels of adoption could result in significant displacement	**Higher capability:** More capable AI systems are likely to cause more labour displacement **Higher security:** Secure AI systems are likely to be given greater access and autonomy and therefore require less human supervision, resulting in greater labour displacement. **Higher adoption:** Higher adoption would be required for significant labour displacement. Though labour displacement could still be lower even with higher adoption.

6. Global cooperation

6.1. International cooperation: International cooperation on AI governance is increasingly fragmented, particularly with the proliferation of open-source models reducing centralised control and subverting governance standards (Science Business, 2025). There have been positive signs of cooperation, including the Bletchley Declaration, the OECD AI Principles, and the Global Partnership on AI. However, competition and national security concerns have led more countries to assert control over AI infrastructure, talent, and governance frameworks (Centre for Geopolitics, JP Morgan Chase, 2025).

Only US or Chinese labs have the funds, talent, power and infrastructure to achieve AGI in the next five years (Federal Reserve Board, 2025). China is pursuing state-led self-reliance and lower cost open-source exports and is expanding its influence via BRICS+ and the Digital Silk Road (International Services Shanghai, 2025; East Asia Forum, 2025). While the US favours private-sector innovation, infrastructure buildout, and defence integration, and is forming tech bilateral alliances through technology prosperity deals (The White House, 2025b; UK Government, 2025a). Currently, the US has a decisive lead over China in terms of compute. However, China's unmatched energy production capacity will likely provide an advantage over longer timelines (Altman, 2025b).

The EU has made concerted moves towards tech sovereignty, striving to bolster their own AI capabilities and assert regulatory leadership (Carnegie Endowment for International Peace, 2025). The UK has made similar efforts, but has aligned more closely with the US, for instance, in being the only participating nation to join the US in declining to sign the joint principles of the 2025 Global AI Summit (BBC News, 2025). The Middle East is also building out compute hubs and India is positioning itself as a leader in the Global South.

6.2. Global conflict: Global tensions are at their highest level since the aftermath of World War II, making cooperation more difficult. The Global Peace Index has declined almost every year since 2014, signalling a structural shift toward a more violent and fragmented world (Institute for Economics and Peace, 2025; Peace Research Institute Oslo, 2025). AI itself is also becoming increasingly weaponised as it becomes more integrated into military systems, compressing decision cycles and enhancing battlefield autonomy, such as in Ukraine (Centre for Geopolitics, JP Morgan Chase, 2025). It is possible that an AI-based weapon system disrupts the military balance of power (Ding and Dafoe, 2021).

6.3. Globalised supply chains: AI infrastructure is now a critical strategic resource. While the US, EU, and China have recently invested heavily in domestic chip production, the world is increasingly dependent on Taiwan for its semiconductors. Taiwan produces over 90% of the most advanced semiconductor chips and just one of its companies (TSMC) occupies over 70%

of global market share (The Economist, 2023). Export controls have become key levers amid this level of concentration. Meanwhile, gulf states are leveraging cheap energy and capital to position themselves as AI hubs, while access to critical minerals, including rare earths and lithium, remains a challenge for resource-poor nations (Data Centre Magazine, 2025). The geographic concentration of the AI supply chain creates significant vulnerability to single points of failure and geopolitical tensions.

Relationship to other CUs	
Lower global cooperation	**Higher global cooperation**
Lower security: Low levels of cooperation will likely mean replication, confusion, or opposition around safety standards, with race dynamics possibly expediting the development of dangerous capabilities and encouraging a disregard for security. **Higher capability:** The development of high capability systems could drive an arms race dynamic and more fragmentation. Though it could also mobilise cooperation.	**Higher security:** Higher levels of cooperation would be required for global safety standards. However, low levels of security could mobilise cooperation. **Higher adoption:** Global cooperation on AI will help to avoid widespread malicious use and build public trust.

7. Underpinning economic assumptions

Table 7 below sets out the assumptions underpinning economic elements of the scenarios, bringing together elements of the above paragraphs on critical uncertainties.

Scenario	AI deployment capability & adoption	Supply side productivity & investment	Demand side growth & demand
Slow Burn	AI outperforms humans at a minority of cognitive tasks. AI deployment occurs incrementally, at a similar pace to electrification in the early 20th century. Only a small proportion of firms have successfully integrated AI.	Modest productivity growth; productivity gains are confined to early-adopting sectors. Modest impact on labour market churn.	**Modest growth uplift.** Consumer demand broadly stable, though some deflationary pressure from the real economic spillovers arising from a global market-correction in AI asset prices.
Open Frontier	AI outperforms humans at a minority of cognitive tasks. AI deployment occurs moderately fast, at a similar pace to internet adoption in the early 21st century. A sizeable minority of firms have integrated AI into their workflows.	Considerable aggregate productivity gains but high variance across firms and sectors. Considerable impact on labour-market churn.	**Considerable growth uplift.** Consumer demand under pressure from disruption of highly exposed sectors. However, wage and employment gains in some sectors enable economic growth to continue.

