Microsoft

Governing AI

A blueprint for the UK

AI

Governing AI

A blueprint for the UK

Contents

04 The promise of AI
05 Foreword by Clare Barclay

06 AI explained
07 What exactly is AI?
08 AI-powered possibility
09 AI at work
10 A brief history of AI: From Turing to ChatGPT
12 Developing and scaling AI

14 AI in action
16 AI for a healthier future
19 AI for a more sustainable future
22 AI for a growing future in agriculture
24 AI for smarter manufacturing
26 AI for better education
29 The value of AI

30 Responsible development and use of AI
32 Responsible AI – Q&A with Natasha Crampton
35 Microsoft AI Red Team building future of safer AI
36 The Frontier Model Forum
37 Helping our UK customers develop and deploy AI responsibly – Antony Cook
40 Microsoft's new Customer Copyright Commitment
42 Our commitments to advance safe, secure and trustworthy AI
44 Our investment in AI infrastructure, skills and security

46 AI governance
47 Advancing AI governance – Brad Smith
48 A five-point blueprint for governing AI
54 Protecting elections
56 International partnership to advance AI governance

The promise of AI

Foreword

Clare Barclay, Chief Executive Officer, Microsoft UK

Artificial intelligence is a vital technology that will define our future. It has the potential to solve society's biggest challenges, revolutionise the way we work and – in line with our mission – can help every person and every organisation on the planet achieve more.

As one of the most digitally advanced economies in the world, the UK is well placed to lead the way in AI development and deployment. The home of innovative businesses, research institutions and world leading universities, the UK is quick to adopt emerging technologies and use them to solve real-world challenges. For example, AI is helping improve patient outcomes in the National Health Service; monitor puffin and salmon populations in Scotland; and accelerate clinical research.

At Microsoft, we are optimistic about what AI can do for people, industry and society. We are also acutely aware that if you create technologies with the ability to change the world, you have to do so responsibly. For that reason, we have invested to make sure that our AI systems are responsible by design. To guide our work, we have a set of core principles: fairness, reliability and safety, privacy and security, inclusiveness, transparency and accountability. We've put those principles into practice by developing and implementing our Responsible AI Standard and related tools like our Impact Assessment. And we also share our learnings externally, including our Standard and Impact Assessment template, to contribute to broader societal conversations about AI.

We are not the only ones considering how to democratise access to new technologies and minimise any risks that emerge as we deploy them. Governments around the world want to reap the benefits of AI and do so in a responsible way. They know that there isn't a single country, business or individual that has all the answers.

> "We have a set of core principles: fairness, reliability and safety, privacy and security, inclusiveness, transparency and accountability"

Creating the right regulatory framework starts with asking the right questions: about fairness, bias, inclusion, privacy and we need a wide range of voices in the room to do that. As such, we applaud the UK Government's decision to host the first ever global summit on AI safety, and we were delighted to sign up to the White House commitments in 2023 and provide an update on our policies alongside the summit. Working together, we can set out the international 'rules of the road' on how to develop and deploy AI technologies safely.

Technology has the potential to change people's lives for the better, and the era of AI – the latest in a long line of innovations from the railway to the internet – is no exception to this. The challenge will be to harness AI to deliver greater prosperity across the UK so that its benefits are shared by all. Only by working together, pooling cross-sector and cross-societal expertise, will we be able to realise the promise of AI.

AI explained

AI is changing how we work, live, and play. Microsoft AI provides billions of intelligent experiences every day to people using our business software and services, gaming on our Xbox platform and keeping organisations secure.

Our AI tools and technologies are designed to benefit everyone at every level in every organisation. They're already being used across the public sector and in workplaces, home offices, academic institutions, research labs, and manufacturing facilities around the world. They're helping everyone from scientists and salespeople to farmers, software developers, and security practitioners.

What exactly is AI?

Using maths and logic, an artificially intelligent computer system simulates the reasoning that humans use to learn from data and make decisions.

Algorithms are at the core of every AI system. They are sets of instructions or rules that the system follows to process and analyse data. Different algorithms are used for different tasks, such as recognising images or processing and predicting language.

An AI model is trained on a dataset representative of the task or problem it's being developed to solve.

Data could be in the form of unstructured data, such as text, images, or audio, or structured data that can be stored in a standardised format.

The AI model uses the patterns and relationships it discovers in the data to adjust and optimise its performance resulting in a learned representation of the training data, which can be used to interpret and make inferences from new input or data.

One of the techniques used to train AI is machine learning. It's akin to teaching a computer program to recognise patterns by showing it examples. Just as we learn from experience, machine learning algorithms learn from data.

> "AI development has been inspired by the structure and functioning of the human brain"

AI models look for similarities and relationships in the data to perform tasks such as making predictions, recommendations, classifications and more.

Imagine you want to train a computer program to recognise cats. You would feed a machine learning algorithm a large number of pictures, each labelled as to whether the image contains a cat or not. By analysing these labelled examples, the algorithm can identify patterns such as shapes, colours, and textures associated with cats, and create a model based on this learning. Once trained, the model can then be used to analyse new, unseen images and predict whether they contain a cat or not. The more examples the algorithm sees, the better the model becomes at recognising cats accurately.

A subset of machine learning under the broader umbrella of AI is reinforcement learning. It involves asking an AI system to make a sequence of decisions to maximise a reward. The AI system learns through trial and error, receiving feedback in the form of rewards – or penalties – based on its choices. By learning which actions lead to higher rewards, the AI system improves its decision-making over time.

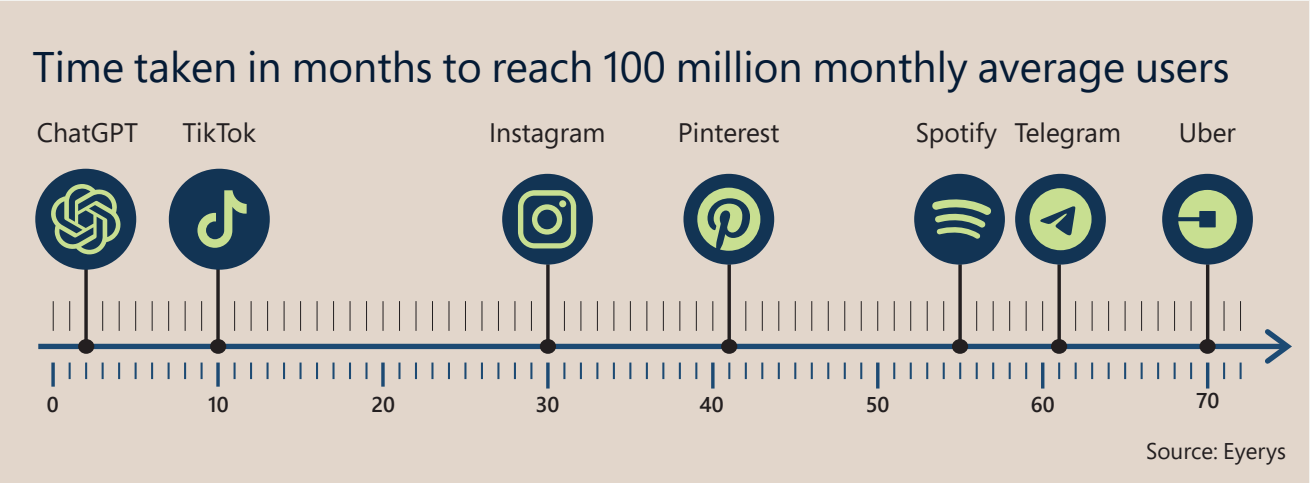
AI development has also been inspired by the structure and functioning of the human brain. Neural networks are computational models that consist of interconnected nodes called neurons, organised in layers. Each neuron receives inputs, performs a computation, and passes the result to the next layer. Neural networks excel at pattern recognition and can learn complex relationships in data.

As you can see, AI is a wide umbrella under which sit a range of different but complementary computer science fields and sub-fields – many of which have been in research for decades. But whichever approach is used to develop an AI system, having access to data and computing power to train your AI model are fundamental building blocks.

AI-powered possibility

AI enables machines to carry out a wide range of tasks, many of which were not possible before its advent.

Natural language processing (NLP), for example, extracts meaning from text or speech data. It enables computers to interpret language and perform tasks like sentiment analysis, language translation, and text generation. AI-powered Computer Vision similarly enables machines to "see" and decode the visual world around us. It involves techniques such as object detection and image recognition, and has a wide range of applications from medical imaging to autonomous vehicles.



These capabilities and more are transforming how we use, benefit from, and interact with computer technology. AI can serve as a "copilot" to augment our human ingenuity and creativity through automating tasks and providing new insight. Over the last decade, the AI field has made significant progress on perception and language tasks. There have been recent advances in the form of generative AI – a class of AI models that can generate new content such as text, images, code, and more. Generative AI is underpinned by a class of large-scale models known as foundation models.

Foundation models are trained on massive amounts of data and are capable of performing a wide range of tasks. With a simple prompt like "describe a scene of the sun rising over the beach", generative AI models can create a detailed description or produce an image, which can then be animated or even turned into video.

Many recent generative AI models such as language models are not only good at generating text but also at generating, explaining, and debugging code. GitHub Copilot, for example, leverages OpenAI's Codex model to assist developers in writing code.

Another example of AI systems in action is Microsoft 365 Copilot, an AI-powered workplace productivity tool. Microsoft 365 Copilot is embedded in Microsoft apps like Word, Excel, PowerPoint, Outlook and Teams. It works alongside the user, helping boost productivity and unleash creativity. Copilot in PowerPoint, to name one example, enables the creation process by turning the user's ideas into a designed presentation.

Another example of AI systems in action is the new AI-powered Bing search engine. The web search experience often involves the time-consuming task of reviewing and synthesising information from a variety of sources identified from different search queries. Now, Bing can do the heavy lifting for you, working behind the scenes to make the necessary queries, collect results, synthesise the information, and present a single complete answer.

AI at work

AI is set to transform the entire knowledge economy, helping workers with repetitive aspects of their work and allowing us more time to be creative, which can make work generally more pleasant and fulfilling. Not only will AI help with tasks like writing, editing or creating reports, it is going to apply to almost anything – designing new molecules to create medicine, or making manufacturing "recipes" from 3D models. AI has the potential to be a "copilot for everything" – a copilot that assists any kind of cognitive work, not just helping us to achieve more, but also enhancing creativity in new and exciting ways.

AI also promises to be a powerful tool to boost the productivity of workers, with our latest Work Trend Index showing that using AI to increase employee productivity is the feature of AI business leaders are most excited about. Their enthusiasm is shared by employees surveyed for our Work Trend Index – 70% of employees said they are excited about the prospect of AI helping them ease their workloads, and an even greater share of employees report interest in using AI for analytical (79%) and even creative work (73%). In short, there is widespread optimism for how AI can make our working lives more fulfilling and productive.

Existing research indicates that this optimism is justified – according to one report, generative AI could enable global labour productivity growth of 0.1 to 0.6 percent annually through to 2040.

Who will be benefiting the most from the integration of AI in the workplace? In a departure from previous technology tools – even recent ones, such as the spreadsheet – which have disproportionately benefited higher-skilled workers, current research suggests that the largest productivity gains are, in fact, seen among the lowest-skilled workers. One study found that access to AI tools increases productivity among novice and low-skilled workers by 14% on average.

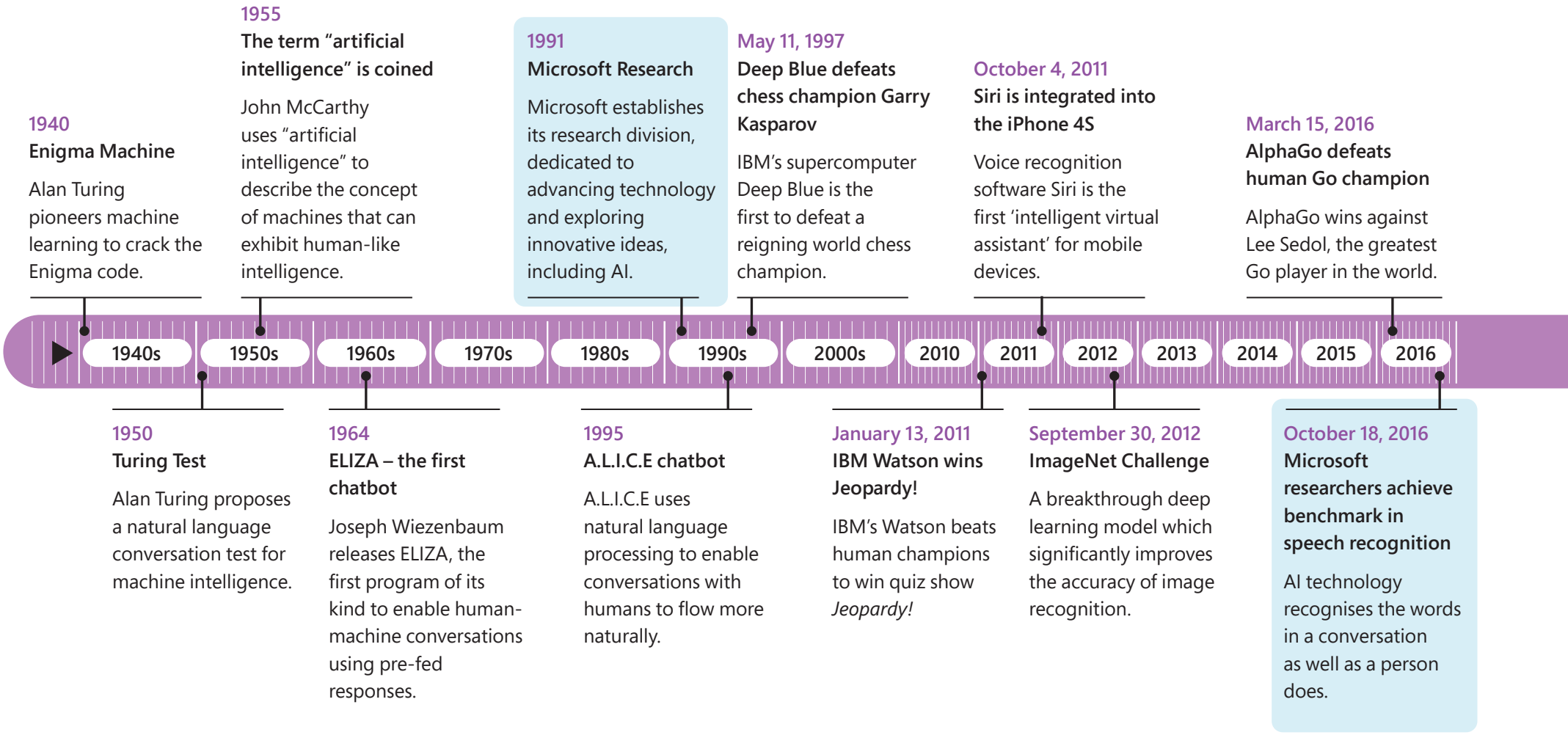
"According to one report, generative AI could enable global labour productivity growth of 0.1 to 0.6 percent annually through to 2040"

The need for generative AI productivity tools is also clear. Microsoft's Work Trend Index found that the number one struggle employees faced in 2022 was not having enough time and energy to do their jobs. This time and energy deficiency is a significant threat to business outcomes, as these employees are twice as likely to leave their jobs and 3.5 times as likely to struggle with being innovative and thinking strategically.