Augmented Growth	AI matches human-level cognitive performance across most knowledge tasks. AI is adopted at the pace of smartphones in the 2010s. Around half of businesses have adopted the technology.	Substantial economy-wide productivity growth from widespread business adoption of labour augmenting technology. Significant increase in investment in intangible assets and compute infrastructure.	**Economic boom.** Labour-augmenting productivity gains produce real wage increases for the majority, resulting in stable aggregate demand and employment despite labour-market churn.
Transformation Economy	AI matches human-level cognitive performance across most knowledge tasks. AI is adopted at an unprecedented pace and is used by the majority of businesses.	Substantial labour-saving efficiency gains for adopting businesses. However, economic gains for UK businesses are partially offset by the cost of accessing models resulting in profit leakage to overseas frontier AI firms.	**Growth is reduced.** Labour displacement from adoption of human-level AI models results in competitive pressure on wages and employment. This reduces consumer spending and destabilises aggregate demand.
Take-Off	AI exceeds human experts across virtually all cognitive tasks. AI is adopted at an unprecedentedly fast pace and is used by most businesses.	Transformative productivity gains without recent historical precedent. Fundamental restructuring of labour markets, and large-scale redistribution of workers. Large scale capital investment boom in UK AI infrastructure.	**Rapid growth with K-shaped distribution.** Productivity gains pool heavily at the top; aggregate demand is maintained by elite consumption and investment rather than broad wage growth. Rising prices for in-person services and manual labour.

Table 7: Assumptions underpinning economic elements of the scenarios

Annex C: Updates since the April 2025 report

New Name	Most similar scenario from Jan. 2024 report	Key Differences Relating to the CUs
Slow Burn	AI Disappoints	Slow Burn reflects more clearly how AI is unlikely to conclusively disappoint, even if the highest capabilities take longer to develop. For instance, more consideration is given to how new ways could be found to integrate and productise existing AI systems, even without radical improvements in capabilities. Access has been changed from 1 to 2 to reflect how open-weight systems would likely still be reasonably accessible given the focus of developers on productisation and open-weight systems being easier to fine-tune for specific uses and environments. Adoption has also been changed from 2 to 1 to reflect that this scenario shows the lowest plausible level of adoption that could be expected by 2030.
Open Frontier	AI Wild West	We have highlighted more of the opportunities highly capable and accessible open-weight systems would present. 'Capability' has also been lowered from 3 to 1 to show how this scenario represents a slowdown in AI progress, rather than continuation.
Augmented Growth	New scenario replacing Unpredictable Advanced AI	This is a new scenario, in which AI systems are highly capable but have not yet replaced most human workers given they still lack some key cognitive abilities. It also illustrates the possibility of significant progress on making AI systems safer. It has replaced "Unpredictable Advanced AI", which has been merged into "AI on a Knife Edge" (now called "Take-Off") given the overlap between these.
Transformation Economy	AI Disrupts the Workforce	'Adoption' has been raised from 3 to 5 given how many firms would need to adopt AI for widespread job displacement to occur.
Take-Off	AI on a Knife Edge (with themes from Unpredictable Advanced AI)	'Safety' has been lowered from 4 to 1 to more accurately reflect the key themes and to integrate themes from "Unpredictable Advanced AI", which had 1 on safety. Both previous scenarios covered misalignment risks, but this is now defined more clearly in this scenario. 'Global Cooperation' has been lowered from 2 to 1 to reflect the AI race being a key theme of this scenario.

References

AI Security Institute (UK Government) (2025) Frontier AI trends report. Available at: https://www.aisi.gov.uk/frontier-ai-trends-report (Accessed: 20 May 2026).

Acemoglu, D., Autor, D. and Johnson, S. (2026) Building Pro-Worker Artificial Intelligence. MIT Economics (PDF). Available at: https://economics.mit.edu/sites/default/files/2026-03/Building%20Pro-Worker%20Artificial%20Intelligence.pdf (Accessed: 20 May 2026).

Ada Lovelace Institute (2025) How do people feel about AI?. Available at: https://attitudestoai.uk/assets/documents/How-do-people-feel-about-AI-2025-Ada-Lovelace-Institute.pdf (Accessed: 20 May 2026).

Allas, T. and Goodman, A. (2025) Not yet productive, already disruptive: AI's uneven effects on UK jobs and talent. McKinsey & Company. Available at: https://www.mckinsey.com/uk/our-insights/the-mckinsey-uk-blog/ai-uneven-effects-on-uk-jobs-and-talent (Accessed: 20 May 2026).

Altman, S. (2025a) Reflections. Available at: https://blog.samaltman.com/reflections (Accessed: 20 May 2026).

Altman, S. (2025b) Thoughts on the AI buildout. Available at: https://blog.samaltman.com (Accessed: 20 May 2026).

Amazon (2025) Q1 2025 earnings call. Available at: https://ir.aboutamazon.com/events/event-details/2025/Q1-2025-Amazoncom-Inc-Earnings-Conference-Call-/default.aspx (Accessed: 20 May 2026).

Amazon Web Services (2025) Amazon Q Developer reaches $260 million milestone. Available at: https://aws.amazon.com/blogs (Accessed: 20 May 2026).

André, C., Bétin, M. and Gal, P. (2025) Dynamism in generative AI markets since the release of ChatGPT. CEPR. Available at: https://cepr.org/voxeu/columns/dynamism-generative-ai-markets-release-chatgpt (Accessed: 20 May 2026).

Anthropic (2025a) Activating AI Safety Level 3 (ASL 3) protections. May. Available at: https://www.anthropic.com/activating-asl3-report (Accessed: 20 May 2026).