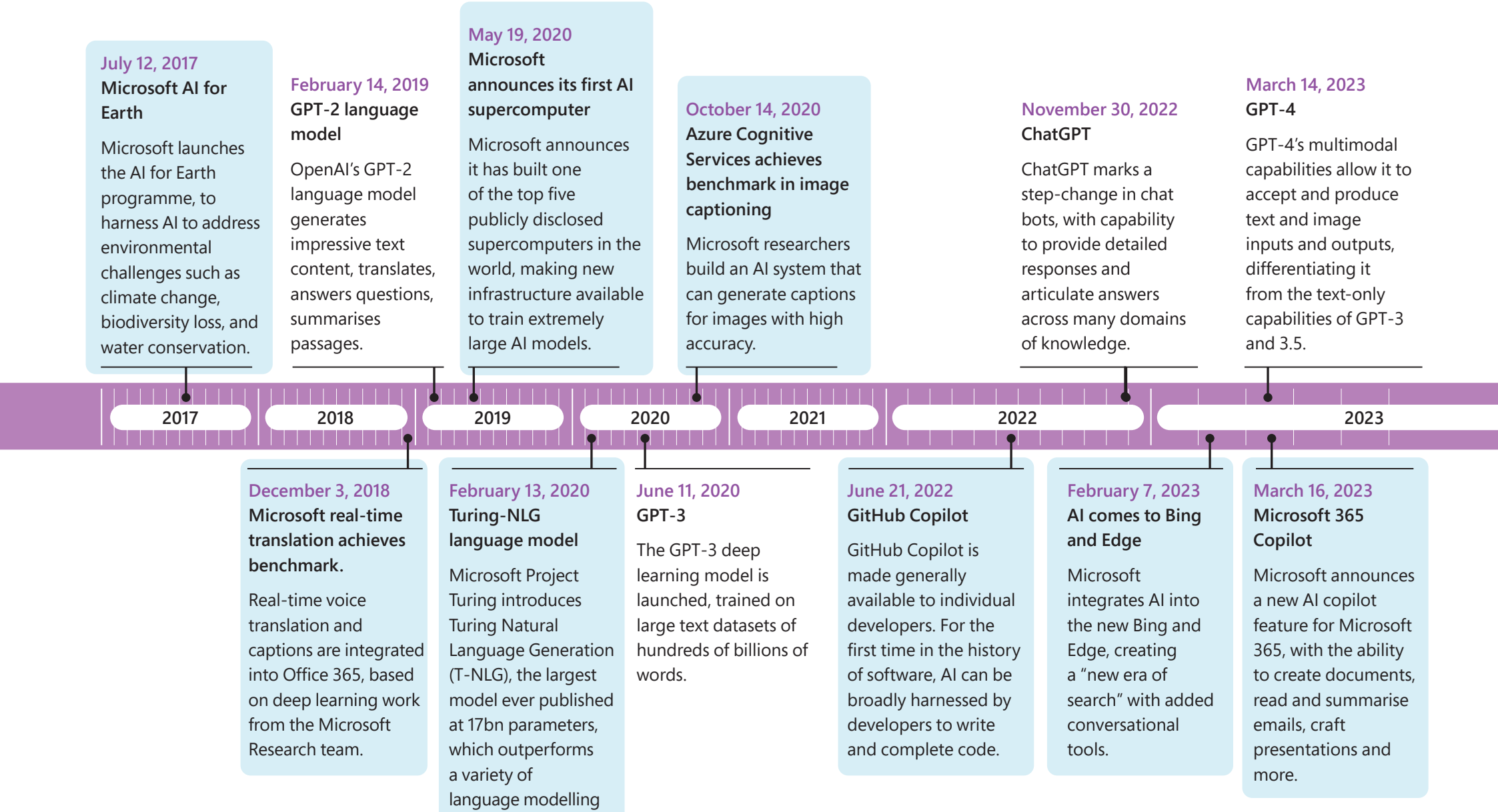
Even though a broad range of productivity tools exists, full use of their capabilities remains untapped. For instance, the average person uses less than 10% of what Microsoft PowerPoint can do. Generative AI, acting as a co-pilot, can unlock the other 90% - while freeing up time to be spent on more productive tasks. Generative AI is the ultimate tool for prototyping, even for work outside our own skillsets. It can make our writing better, even as it makes us better writers. It can help us with the work we find most challenging so we can get back to the work we enjoy.

The rapid advances of AI echo previous technology-based paradigm shifts: technologies like the telephone, the automobile, the internet. Technologies that caused changes in society and the economy, and that required new ways of thinking about work and the skills needed in the workforce.

A brief history of AI: From Turing to ChatGPT



Microsoft projects and collaborations are highlighted in a blue box


July 12, 2017
Microsoft AI for Earth
Microsoft launches the AI for Earth programme, to harness AI to address environmental challenges such as climate change, biodiversity loss, and water conservation.
February 14, 2019
GPT-2 language model
OpenAI's GPT-2 language model generates impressive text content, translates, answers questions, summarises passages.
May 19, 2020
Microsoft announces its first AI supercomputer
Microsoft announces it has built one of the top five publicly disclosed supercomputers in the world, making new infrastructure available to train extremely large AI models.
October 14, 2020
Azure Cognitive Services achieves benchmark in image captioning
Microsoft researchers build an AI system that can generate captions for images with high accuracy.
November 30, 2022
ChatGPT
ChatGPT marks a step-change in chat bots, with capability to provide detailed responses and articulate answers across many domains of knowledge.
March 14, 2023
GPT-4
GPT-4's multimodal capabilities allow it to accept and produce text and image inputs and outputs, differentiating it from the text-only capabilities of GPT-3 and 3.5.
2017
2018
2019
2020
2021
2022
2023
December 3, 2018
Microsoft real-time translation achieves benchmark.
Real-time voice translation and captions are integrated into Office 365, based on deep learning work from the Microsoft Research team.
February 13, 2020
Turing-NLG language model
Microsoft Project Turing introduces Turing Natural Language Generation (T-NLG), the largest model ever published at 17bn parameters, which outperforms a variety of language modelling benchmarks.
June 11, 2020
GPT-3
The GPT-3 deep learning model is launched, trained on large text datasets of hundreds of billions of words.
June 21, 2022
GitHub Copilot
GitHub Copilot is made generally available to individual developers. For the first time in the history of software, AI can be broadly harnessed by developers to write and complete code.
February 7, 2023
AI comes to Bing and Edge
Microsoft integrates AI into the new Bing and Edge, creating a "new era of search" with added conversational tools.
March 16, 2023
Microsoft 365 Copilot
Microsoft announces a new AI copilot feature for Microsoft 365, with the ability to create documents, read and summarise emails, craft presentations and more.

Developing and scaling AI

More than a decade ago, Microsoft forecast an exponential growth in demand for AI systems and started to build special computing infrastructure to handle it. Inspired by early research developments, we have integrated large-scale language models in services ranging from Microsoft Bing to Microsoft 365, and have enabled other companies to take advantage of these technologies through Azure AI and the Azure OpenAI Service. We believe that every person and organisation in the world should benefit from the power of large-scale AI models.

We have developed platforms, tools and a supercomputing infrastructure that can be seen as foundational building blocks that businesses can leverage custom layers on top of to drive economic growth over time. British companies, large and small, and organisations across the public sector use our tools to drive productivity and grow their business. Through Azure OpenAI Service, we provide organisations and developers with access to high-performance AI models, such as GPT-4, Codex and DALL·E 3.

The service is based on the same platform we use to power AI models in our own products, including GitHub Copilot, Power Platform, Microsoft Designer, Bing and Edge.

We continue to enhance our AI infrastructure by incorporating feedback and insights gained from the extensive training and deployment of AI models, and our teams are working in lockstep with industry partners on the design of processors, networks and datacentres optimised for AI.

Microsoft is committed to building Azure into an AI supercomputer for the world, serving as the foundation of our vision to democratise AI as a platform. We pushed the frontier of cloud supercomputing technology, announcing our first top five supercomputer in 2020, and have subsequently constructed multiple AI supercomputing systems at massive scale.

The convergence of innovations in infrastructure, machine learning acceleration software, platform service and modelling powered by cloud technology has created the perfect conditions to accelerate innovation in AI and enable every company to become an AI technology company.

Why the cloud is the critical infrastructure behind AI

Training AI models requires both large datasets and AI-optimised cloud computing infrastructure

Cloud-based AI platforms make it easier for organisations to develop their own AI applications

Cloud computing resources give organisations of all sizes a cost-effective way to run and scale AI innovation

The technology stack for AI foundation models

The technology stack of AI foundation models consists of different components that work together to create powerful AI systems. These AI foundation models are designed to understand and generate language, solve problems and perform various tasks.

Applications

Software programs where the output of an AI model is put to work

API services

APIs (Application Program Interfaces), or endpoints, through which applications access pre-trained models

Powerful pre-trained AI models

Pre-trained models like GPT-4 that can be used to solve similar problems without starting from scratch

Machine learning acceleration software

Software that speeds up the process of developing and deploying large AI models

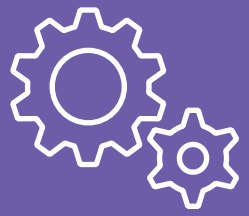
AI datacentre infrastructure

Advanced supercomputing infrastructure, including clusters of advanced GPUs (graphics processing units) with high bandwidth network connections

Putting people at the centre

At every stage from development to deployment and ongoing supervision, people play an essential role in AI: collecting, preparing, and curating data to train AI systems; evaluating which algorithms or models to use; validating system performance in real-world scenarios to assess its behaviour and robustness before deployment; gathering and analysing ongoing user feedback to identify errors, biases, or limitations to make improvements to the data, algorithms, or models; and many more roles.

By prioritising safeguarding throughout, AI developers, deployers, and users can put in place measures to mitigate potential harm, ensure fairness, and promote transparency. This hands-on involvement is critical to shaping and guiding AI systems to achieve reliable and beneficial outcomes.