Anthropic (2025b) Claude Code. Available at: https://code.claude.com/docs/en/overview (Accessed: 20 May 2026).

Anthropic (2025c) Project Vend: can Claude run a small shop? Available at: https://www.anthropic.com/research/project-vend-1 (Accessed: 20 May 2026).

Anthropic (2025d) Agentic misalignment. Available at: https://www.anthropic.com/research/agentic-misalignment (Accessed: 20 May 2026).

Anthropic (2025e) Disrupting AI-enabled espionage. Available at: https://www.anthropic.com/news/disrupting-AI-espionage (Accessed: 20 May 2026).

Anthropic (2025f) Detecting and countering misuse of AI: August 2025. Available at: https://www.anthropic.com/news/detecting-countering-misuse-aug-2025 (Accessed: 20 May 2026).

Anthropic (2025g) Anthropic Economic Index: AI's impact on software development. Anthropic Research. Available at: https://www.anthropic.com/research/impact-software-development (Accessed: 20 May 2026).

Anthropic (2026a) Pick up where you left off with Claude. Available at: https://www.youtube.com (Accessed: 20 May 2026).

Anthropic (2026b) Understanding Claude's personalization features. Available at: https://support.claude.com/en/articles/10185728-understanding-claude-s-personalization-features (Accessed: 16 May 2026).

Anthropic (2026c) Economic Index. Available at: https://www.anthropic.com/economic-index#country-usage (Accessed: 20 May 2026).

Anthropic (2026d) Labor market impacts of AI: A new measure and early evidence. Anthropic Research. Available at: https://www.anthropic.com/research/labor-market-impacts (Accessed: 20 May 2026).

Apart Research (2024) For-Profit AI Safety. Available at: https://apartresearch.com/news/ai-safety-needs-to-scale-and-heres-how-you-can-do-it (Accessed: 20 May 2026).

Artificial Analysis (2026a) AI model performance trends. Available at: https://artificialanalysis.ai/trends (Accessed: 20 May 2026).

Artificial Analysis (2026b) Language model inference price. Available at: https://artificialanalysis.ai/trends#language-model-inference-price (Accessed: 20 May 2026).

The Atlantic (2025) Elon Musk and Tesla Optimus. Available at: https://www.theatlantic.com/technology/2025/11/elon-musk-tesla-optimus/684968/ (Accessed: 20 May 2026).

Bai, Y. et al. (2022) Constitutional AI: Harmlessness from AI Feedback. Available at: https://arxiv.org/abs/2212.08073 (Accessed: 20 May 2026).

Balesni, M. et al. (2024) Towards evaluations-based safety cases for AI scheming. Available at: https://arxiv.org/abs/2411.03336 (Accessed: 20 May 2026).

BBC News (2025) UK and US refuse to sign international AI declaration. Available at: https://www.bbc.co.uk/news/articles/c8edn0n58gwo (Accessed: 20 May 2026).

BBC News (2026) AI firm Anthropic agrees to pay authors $1.5bnto settle piracy lawsuit. Available at: https://www.bbc.co.uk/news/articles/c5y4jpg922qo (Accessed: 20 May 2026).

Bean. A, Payne. R, Parsons. G, et al (2026) Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Available at: https://www.nature.com/articles/s41591-025-04074-y (Accessed: 20 May 2026).

Bengio, Y. (chair) et al. (2026) International AI Safety Report 2026. International AI Safety Report. February. Available at: https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026 (PDF: https://internationalaisafetyreport.org/sites/default/files/2026-02/international-ai-safety-report-2026_1.pdf) (Accessed: 20 May 2026).

Benton, G. et al. (2024) Sabotage evaluations for frontier models. Available at: https://www.anthropic.com/research/sabotage-evaluations (Accessed: 20 May 2026).

Bondarenko, Y. et al. (2025) Demonstrating specification gaming in reasoning models. Available at: https://arxiv.org/abs/2502.13295 (Accessed: 20 May 2026).

Brynjolfsson, E. et al. (2024) Generative AI at Work. Available at: https://arxiv.org/abs/2304.11771 (Accessed: 20 May 2026).

Brynjolfsson, E., Chandar, B. and Chen, R. (2025) Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. Stanford Digital Economy Lab (Working Paper). Available at: https://digitaleconomy.stanford.edu/publication/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence/ (Accessed: 20 May 2026).

Burnham, G. and Adamczewski, T. (2025) LLMs now accept longer inputs, and the best models can use them more effectively. Epoch AI. Available at: https://epoch.ai/data-insights/context-windows (Accessed: 20 May 2026).

Burton, J. (2023) Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0160791X23000672 (Accessed: 20 May 2026)

Carnegie Endowment for International Peace (2025) The EU's AI power play: between deregulation and innovation. Available at: https://carnegieendowment.org (Accessed: 20 May 2026).

Casper, S. et al. (2023) Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. Available at: https://arxiv.org/abs/2307.15217 (Accessed: 20 May 2026)

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E.J. and Tavares, M.M. (2024) Gen-AI: Artificial Intelligence and the Future of Work. International Monetary Fund (IMF) Staff Discussion Note No. 2024/001. Available at: https://www.imf.org/en/publications/staff-discussion-notes/issues/2024/01/14/gen-ai-artificial-intelligence-and-the-future-of-work-542379 (Accessed: 20 May 2026).

Centre for the Future of Intelligence (n.d.) ADELE: A dataset for evaluating long-term AI capabilities. Available at: https://kinds-of-intelligence-cfi.github.io/ADELE/ (Accessed: 20 May 2026).

Centre for Geopolitics, JP Morgan Chase (2025) The geopolitics of AI: decoding the new global operating system. Eurasia Group. Available at: https://www.jpmorganchase.com/content/dam/jpmorganchase/documents/center-for-geopolitics/decoding-the-new-global-operating-system.pdf (Accessed: 20 May 2026).

Cissé, A., Cooper, M.E., Zhu, M., Evangelopoulos, X. and Cooper, A.I. (2026) Can we automate scientific reasoning in closed-loop experiments using large language models? Digital Discovery. Available at: https://pubs.rsc.org/en/content/articlelanding/2025/dd/d5dd00520e (Accessed: 20 May 2026).

CNBC (2025a) Satya Nadella says up to 30% of Microsoft code is written by AI. Available at: https://www.cnbc.com/2025/04/29/satya-nadella-says-as-much-as-30percent-of-microsoft-code-is-written-by-ai.html (Accessed: 20 May 2026).

CNBC (2025b) Scale AI announces multimillion-dollar defence deal, a major step in U.S. military automation. Available at: https://www.cnbc.com/2025/03/05/scale-ai-announces-multimillion-dollar-defense-military-deal.html?msockid=172bd1a115326f6b1a27c48011326458 (Accessed: 20 May 2026).

CNBC (2025c) Report on OpenAI business user growth. Available at: https://www.cnbc.com/2025/06/04/openai-chatgpt-enterprise-ai.html?msockid=0d1007ee6c9a6d03391611976d1f6cdb (Accessed: 20 May 2026).

Cui, Z. et al. (2024) The productivity effects of generative AI: Evidence from a field experiment with GitHub Copilot. Available at: https://mit-genai.pubpub.org/pub/v5iixksv/release/2 (Accessed: 20 May 2026).

Data Centre Magazine (2025) How the Middle East data centre sector can spur AI dominance. Available at: https://www.datacentremagazine.com (Accessed: 20 May 2026).

DeepMind (2024) AlphaFold. Available at: https://deepmind.google/science/alphafold/ (Accessed: 20 May 2026).

Dell'Acqua, F. et al. (2023) Navigating the jagged technological frontier. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321 (Accessed: 20 May 2026).

Denison, C. et al. (2024) Sycophancy to Subterfuge: Investigating Reward-Tampering in Large Language Models. Available at: https://arxiv.org/abs/2406.10162 (Accessed: 20 May 2026).

Department for Science, Innovation and Technology (2024) Public attitudes to data and AI: Tracker survey (Wave 4) report. Available at: https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-4/public-attitudes-to-data-and-ai-tracker-survey-wave-4-report (Accessed: 20 May 2026).

Department for Science, Innovation & Technology (2025) Economic Estimates: Employment in the Digital Sector, January 2024 to December 2024. GOV.UK. Available at: https://www.gov.uk/government/statistics/economic-estimates-employment-in-the-digital-sector-january-2024-to-december-2024/economic-estimates-employment-in-the-digital-sector-january-2024-to-december-2024 (Accessed: 20 May 2026).

Department for Science, Innovation and Technology (2026) AI adoption research. Available at: https://www.gov.uk/government/publications/ai-adoption-research/ai-adoption-research (Accessed: 27 May 2026).

de Souza, G. (2025) Artificial intelligence in the office and the factory: Evidence from administrative software registry data. CEPR / Federal Reserve Bank of Chicago Working Paper WP 2025-11. Available at: https://cepr.org/voxeu/columns/artificial-intelligence-office-and-factory-evidence-administrative-software-registry (Accessed: 20 May 2026).

Ding J. Dafoe, A (2021) Engines of Power: Electricity, AI, and General-Purpose Military Transformations. Available at: https://arxiv.org/pdf/2106.04338 (Accessed: 20 May 2026).

East Asia Forum (2025) China expands AI globally through the Digital Silk Road; Available at: https://eastasiaforum.org/2025/04/11/china-expands-ai-globally-through-the-digital-silk-road/ (Accessed: 20 May 2026)

The Economist (2023) Taiwan's dominance of the chip industry makes it more important. The Economist. Available at: https://www.economist.com/special-report/2023/03/06/taiwans-dominance-of-the-chip-industry-makes-it-more-important (Accessed: 20 May 2026).

Engineering and Technology Jobs (2025) Surge in demand for AI professionals sends job vacancies soaring. Available at: https://engineering-jobs.theiet.org/article/surge-in-demand-for-ai-professionals-sends-job-vacancies-soaring- (Accessed: 20 May 2026).

Epoch AI (2023) How predictable is language model benchmark performance? Available at: https://epoch.ai/blog/how-predictable-is-language-model-benchmark-performance (Accessed: 20 May 2026).

Epoch AI (2025a) AI companies' revenue. Available at: https://epoch.ai/data-insights/ai-companies-revenue (Accessed: 27 May 2026).

Epoch AI (2025b) What you need to know about AI data centres. Available at: https://epoch.ai/blog/what-you-need-to-know-about-ai-data-centers (Accessed: 20 May 2026).

Epoch AI (2025c) AI in 2030-Extrapolatingcurrent trends. Available at: https://www.searchyour.ai/en/ai-in-2030-extrapolating-current-trends (Accessed: 20 May 2026).