AI in action

Recent months have put a spotlight on advancements in generative AI, and the possibilities of how it can shape the future of industry. The pace and scale of technological innovation happening today is unprecedented and breathes optimism into a demanding economy. It is also providing a powerful boost to our efforts to address the major societal challenges we face.

And that matters to all of us, as it is ultimately about improving people's lives.

Health

We are at a unique moment in history where medicine, biology, and technology are converging on a large scale. Healthcare organisations and life sciences firms are using AI to develop new treatments and medicines, and helping doctors and nurses improve patient care while tackling backlogs across public health services.

Sustainability

The climate challenge is a race against the clock to decarbonise. AI is helping accelerate and scale our efforts to transition to clean energy and mitigate the impact of global warming on people and communities.

Agriculture

Ensuring we have access to affordable food even as the world's population grows and weather patterns become more extreme requires us to think differently about farming. AI is helping farmers improve yields even as they use fewer resources, especially water.

Manufacturing

The UK is the 8th largest industrial nation in the world . AI is helping to make the way we produce things safer, more efficient, and more profitable.

Education

AI is transforming the way we learn and train, making it more accessible, personalised and impactful, benefitting educators and students alike.

AI for a healthier future

The UK is at a health crossroads. The Office for National Statistics has forecast that by 2050, the number of centenarians in the UK will triple to over 50,000. Health systems will need to shift from episodic care to an approach that is much more proactive and focused on long-term care management. A recent report surveyed health care professionals, who said AI solutions can ease administrative tasks, facilitate home-based care and improve clinical practice.

According to data published by the Kings Fund, prior to the onset of the coronavirus pandemic in 2019, the UK Government spent the equivalent of 9.9% of UK GDP on health. And nearly two thirds of health care expenditure in the UK in 2019 was for curative and rehabilitative care. Prevention is better than a cure, and AI can help accelerate public health and clinical research to identify solutions to some of the most pressing health care challenges. In addition, the sector faces a staffing crisis. Europe faces an ageing health workforce, with many doctors approaching retirement age and a surge in chronic illnesses. Dr Hans Henri P Kluge, WHO Regional Director for Europe stated in March 2023: "Health providers and workers across our region are clamouring for help and support." AI and Cloud technologies can help health care staff tackle this workload.

AI speeds up cancer treatment

Doctors at Addenbrooke's hospital in Cambridge are using AI to drastically cut waiting times for radiotherapy patients.

Proportion of UK Government expenditure spent on health in 2019, prior to onset of the coronavirus pandemic

"In 2019, the UK Government spent the equivalent of 9.9% of UK GDP on health"

Source: Kings Fund

The new AI system, named "OSAIRIS," has revolutionised the preparation of scans, reducing the time patients must wait between referral and treatment initiation. "OSAIRIS does much of the work in the background so that when the oncologist sits down to start planning treatment, most of the heavy lifting is done," Dr. Raj Jena, the researcher who led the project, explains. "It is the first cloud-based AI technology to be developed and deployed within the NHS, which we will be able to share across the NHS for patient benefit."

The OSAIRIS technology is built on Project InnerEye, a decade-long research collaboration project between Cambridge-based Microsoft Research and Addenbrooke's. Its aim is to save clinicians time laboriously marking up patient scans prior to radiotherapy. To broaden access to this technology, Microsoft Research made Project InnerEye toolkits available as open-source software in September 2020. The team has demonstrated how models built using the InnerEye toolkits can cut preparation time by up to 90% – meaning that waiting times for starting potentially life-saving radiotherapy treatment can be dramatically reduced.

Nuance AI tool

Doctors have been taking notes since Ancient Egypt, when they used papyrus to note medical histories. However, even with digital tools, writing them up takes up precious time, making it a perfect job for AI support. Nuance, which develops AI and voice recognition for medical applications, is used by more than 10,000 health care organisations worldwide.

Oxford University Hospitals NHS Foundation Trust (OUH) is a world-renowned centre of clinical excellence and one of the largest National Health Service teaching trusts in the UK. It relied on writing and sending printed letters to patients in many departments. In 2017, one department at the OUH was averaging a 12-day turnaround of letters from frontline doctors. Their target was five days. The issue was a chronic shortage of administration staff and the complex, costly work of transcription, printing, and mailing.

Nuance's Dragon Medical One is a tool that uses AI to input voice-generated content directly into clinical systems to develop structured clinical letter templates. Giving this mundane, but indispensable, job to AI cut turnaround for letters from an average of 12 days to three. Where no blood results are required, letters can be sent in real-time.

Leading innovation in the UK's NHS

The NHS Business Services Authority provides a range of critical central services to NHS organisations, NHS contractors, patients and the public, including: the NHS Prescription Service, NHS pensions, NHS penalty charges, NHS dental services, Total Reward statements, NHS injury benefits scheme, England Infected Blood Support Scheme, NHS Jobs, NHS ESR staff record, NHSBSA statistics, student services and information services.

Among its activities, NHS BSA operates the NHS prescription service, including processing more than 54 million prescriptions per month. Of these, 24 million are electronic, but 30 million are paper forms. The data from many of these paper forms will need to be manually keyed in by an operator. NHS BSA decided to explore the opportunities to streamline this process through greater automation using Microsoft AI and machine learning technologies and a machine vision solution. Following the success of this pilot project, NHS BSA is exploring new ways to leverage the machine vision technologies within its operations.

> "Cloud services are helping to create one of the most detailed pictures we've ever had of people's health"

AI for a better tomorrow

In the UK, Our Future Health is a collaboration between the private, charity, and public sectors to address some of the leading causes of death and serious illness in the UK, including dementia, cancer, diabetes, heart disease, arthritis, and strokes. The program is recruiting five million volunteers to share health information about themselves, which AI-powered tools will turn into health insights. Our Future Health will use Microsoft Cloud to securely store the data in a Trusted Research Environment (TRE) provided by DNAnexus, which will sit in a UK Azure region.

Cloud services "will be an integral part of Our Future Health, underpinning so many important systems that are essential to the running of the program and ultimately helping to create one of the most detailed pictures we've ever had of people's health," says Andrew Roddam, Chief Executive of Our Future Health.

AI for a more sustainable future

Sustainability is the challenge of the century. According to the World Meteorological Organization (WMO), there is a 98% likelihood that at least one of the next five years, and the five-year period as a whole, will be the warmest on record. To curb the progression of climate change, humans urgently need to change the way the planet's resources are used and accelerate decarbonisation.

Microsoft selects AI start-ups to accelerate progress on environmental sustainability

Ten companies focused on reducing carbon emissions and waste, preserving water and protecting ecosystems have been selected to take part in Microsoft's AI for Environmental Sustainability Accelerator programme. The programme, launched in collaboration with the Met Office and Social Tech Trust, will support the cohort through a four-month programme to advance their environmental solution using AI. The ten companies will develop their technological capabilities and commercial strategy, build the relationships they need to scale their solution and, crucially, learn how to measure the impact it will have for environmental progress. Among the accelerator participants is Exeter-based Quanterra Systems, which is creating a measurement solution for the indicators of ecosystem health.

A Planetary Computer for a Sustainable Future

The Microsoft Planetary Computer is a platform that leverages the power of the cloud to accelerate environmental sustainability. It puts global-scale environmental monitoring capabilities in the hands of scientists, developers, and policy makers, enabling data-driven decision making.

> "The Microsoft Planetary Computer puts global-scale monitoring capabilities in the hands of scientists"

From gaining new insights about global land use to using data-driven analysis on past and future carbon offset projects, hosted in the Azure Cloud, our open source AI driven software can be used by developers to support policy makers around the world.

AI for Earth

Since it was launched in the summer of 2017, the program has empowered individuals and organisations to develop innovative solutions to the way we monitor, model, and ultimately manage Earth's natural systems.

Of the over 950 projects supported so far, one project, 'To the Poles' from Aberystwyth University, aims to increase understanding about the Arctic glaciers. Melting glaciers provide some of the clearest visual evidence of the Earth's warming temperatures.

Measuring this change is likely to provide important information about the pace and magnitude of climate change. By applying machine learning to data from drones and satellites, the project will produce high resolution mapping of ice surfaces that over time can be scaled to glaciers.

The ultimate aim being that with deeper knowledge, further investigation and public interest, polar regions that are fundamental to the climate will be better protected.