Epoch AI (2025d) Open weights vs closed weights models. Available at: https://epoch.ai/data-insights/open-weights-vs-closed-weights-models (Accessed: 20 May 2026).

Epoch AI (2025e) How much does it cost to train frontier AI models? Available at: https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models (Accessed: 20 May 2026).

Epoch AI (2026a) MirrorCode: Evidence that AI can already do some weeks long coding tasks. 10 April. Available at: https://epoch.ai/blog/mirrorcode-preliminary-results/ (Accessed: 20 May 2026).

Epoch AI (2026b) Data on AI companies. Updated 28 April. Available at: https://epoch.ai/data/ai-companies (Accessed: 20 May 2026).

Epoch AI (2026c) AI Benchmarking Hub. Available at: https://epoch.ai/benchmarks (Accessed: 20 May 2026).

Epoch AI (2026d) GPU clusters database. Available at: https://epoch.ai/data/gpu-clusters?view=map&tab=point (Accessed: 20 May 2026).

Epoch AI (2026e) Trends in Artificial Intelligence. Epoch AI. Available at: https://epoch.ai/trends (Accessed: 20 May 2026).

Edelman, Y and Ho, A (2025) Compute scaling will slow down due to increasing lead times. Available at: https://epoch.ai/gradient-updates/compute-scaling-will-slow-down-due-to-increasing-lead-times (Accessed: 20 May 2026).

Federal Reserve Board (2025) The state of AI competition in advanced economies. Available at: https://www.federalreserve.gov/econres/notes/feds-notes/the-state-of-ai-competition-in-advanced-economies-20251006.html (Accessed: 20 May 2026).

Filiz, I., Judek, J.R., Lorenz, M. and Spiwoks, M. (2023) The extent of algorithm aversion in decision-making situations with varying gravity. PLOS ONE, 18(2). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9942970/ (Accessed: 20 May 2026).

Föyen, L.F., Zapel, E., Lekander, M., Hedman-Lagerlöf, E. and Lindsäter, E. (2025) Artificial intelligence vs. human expert: Licensed mental health clinicians' blinded evaluation of AI-generated and expert psychological advice on quality, empathy, and perceived authorship. Internet Interventions, 41, 100841. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC12169703/ (Accessed: 20 May 2026).

Frank, D.-A., Chrysochou, P., Mitkidis, P., Otterbring, T. and Ariely, D. (2024) Navigating uncertainty: Exploring consumer acceptance of artificial intelligence under self-threats and high-stakes decisions. Technology in Society, 79. Available at: https://doi.org/10.1016/j.techsoc.2024.102732 (Accessed: 20 May 2026).

Gade, P. et al. (2024) BadLlama: cheaply removing safety fine-tuning from Llama 2-Chat 13B. Available at: https://arxiv.org/abs/2311.00117 (Accessed: 20 May 2026).

Gemini Team, Google (2023) Gemini: A family of highly capable multimodal models. arXiv preprint arXiv:2312.11805. Available at: https://arxiv.org/abs/2312.11805 (Accessed: 20 May 2026).

Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K. and Troszyński, M. (2025) Generative AI and Jobs: A Refined Global Index of Occupational Exposure. International Labour Organization. Available at: https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure (Accessed: 20 May 2026).

Goldman Sachs (2024) AI: to buy, or not to buy, that is the question. Available at: https://www.goldmansachs.com/insights/goldman-sachs-research/global-strategy-paper-ai-to-buy-or-not-to-buy-that-is-the-question (Accessed: 20 May 2026).

Google Research (2025) Introducing nested learning: a new ML paradigm for continual learning. Available at: https://research.google/blog/introducing-nested-learning-a-new-ml-paradigm-for-continual-learning/ (Accessed: 20 May 2026).

Government Office for Science (2024) Futures toolkit for policymakers and analysts. Available at: https://www.gov.uk/government/publications/futures-toolkit-for-policy-makers-and-analysts (Accessed: 20 May 2026).

Government Office for Science (2025) AI 2030 scenarios. Available at: https://assets.publishing.service.gov.uk/media/6808fc002a86d6dfb2b52772/AI_2030_Scenarios_Report.pdf (Accessed: 20 May 2026).

Greenblatt, R. et al. (2023) AI Control: Improving Safety Despite Intentional Subversion. Available at: https://arxiv.org/abs/2312.06942 (Accessed: 20 May 2026)

Guan, M. et al. (2024) Deliberative Alignment: Reasoning Enables Safer Language Models. Available at: https://arxiv.org/abs/2412.16339 (Accessed: 20 May 2026).

Gundlach, H., Lynch, J., Mertens, M. and Thompson, N. (2025) The Price of Progress: Algorithmic Efficiency and the Falling Cost of AI Inference. arXiv. Available at: https://arxiv.org/abs/2511.23455 (Accessed: 20 May 2026).

Hadfield-Menell, D. et al. (2016) The Off-Switch Game. Available at: https://arxiv.org/abs/1611.08219 (Accessed: 20 May 2026).

Hassabis, D., Kumaran, D., Summerfield, C. and Botvinick, M. (2017) Neuroscience-inspired artificial intelligence. Neuron, 95(2), pp. 245–258. Available at: https://doi.org/10.1016/j.neuron.2017.06.011 (Accessed: 20 May 2026).

Heitmann, M., Hinrichsen, T., Africa, D. and Sandbrink, J. (2025) Understanding AI trajectories: mapping the limitations of current AI systems. UK AI Security Institute. Available at: https://www.aisi.gov.uk/research/understanding-ai-trajectories-mapping-the-limitations-of-current-ai-systems (Accessed: 20 May 2026).

Ho, A and Berg, A (2025) Quantifying the algorithmic improvement from reasoning models. Available at: https://epoch.ai/gradient-updates/quantifying-the-algorithmic-improvement-from-reasoning-models (Accessed: 20 May 2026).

Hobbs, H. et al. (2026), "Exploring possible AI trajectories through 2030", OECD Artificial Intelligence Papers, No. 55, OECD Publishing, Paris, https://doi.org/10.1787/cb41117a-en (Accessed: 27 May 2026).

Hoffmann, J. et al. (2022) Training compute-optimal large language models. Available at: https://arxiv.org/abs/2203.15556 (Accessed: 20 May 2026).

Hu, Y. et al. (2023) Toward general-purpose robots via foundation models: a survey and meta-analysis. Available at: https://arxiv.org/abs/2312.08782 (Accessed: 20 May 2026).

Hugging Face (2025) Vectara hallucination evaluation model. Available at: https://huggingface.co/vectara/hallucination_evaluation_model (Accessed: 20 May 2026).

The Independent (2025) Google Gemini update adds memory features. Available at: https://www.independent.co.uk/tech/google-gemini-update-ai-memory-b2807453.html (Accessed: 20 May 2026).

Institute for Economics and Peace (2025) Global Peace Index 2025. Available at: https://www.visionofhumanity.org/wp-content/uploads/2025/06/Global-Peace-Index-2025-web.pdf (Accessed: 20 May 2026).

International Services Shanghai (2025) Key alliances launched to boost China–BRICS AI cooperation; Available at: https://english.shanghai.gov.cn/en-WAICLatest/20250729/eac167d7c7c6452bb57f05440b6e6345.html (Accessed: 20 May 2026).

Ipsos (2024) Public trust in AI. Available at: https://www.ipsos.com/sites/default/files/ct/news/documents/2024-09/Ipsos%20Public%20Trust%20in%20AI.pdf (Accessed: 20 May 2026).

Jarviniemi, J. and Hubinger, E. (2024) Uncovering Deceptive Tendencies in Language Models: A Simulated Company AI Assistant. Available at: https://arxiv.org/abs/2405.01576 (Accessed: 20 May 2026)

Kaplan, J. et al. (2020) Scaling laws for neural language models. Available at: https://arxiv.org/abs/2001.08361 (Accessed: 20 May 2026).

Krakovna, V. et al. (2023) Power-seeking can be probable and predictive for trained agents. Available at: https://arxiv.org/pdf/2304.06528 (Accessed: 20 May 2026).

Kwa, T., West, B., Becker, J. et al. (2025) Measuring AI Ability to Complete Long Software Tasks. arXiv preprint arXiv:2503.14499. Available at: https://arxiv.org/abs/2503.14499 (Accessed: 20 May 2026).

Lambert, N. et al. (2023) The Alignment Ceiling: Objective Mismatch in Reinforcement Learning from Human Feedback. Available at: https://arxiv.org/abs/2311.00168 (Accessed: 20 May 2026)

Lewis, P. et al. (2020) Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Processing Systems, 33, pp. 9459–9474. (Accessed: 20 May 2026).

Liu, P., Wang, L. and Vincent, C. (2020) Self-driving vehicles against human drivers: Equal safety is far from enough. Journal of Experimental Psychology: Applied, 26(4), pp. 692–704. Available at: https://ora.ox.ac.uk/objects/uuid:d39c1bf4-cf54-4c1e-9a1d-d988376f50d0 (Accessed: 20 May 2026).

LMSYS (2026) Chatbot Arena leaderboard. Available at: https://lmarena.ai/leaderboard (Accessed: 20 May 2026).

MacDiarmid, M. et al (2025) Natural emergent misalignment from reward hacking. Available at: https://www-cdn.anthropic.com/daad4360a8bdc707f8b22e3e745796ba27e57fb3.pdf (Accessed: 20 May 2026)

McKinsey & Company (2025) The cost of compute power: A $7 trillion race to scale data centers. McKinsey & Company. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers (Accessed: 20 May 2026).

Meta Platforms (2025) Q2 2025 earnings call. Available at: https://investor.atmeta.com/investor-events/event-details/2025/Q2-2025-Earnings-Call/default.aspx (Accessed: 20 May 2026).

Microsoft (2024) Q4 2024 earnings call. Available at: https://www.microsoft.com/en-us/investor/events/fy-2024/earnings-fy-2024-q4?msockid=0d1007ee6c9a6d03391611976d1f6cdb (Accessed: 20 May 2026).

Mirzadeh, I. et al (2025) GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models. arXiv preprint arXiv:2410.05229. Available at: https://arxiv.org/abs/2410.05229 (Accessed: 20 May 2026)

Model Evaluation & Threat Research (METR) (2025) How does time horizon vary across domains? Available at: https://metr.org/blog/2025-07-14-how-does-time-horizon-vary-across-domains/ (Accessed: 20 May 2026).