AI may hold a key to the preservation of the Amazon rainforest

Protecting the Amazon rainforest is a large-scale issue that no one organisation or entity can tackle on its own. Through using machine learning, cloud technology, and data science, researchers are developing new programs that will make recognising deforestation patterns easier and provide tools for policymakers to tackle deforestation. Researchers at the Microsoft AI for Good lab, together with organisations in Colombia, are using AI models to monitor deforestation and protect the biodiversity of the ecosystem, which allows analysis to be developed 90% faster than traditional methods.

> "Researchers are using AI models to protect the biodiversity of the ecosystem"

In Brazil, where the effects of deforestation and mining in the Amazon have been most prominent, a collaboration between Microsoft and Brazilian environmental organisations is also addressing the problem using AI.

Using satellite images, stored in the cloud, AI algorithms detect unofficial roads and other risk factors of deforestation. High risk areas are highlighted on an interactive map, which is then used to support decision-making for safeguarding the rainforest.

AI is helping animal conservation in Scotland

Together with renewable energy provider SSE, in Scotland on the Isle of May, Microsoft supported a project to use AI to trail new ways of counting puffins. Using cameras and AI, the project is able to count the number of puffins and monitor their burrows on the Isle of May all day, every day, without going near them. While they are not at imminent risk of extinction, there are serious concerns over numbers of puffins in the wild.

This makes understanding whether their population is growing or shrinking important in preserving biodiversity and monitoring population levels a vital part of this question. The use of AI, machine learning and image recognition technologies has the potential to significantly increase the ease and accuracy involved.

AI for a growing future in agriculture

Humans have been farming crops for millennia, but this ancient practice can stand to gain a great deal from modern technologies as the planet adjusts to feeding 8 billion people.

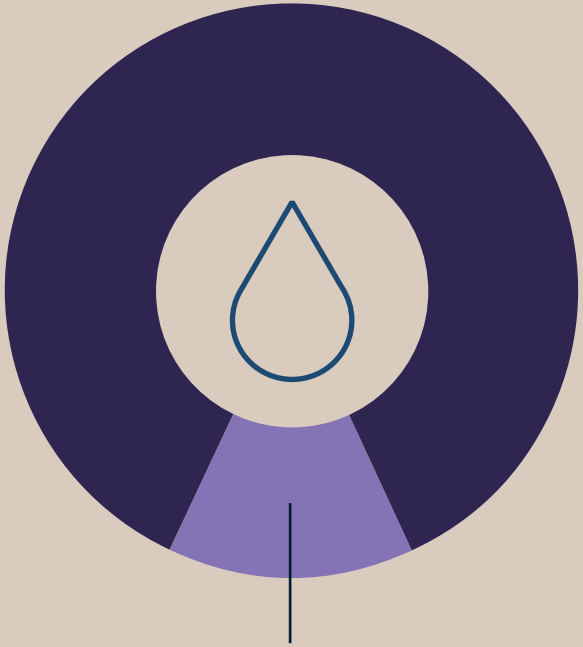
AI, Cloud, and Internet of Things (IoT) technologies have the power to transform farming and agribusinesses. And transformation is needed.

According to the OECD-FAO Agricultural Outlook to 2031, to achieve the UN Sustainable Development Goals Zero Hunger target while reducing direct greenhouse gas emissions by 6%, overall agricultural productivity would need to increase by 28% over the next decade. At the same time, nearly one in three people worldwide lack regular access to adequate food. Optimising harvests while conserving precious resources, including water, is becoming essential.

In the UK alone, the agricultural sector accounts for 14% of water use and is now looking towards technology, powered by data and AI, to reduce its footprint.

"The soil is not getting any richer; the water levels are receding; there is climate change – these make the farmer's life much harder," says Microsoft's Chief Technology Officer for Agri-Food, Ranveer Chandra. "One approach that can help is that of data-driven agriculture, where our goal is not to replace the farmer but to augment the farmer's knowledge with data and AI."

Proportion of water use in the UK that is agricultural



"In the UK alone, the agricultural sector accounts for 14% of water use"

Feeding the world

Farmers know their land through decades, if not generations, of experience. Still, many decisions, such as when to sow seeds or fertilise soil, are based on rough estimates and gut feeling. AI can remove the guesswork.

Project FarmVibes.AI, which runs on Microsoft Azure, makes it possible to predict and plot the ideal amounts of fertiliser and herbicide required based on the level of soil nutrients and to forecast temperatures and wind speeds across their fields. The technology can also help determine the ideal depth to plant seeds, estimate irrigation needs based on soil moisture and guide how different crops and practices can keep carbon sequestered in the soil.

Through these digital tools, farmers can augment their capabilities and knowledge about their farm with data and AI, helping them to make the best choices for optimising harvests with minimal resources.

Food Security Analysis and Forecasting: A Machine Learning Case Study in Southern Malawi

According to the Food and Agriculture Organization, up to 783 million people faced hunger in 2022. Extreme weather events, like floods or droughts, will exacerbate the amount of people who go hungry. Charities and humanitarian agencies' ability to respond quickly and accurately makes it easier to provide people with basic assistance.

By using AI, in combination with up to date information, our research suggests that response times can be shortened through anticipating problems and targeting resources accordingly.

When used to predict food security outcomes, the model in our research project, supported through Microsoft's AI for Good lab, had an 83% accuracy rate and generated accurate forecasts up to four months in the future.

AI can make existing early warning systems more accurate in vulnerable countries, improving our ability to respond.

"When used to predict food security outcomes, the model had an 83% accuracy rate"

AI for smarter manufacturing

AI, Cloud and Big Data are already powering Industry 4.0, the next evolution of industry. Intelligent machines are already in place in smart factories, enabling manufacturing companies to be safer, more efficient, more productive, and more profitable.

AI can support simulations that prepare for real-world scenarios (digital twins), support predictive maintenance and optimise processes to save energy. Smarter factories, in turn, can produce smarter devices that can send data back to manufacturers, who can update the software, improve the device, and design new features for the next iteration. We are proud that companies including Rolls-Royce, Siemens, Mercedes-Benz and Bayer have trusted Microsoft to support them in their own innovation journeys – using our AI, sustainability, and cloud offerings to digitally transform their organisations, while improving service and experience for their customers.

State of the nation

Manufacturing contributes around 11% to the global economy, and the UK remains the 8th largest industrial nation in the world, with a robust and growing manufacturing industry. Our manufacturing is a global leader when adopting emerging digital transformation strategies such as AI, with one report revealing more than half of UK manufacturers (51%) are currently using AI to some degree. However, the sector faces challenges regarding the pace and scale of AI adoption. A key challenge for the sector is the perception among workers about the risks automation poses to jobs, despite acknowledgement that these emerging technologies could give a much needed boost to productivity and safety. These concerns must be addressed and allayed if the industry is to truly move forward on its AI-led digital transformation.

> “More than half of UK manufacturers are currently using AI to some degree”

AI for better education

AI has immense potential to transform education and training for students, teachers and school staff. It can help students with learning difficulties and support teachers in creating individualised learning. It could reduce school dropout rates and enable differentiated education programmes. It can help analyse vast swathes of data to see which approaches work best.

There is an urgent need to transform education: as the world is changing, so are careers. The next generation will be working closely with AI, yet there is a desperate shortage of digital skills.

According to the Office for National Statistics, 8.8% of businesses have had a high demand for advanced digital skills in the last 12 months. And in 2020, a Microsoft report revealed that 82% of UK jobs already require digital skills, and that 69% of leaders feel their organisation suffers from a digital skills gap. With the rapid advances in AI technology, the digital skills gap is likely to widen unless concerted efforts are made to accelerate upskilling. Our Skills White Paper published in May 2023 is informed by granular LinkedIn data on regional skills gaps and in-demand skills across the UK and sets out a number of recommendations aimed at transforming skills delivery in the UK.

UNESCO wants countries to harness the potential of AI for learning "while ensuring that its application in educational contexts is guided by the core principles of inclusion and equity". It notes in the Beijing Consensus, which sets out a common position on AI and learning, that the promise of "AI for all" must be that everyone can take advantage of the technological revolution underway and access its benefits, notably in terms of innovation and knowledge.

And for this to happen, there is a growing need for researchers, educators, and students to have a foundational understanding of AI and data use to be able to engage positively, critically, and responsibly with this technology.

"UNESCO wants countries to harness the potential of AI technologies for learning"

New AI skills

AI offers tremendous potential to empower workers around the world – but only if everyone, everywhere has the skills to use it. It's little surprise that, according to the World Economic Forum, AI skills represent the third-highest priority for companies' training strategies, right alongside analytical and creative thinking. To begin addressing this need, Microsoft launched a new AI Skills Initiative to help people and communities around the world learn how to harness the power of AI.

The Microsoft AI Skills Initiative includes new, free coursework developed with LinkedIn, including the first Professional Certificate on Generative AI in the online learning market; a new open global grant challenge in coordination with data.org to uncover new ways of training workers on generative AI; and greater access to free digital learning events and resources for everyone to improve their AI fluency.

AI-powered Microsoft Bing

In February 2023, Microsoft integrated AI into Bing and Edge, delivering better search, more complete answers, a new chat experience and the ability to generate content. We think of these tools as an AI copilot for the web.

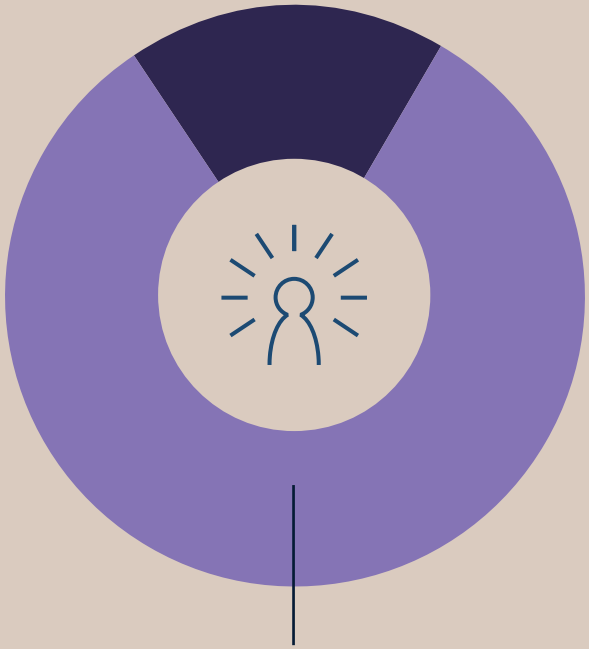
Bringing together search, browsing and chat into one experience helps address the shortfalls of traditional search.

There are 10 billion search queries a day, but we estimate half of them go unanswered. That's because people are using search to do things it was not originally designed to do.

The groundbreaking new search experiences are possible because Microsoft has committed to building Azure into an AI supercomputer, and OpenAI has used this infrastructure to train through the breakthrough models that are now being optimised.

We have been intentional in implementing safeguards to defend against harmful content. Our teams are working to address issues such as misinformation and disinformation, and preventing the promotion of harmful or discriminatory content in line with our AI principles.

Proportion of UK jobs requiring digital skills



"In 2020, a Microsoft report revealed that 82% of UK jobs already require digital skills"

Source: Microsoft

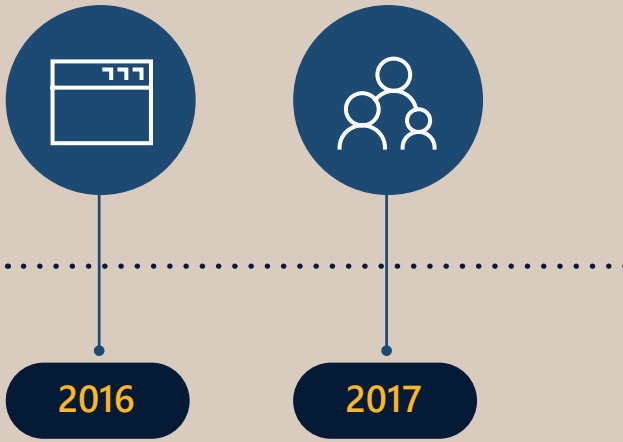
The value of AI

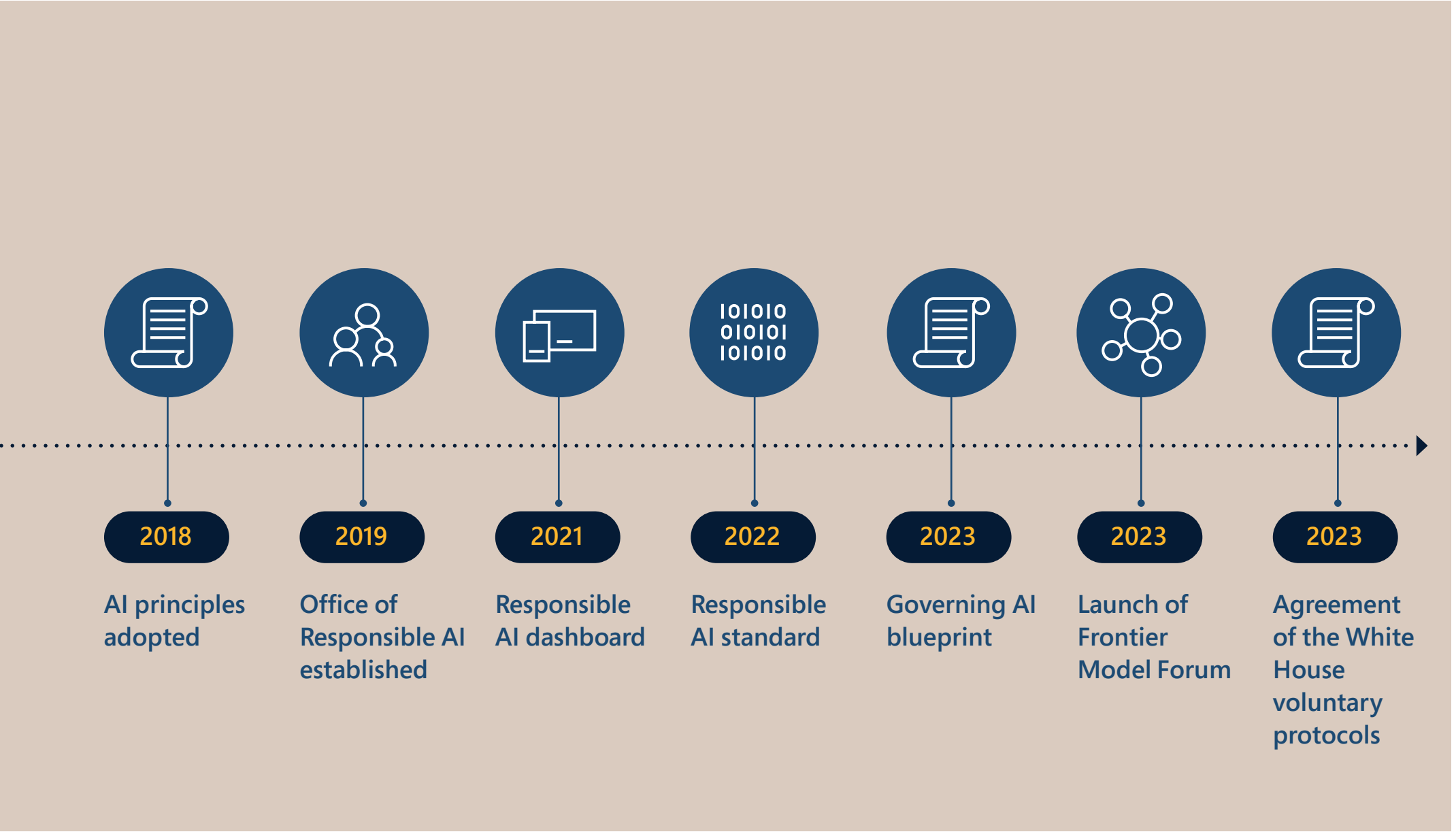
The scale of value that AI could help British organisations and entrepreneurs create for the UK economy and citizens is truly exciting. According to PwC, the UK's GDP will be 10.3% higher in 2030 as a result of AI. This is the equivalent of an additional £232 billion.

Key to this will be fostering a thriving 'AI ecosystem' and ensuring AI is an affordable and accessible platform for British organisations of all kinds and sizes. The true value of AI will only be realised through diffusing its use throughout the economy; through the development of AI applications that can be harnessed by businesses and organisations across the country; and by ensuring the training and skills are available to enable everyone to leverage the benefits of AI.

Responsible development and use of AI

Our responsible AI journey




2018
AI principles adopted
2019
Office of Responsible AI established
2021
Responsible AI dashboard
2022
Responsible AI standard
2023
Governing AI blueprint
2023
Launch of Frontier Model Forum
2023
Agreement of the White House voluntary protocols

Responsible AI – Q&A with Natasha Crampton

Natasha Crampton leads Microsoft's Office of Responsible AI, as the company's first Chief Responsible AI Officer. The Office of Responsible AI puts Microsoft's AI principles into practice by defining, enabling, and governing the company's approach to responsible AI. The Office of Responsible AI also collaborates with stakeholders within and outside the company to shape new laws, norms, and standards to help ensure that the promise of AI technology is realised for the benefit of all.