Model Evaluation & Threat Research (METR) (2026) Task completion time horizons of frontier AI models. Updated 15 April. Available at: https://metr.org/time-horizons/ (Accessed: 20 May 2026).

Morris, M. et al. (2024) Levels of AGI for Operationalizing Progress on the Path to AGI. Available at: https://arxiv.org/abs/2311.02462 (Accessed: 20 May 2026)

Mu, J. (n.d.) Statement on AI-driven software engineering (post on X). Available at: https://x.com (Accessed: 20 May 2026).

National Bureau of Economic Research (2025) How people use ChatGPT (Working Paper w34255). Available at: https://www.nber.org/papers/w34255 (Accessed: 20 May 2026).

National Cyber Security Centre (2025) The impact of AI on cyber threat (now to 2027). Available at: https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027 (Accessed: 20 May 2026).

Newman, S. (2025) GPT-5: the case of the missing agent. Available at: https://secondthoughts.ai/p/gpt-5-the-case-of-the-missing-agent (Accessed: 20 May 2026).

The New York Times (2025) OpenAI Uncovers Evidence of A.I.-Powered Chinese Surveillance Tool. Available at: https://www.nytimes.com/2025/02/21/technology/openai-chinese-surveillance.html?unlocked_article_code=1.zE4.VE1v.QxVmvFuBeQtX&smid=url-share (Accessed: 20 May 2026).

Ngo, R. et al. (2022) The Alignment Problem from a Deep Learning Perspective. Available at: https://arxiv.org/abs/2209.00626 (Accessed: 20 May 2026).

Noy, S. and Zhang, W. (2023) Experimental evidence on the productivity effects of generative AI. Available at: https://www.science.org/doi/10.1126/science.adh2586 (Accessed: 20 May 2026).

Office for National Statistics (2026) Business Insights and Conditions Survey (BICS), January 2026. Available at: https://www.ons.gov.uk/economy/economicoutputandproductivity/output/

OpenAI (2025a) GPT-4.5 system card. Available at: https://openai.com/index/gpt-4-5-system-card/ (Accessed: 20 May 2026).

OpenAI (2025b) o1 preview. Available at: https://openai.com/index/introducing-openai-o1-preview/ (Accessed: 20 May 2026).

OpenAI (2025c) ChatGPT Agent. Available at: https://openai.com/index/introducing-chatgpt-agent/ (Accessed: 20 May 2026).

OpenAI (2025d) Introducing GPT-OSS. Available at: https://openai.com/index/introducing-gpt-oss/ (Accessed: 20 May 2026).

OpenAI (2026) Memory FAQ. Available at: https://help.openai.com/en/articles/8590148-memory-faq (Accessed: 20 May 2026).
Our World in Data (2024) Mobile phone subscriptions data. Available at: https://ourworldindata.org/grapher/number-of-internet-users (Accessed: 20 May 2026).

Our World in Data (2026) External funding for privately held AI companies raising above $1.5 million. Our World in Data. Available at: https://ourworldindata.org/grapher/private-investment-in-artificial-intelligence (Accessed: 20 May 2026).

Peace Research Institute Oslo (2025) New data shows conflict at historic high as U.S. signals retreat from world stage. Available at: https://www.prio.org/news/3616 (Accessed: 20 May 2026).

Penadés, J.R., Gottweis, J., He, L. et al. (2025) 'AI mirrors experimental science to uncover a mechanism of gene transfer crucial to bacterial evolution', Cell, 188(23), pp. 6654–6665.e2. Available at: https://www.cell.com/cell/fulltext/S0092-8674(25)00973-0 (Accessed: 20 May 2026).

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023) The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv preprint arXiv:2302.06590. Available at: https://arxiv.org/abs/2302.06590 (Accessed: 27 May 2026).

Rahman, R. and Owen, D. (2024) The training compute of notable AI models has been doubling roughly every six months. Epoch AI. Available at: https://epoch.ai/data-insights/compute-trend-post-2010 (Accessed: 20 May 2026).

RAND Corporation (2025) The United States needs to stress test critical infrastructure for different AI adoption scenarios. Available at: https://www.rand.org/pubs/commentary/2025/01/the-united-states-needs-to-stress-test-critical-infrastructure.html (Accessed: 20 May 2026).

RBC Wealth Management (2026) Big tech's AI expansion: from investment to scalable returns. Available at: https://www.rbcwealthmanagement.com/en-us/insights/big-techs-ai-expansion-from-investment-to-scalable-returns (Accessed: 27 May 2026).

Ren, J. et al. (2025) The MASK Benchmark: Disentangling Honesty From Accuracy in AI Systems. Available at: https://arxiv.org/abs/2503.03750 (Accessed: 20 May 2026).

Research and Markets (2025) Embodied AI and humanoid robot market research 2024–2025. Available at: https://www.researchandmarkets.com/reports/6078204/embodied-ai-humanoid-robot-market-research (Accessed: 20 May 2026).

Ritchey, T. (2011) 'General Morphological Analysis (GMA)'. In: Ritchey, T. Wicked problems – social messes: decision support modelling with morphological analysis. Berlin: Springer, pp. 7–18. Available at: https://link.springer.com/chapter/10.1007/978-3-642-19653-9_2 (Accessed: 20 May 2026).