Why is responsible AI so important?

AI offers tremendous opportunities to improve our world – stimulating economic growth, making work more enjoyable and helping address pressing societal issues like sustainability. However, like every technology before it, some people, communities and countries will turn this advancement into both a tool and a weapon. Some people will use this technology to exploit the flaws in human nature, deliberately target people with false information, undermine democracy and explore new ways of pursuing criminal endeavours.

This creates a profound sense of responsibility, both in terms of AI development and its deployment. Since 2017, Microsoft has invested in a cross-company programme to ensure that our AI systems are responsible by design.

How does Microsoft approach responsible AI?

For Microsoft, responsible AI development is about both a practice and a culture.

Practice is how we formally operationalise responsible AI across the company, through governance processes, policy requirements, and tools and training to support implementation. Culture is how we empower our employees to actively champion responsible AI.

We see three essential tenets of upholding the right culture and practices:

1. Leadership from the top
Microsoft's responsible AI journey began in 2016 when Microsoft's Chairman and CEO, Satya Nadella, penned an article articulating his vision for humanity powered by AI. Satya's vision set in motion the beginnings of our core AI principles. Microsoft's responsible AI efforts continue to be championed by Microsoft's senior leadership team today. Microsoft's Responsible AI Council is chaired by Microsoft's Vice Chair and President, Brad Smith and by Chief Technology Officer Kevin Scott and convenes senior staff from research, policy, and engineering teams who lead responsible AI efforts across the company. The council serves as a forum for leadership alignment and oversees the progress we are making on our responsible AI commitments.

2. Inclusive governance models
Since starting our work on responsible AI seven years ago, we learned the need to create a governance model that was inclusive and encouraged multi-disciplinary collaboration across engineering, policy, and research. The governance model established by Microsoft's Office of Responsible AI ensures that there are senior leaders tasked with spearheading responsible AI across every business group. These leaders are supported by a large network of trained "responsible AI champions". Our continually updated Responsible AI Standard, which is publicly available, offers this community guidance for building AI systems responsibly. Taking lessons from long-standing, cross-company commitments to privacy,

security, and accessibility, we realised that responsible AI must be supported by the highest levels of leadership in the company and championed at every level across Microsoft. Responsible AI is a company-wide mandate.

3. Investment in people
Microsoft currently has nearly 350 people working on responsible AI, with 129 dedicated to it full-time. These team members hold positions across the company in policy, engineering, research, sales, and other core functions. We are committed to continue investing in hiring diverse talent, assigning additional talent to focus on responsible AI full time, and upskilling more people throughout the company.

How does the Responsible AI Standard help teams put Microsoft's principles into practice?

From crafting an AI system's purpose to designing how people interact with it, we must keep people at the centre of all AI decisions. While our responsible AI principles state the enduring values we seek to uphold, we needed more specific guidance on how to build and deploy AI systems responsibly. This is why we developed our Responsible AI Standard, a more practical guide that memorialises a set of "rules of the road" for our engineering teams so that upholding our AI principles is a daily practice.

Microsoft's responsible AI principles

Fairness:
AI systems should treat all people fairly

Reliability & safety:
AI systems should perform reliably and safely

Privacy & security:
AI systems should be secure and respect privacy

Inclusiveness:
AI systems should empower everyone and engage people

Transparency:
AI systems should be understandable

Accountability:
People should be accountable for AI systems

The Responsible AI Standard was the result of a multi year, cross-company effort that reflected a vast array of input from researchers, engineers, lawyers, designers, and policy experts. We consider it to be a significant step forward for our practice of responsible AI because it sets out concrete, practical guidance on how to identify, measure, and mitigate harms ahead of time. It also requires teams to adopt tools and controls to secure beneficial uses while guarding against potential misuses of their products.

When building and updating the Responsible AI Standard, we recognised early on that it is impossible to reduce all the complex sociotechnical considerations – for many different use-cases – into an exhaustive set of pre-defined rules. This led us to create a programme and process for ongoing review and oversight of high-impact cases and rising issues and questions, which we call Sensitive Uses. The Sensitive Uses review process is triggered when Microsoft personnel are involved in developing or deploying an AI system and the foreseeable use or misuse of that AI system could have a consequential impact on a user's legal status or life opportunities, present the risk of significant physical or psychological injury, or restrict, infringe upon, or undermine the ability to realise an individual's human rights.

Our Sensitive Uses programme provides a critically important additional layer of oversight for teams working on higher-risk use-cases of our AI systems.

The Rome Call for AI Ethics

The Rome Call for AI Ethics is an important global effort to promote an ethical approach to artificial intelligence. Microsoft has been involved in this initiative since 2020, which stands for three key principles:

Importance of ethical safeguards to ensure AI technology is transparent, inclusive, and governed with human interests in mind.

Shared responsibility among tech companies and developers to create and implement AI with a broad perspective on ensuring a positive societal impact.

The importance of AI advancing the global community's sustainability efforts.

Microsoft AI Red Team building future of safer AI

An essential part of shipping software securely is red teaming. It broadly refers to the practice of emulating real-world adversaries and their tools, tactics, and procedures to identify risks, uncover blind spots, validate assumptions, and improve the overall security posture of systems. Microsoft has a rich history of red teaming emerging technology with a goal of proactively identifying failures in the technology. As AI systems became more prevalent, in 2018, Microsoft established the AI Red Team: a group of interdisciplinary experts dedicated to thinking like attackers and probing AI systems for failures.

We're sharing best practices from our team so others can benefit from Microsoft's learnings. These best practices can help security teams proactively hunt for failures in AI systems, define a defense-in-depth approach, and create a plan to evolve and grow your security posture as generative AI systems evolve.

The practice of AI red teaming has evolved to take on a more expanded meaning:

it not only covers probing for security vulnerabilities, but also includes probing for other system failures, such as the generation of potentially harmful content. AI systems come with new risks, and red teaming is core to understanding those novel risks, such as prompt injection and producing ungrounded content. AI red teaming is not just a nice to have at Microsoft; it is a cornerstone to responsible AI by design: as Microsoft President and Vice Chair, Brad Smith, announced, Microsoft has committed that all high-risk AI systems will go through independent red teaming before deployment.

The Frontier Model Forum

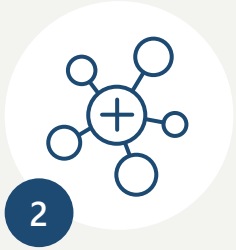
In July 2023, Microsoft, Google, Anthropic and OpenAI announced the formation of the Frontier Model Forum, a new industry body focused on ensuring safe and responsible development of frontier AI models.

Governments and industry agree that, while AI offers tremendous promise to benefit the world, appropriate guardrails are required to mitigate risks. Important contributions to these efforts have already been made by the US and UK Governments, the European Union, the OECD, the G7 (via the Hiroshima AI process), and others.

To build on these efforts, further work is needed on safety standards and evaluations to ensure frontier AI models are developed and deployed responsibly. The Forum will be one vehicle for cross-organisational discussions and actions on AI safety and responsibility.

The core objectives for the Frontier Model Forum are:

1. **Advancing AI safety research to promote responsible development of frontier models, minimise risks, and enable independent, standardised evaluations of capabilities and safety.**



2. **Identifying best practices for the responsible development and deployment of frontier models, helping the public understand the nature, capabilities, limitations, and impact of the technology.**

3. **Collaborating with policymakers, academics, civil society, and companies to share knowledge about trust and safety risks.**

4. **Supporting efforts to develop applications that can help meet society's greatest challenges, such as climate change mitigation and adaptation, early cancer detection and prevention, and combating cyber threats.**

Over the coming months, the Frontier Model Forum will establish an Advisory Board to help guide its strategy and priorities, representing a diversity of backgrounds and perspectives. The founding Frontier Model Forum companies will also establish key institutional arrangements including a charter, governance, and funding with a working group and executive board to lead these efforts.

Helping our UK customers develop and deploy AI responsibly – Antony Cook

Antony Cook leads Microsoft Customer and Partner Solutions group, a team comprising more than 350 legal and government affairs professionals located in more than 50 countries around the world. They focus on legal support for the global sales organisation including pioneering new legal and corporate affairs solutions that advance business goals to benefit consumers, industry partners, and communities.