Rose Kirk. H et al (2025) Neural steering vectors reveal dose and exposure-dependent impacts of human-AI relationships. Available at: https://arxiv.org/abs/2512.01991 (Accessed: 20 May 2026).

Scheurer, J. et al. (2023) Large Language Models can Strategically Deceive their Users when Put Under Pressure. Available at: https://arxiv.org/abs/2311.07590 (Accessed: 20 May 2026)

Science Business (2025) Experts warn of fragmentation in global collaboration for innovation. Available at: https://sciencebusiness.net/news/synergies/experts-warn-fragmentation-global-collaboration-innovation#:~:text=The%20geopolitical%20context%2C%20changes%20in,us%20more%2C%E2%80%9D%20Flanagan%20said (Accessed: 20 May 2026).

Shao, Y., Zope, H., Jiang, Y., Pei, J., Nguyen, D., Brynjolfsson, E. and Yang, D. (2025) Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce. arXiv. Available at: https://arxiv.org/abs/2506.06576 (Accessed: 20 May 2026).

Smaldone, A.M. et al. (2025) Quantum machine learning in drug discovery: applications in academia and pharmaceutical industries. Review article, 6 June. Available at: https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.4c00678?utm_source=chatgpt.com (Accessed: 20 May 2026)

Stiefenhofer, P. (2025) The Future of Work and Capital: Analyzing AGI in a CES Production Model. arXiv. Available at: https://arxiv.org/abs/2502.07044 (Accessed: 20 May 2026).

Tech in Asia (2025) China's Agibot targets 5,000 robots this year. Available at: https://www.techinasia.com/news/chinas-agibot-targets-142m-revenue-ceo-says (Accessed: 20 May 2026).

TechCrunch (2025a) Sam Altman says ChatGPT has hit 800m weekly active users. Available at: https://techcrunch.com/2025/10/06/sam-altman-says-chatgpt-has-hit-800m-weekly-active-users/ (Accessed: 20 May 2026).

TechCrunch (2025b) Sam Altman says OpenAI will have a legitimate AI researcher by 2028. Available at: https://techcrunch.com/2025/10/28/sam-altman-says-openai-will-have-a-legitimate-ai-researcher-by-2028/ (Accessed: 20 May 2026).

TechSpot (2025) Nvidia reaches historic 92% GPU market share. Available at: https://www.techspot.com/news/108225-nvidia-reaches-historic-92-gpu-market-share-leaves.html (Accessed: 20 May 2026).

Time (2025) Parents allege ChatGPT is responsible for their teenage son's death by suicide. Available at: https://time.com/7312484/chatgpt-openai-suicide-lawsuit/ (Accessed: 20 May 2026).

Toner, H. (2025). Taking jaggedness seriously: Why we should expect AI capabilities to remain extremely uneven. Rising Tide (Substack). Available at: https://helentoner.substack.com/p/taking-jaggedness-seriously (Accessed: 20 May 2026).

Turner, A. et al. (2023) Optimal Policies Tend to Seek Power. Available at: https://arxiv.org/abs/1912.01683 (Accessed: 20 May 2026).

UK AI Security Institute (2025) Automation of AI R&D (AISI Deep Dive AISI_2025_07_24_DD_01). Available at: Author website (Accessed: 20 May 2026).

UK Government (2025a) US–UK pact will boost advances in drug discovery, create tens of thousands of jobs and transform lives. Available at: https://www.gov.uk/government/news/us-uk-pact-will-boost-advances-in-drug-discovery-create-tens-of-thousands-of-jobs-and-transform-lives (Accessed: 20 May 2026).

UK Government (2025b) International AI safety report 2025. Available at: https://www.gov.uk/government/publications/international-ai-safety-report-2025 (Accessed: 20 May 2026).

UK Government (2025c) International AI safety report 2025. Available at: https://internationalaisafetyreport.org/sites/default/files/2025-10/international_ai_safety_report_2025_english.pdf (Accessed: 20 May 2026).

Vestager, M., Cardell, S., Kanter, J. and Khan, L. (2024) Joint statement on competition in generative AI foundation models and AI products. Federal Trade Commission. Available at: https://www.ftc.gov/legal-library/browse/joint-statement-competition-generative-ai-foundation-models-ai-products (Accessed: 20 May 2026).

Vinyals, O. (2025) Social media post on AI systems. Available at: https://x.com/OriolVinyalsML/status/1990854455802343680 (Accessed: 20 May 2026).

The White House (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Executive Office of the President, 30 October. https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (Accessed: 20 May 2026).

The White House (2025a). America's AI Action Plan. Executive Office of the President, July. https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (Accessed: 20 May 2026).

The White House (2025b) The United States signs technology prosperity deals with Japan and Korea. Available at: https://www.whitehouse.gov/articles/2025/10/the-united-states-signs-technology-prosperity-deals-with-japan-and-korea/ (Accessed: 20 May 2026).

Whitfill, J. et al. (2025) Forecasting AI Time Horizon Under Compute Slowdowns. Available at: https://joel-becker.com/images/publications/forecasting_time_horizon_under_compute_slowdown.pdf (Accessed: 20 May 2026)

Zheng, X. et al. (2025) Lifelong Learning of Large Language Model based Agents: A Roadmap. Available at: https://arxiv.org/abs/2501.07278 (Accessed: 20 May 2026).

Government Office for Science

Foresight

foresight@go-science.co.uk