Microsoft has a responsible AI process that our AI systems go through, which includes reviews by a multidisciplinary team of experts, that helps us understand potential harms and find mitigations. Examples of mitigations include refining the dataset used to train models, deploying filters to limit the generation of harmful content, integrating techniques like query blocking on sensitive topics that helps to prevent misuse by bad actors, or applying technology that can return more helpful, representative, and diverse responses and results.

Another very important safeguard is intentional and iterative deployment. We take a measured approach to rolling out many of these new AI-powered experiences. We may start with a limited preview with a select number of customers with well-defined use-cases in mind. Collaborations with these early customers will help us make sure the responsible AI safeguards are working in practice so we can scale adoption.

One of our most important responsible AI commitments is sharing our learnings with customers. We provide transparency documentation for our platform AI services. Transparency Notes communicate in clear, everyday language the purposes, capabilities, and limitations of AI systems so our customers can understand when and how to deploy our platform technologies. They also identify use-cases that fall outside the solution's capabilities and the Responsible AI Standard. Transparency Notes fill the gap between marketing and technical documentation, proactively communicating information that our customers need to know to deploy AI responsibly.

Customers also need practical tools to operationalise responsible AI practices. Over the years, responsible AI research at Microsoft has led to the incubation of tools such as Fairlearn and InterpretML. The collection of tools has grown in capability, spanning many facets of responsible AI practice including the ability to identify, diagnose, and mitigate potential errors and limitations of AI systems. Since their original conception within Microsoft, these tools continue to improve and evolve through the contributions of active open-source communities. Our latest tool, which is in preview, is Azure Content Safety. It helps businesses create safer online environments and communities through models that are designed to detect hate, violent, sexual, and self-harm content across languages in both images and text.

Building on the Responsible AI Toolbox, Microsoft's responsible AI programme has invested in integrating some of the more mature responsible AI tools directly into Azure Machine Learning so our customers will also benefit from the development of engineering systems and tools. The collection of capabilities, known as the Responsible AI Dashboard, offers a single pane of glass for machine learning practitioners and business stakeholders to debug models and make informed,

responsible decisions as they build AI systems or customise existing ones.

The community involved in developing, evaluating, and using AI extends beyond our direct customers. To serve this broad ecosystem, we publicly share key artefacts from our responsible AI programme, including our Responsible AI Standard, Impact Assessment Template and Guide, detailed primers on the implementation of our responsible AI by design approach, and collections of cutting-edge research. Our digital learning paths further empower leaders to craft an effective AI strategy, foster an AI-ready culture, innovate responsibly and more.

The bottom line is that every organisation that creates or uses AI systems will need to develop and implement its own governance systems. To help them do so, we will continue to share our knowledge and tools. We will share the work we are doing to build a practice and culture of responsible AI at Microsoft, including key parts of the curriculum that we use to train Microsoft employees. And we continue to invest in dedicated resources and expertise in regions around the world to respond to customer questions about deploying and using AI responsibly.

Furthermore, we have announced that we are creating an AI Assurance Program to help customers ensure that the AI applications they deploy on our platforms meet the legal and regulatory requirements for responsible AI. This programme is open to our customers around the world.

This programme will include:

Regulator engagement support: We have extensive experience helping customers in the public sector and highly regulated industries manage the spectrum of regulatory issues that arise when dealing with the use of information technology. For example, in the global financial services industry, we worked closely for a number of years with both customers and regulators to ensure that this industry could pursue digital transformation on the cloud while complying with its regulatory obligations.

One learning from this experience has been the industry's requirement that financial institutions verify customer identities, establish risk profiles, and monitor transactions to help detect suspicious activity, the "know your customer" requirements. We believe that this approach can apply to AI in what we are calling "KY3C", an approach that creates certain obligations to know one's cloud, one's customers, and one's content. We want to work with our customers to apply KY3C as part of our AI Assurance Program.

Customer councils: We will bring customers together in customer councils to hear their views on how we can deliver the most relevant and compliant AI technology and tools.

KY3C

Applying to AI services the "know your customer" concept developed for financial services

1. Know your cloud

2. Know your customer

3. Know your content

Regulatory advocacy: Finally, we'll play an active role in engaging with governments to promote effective and interoperable AI regulation. The recently launched Microsoft's blueprint for AI governance, (see page 49) presents our proposals to governments and other stakeholders for appropriate regulatory frameworks for AI.

We are committed to supporting customers in the UK and around the world implement their own AI systems responsibly, and we will develop responsible AI programmes for our partner ecosystem.

Many of our partners have already created comprehensive practices to help customers evaluate, test, adopt, and commercialise AI solutions, including creating their own responsible AI systems. We are launching a programme with selected partners to leverage this expertise to assist our mutual customers in deploying their own responsible AI systems. PwC and EY are our initial launch partners and we will be looking to add European partners to the programme.

Ultimately, we know that these commitments are only the start, and we will have to build on them as both the technology and regulatory conditions evolve. But we are also excited by this opportunity to partner more closely with our customers as we continue on the responsible AI journey together.

Advancing trustworthy AI

For more than 30 years, Microsoft Research has been advancing the foundations of computing and translating new scientific understanding into innovative technologies to create value for our customers and broad benefit to society.

Our researchers collaborate across disciplines, institutions, and geographies to deliver cutting-edge advances in vision, speech, language, decision-making and machine learning. They have pioneered AI breakthroughs in conversational speech recognition, machine translation, image captioning, natural language understanding and commonsense question answering.

As part of our ongoing commitment to building AI responsibly, research scientists and engineers at Microsoft are also pursuing methods and technologies aimed at helping builders of AI systems cultivate appropriate trust – that is, building trustworthy models with reliable behaviours and clear communication that set proper expectations.

When AI builders plan for failures, work to understand the nature of the failures and implement ways to effectively mitigate potential harms, they help engender trust that can lead to a greater realisation of AI's benefits.

Facilitating trustworthy measurement, improving human-AI collaboration, designing for natural language processing (NLP), advancing transparency and interpretability, and exploring the open questions around AI safety, security, and privacy are key considerations for developing AI responsibly.

The goal of trustworthy AI requires a shift in perspective at every stage of the AI development and deployment life cycle. We're actively developing a growing number of best practices and tools to help with the shift to make responsible AI more available to a broader base of users. Many open questions remain, but as innovators, we are committed to tackling these challenges with curiosity, enthusiasm, and humility.

Microsoft's new Customer Copyright Commitment

Microsoft's AI-powered Copilots are changing the way we work, making customers more efficient while unlocking new levels of creativity. While these transformative tools open doors to new possibilities, they are also raising new questions. Some customers are concerned about the risk of IP infringement claims if they use the output produced by generative AI. This is understandable, given recent public inquiries by authors and artists regarding how their own work is being used in conjunction with AI models and services.

To address this customer concern, Microsoft has announced our new Customer Copyright Commitment. As customers ask whether they can use Microsoft's Copilot services and the output they generate without worrying about copyright claims, we are providing a straightforward answer: yes, you can, and if you are challenged on copyright grounds, we will assume responsibility for the potential legal risks involved.

This new commitment extends our existing intellectual property indemnity support to commercial Copilot services and builds on our previous AI Customer Commitments. Specifically, if a third party sues a commercial customer for copyright infringement for using Microsoft's Copilots or the output they generate, we will defend the customer and pay the amount of any adverse judgments or settlements that result from the lawsuit, as long as the customer used the guardrails and content filters we have built into our products. Let's start with why we are offering this program:

1 **We believe in standing behind our customers when they use our products**.

We are charging our commercial customers for our Copilots, and if their use creates legal issues, we should make this our problem rather than our customers' problem. This philosophy is not new: For roughly two decades we've defended our customers against patent claims relating to our products, and we've steadily expanded this coverage over time. Expanding our defence obligations to cover copyright claims directed at our Copilots is another step along these lines.

> "It is critical for authors to retain control of their rights under copyright law"

2 **We are sensitive to the concerns of authors, and we believe that Microsoft rather than our customers should assume the responsibility to address them.**

Even where existing copyright law is clear, generative AI is raising new public policy issues and shining a light on multiple public goals. We believe the world needs AI to advance the spread of knowledge and help solve major societal challenges.

Yet it is critical for authors to retain control of their rights under copyright law and earn a healthy return on their creations. And we should ensure that the content needed to train and ground AI models is not locked up in the hands of one or a few companies in ways that would stifle competition and innovation. We are committed to the hard and sustained efforts that will be needed to take creative and constructive steps to advance all these goals.

3 **We have built important guardrails into our Copilots to help respect authors' copyrights.**

We have incorporated filters and other technologies that are designed to reduce the likelihood that

Copilots return infringing content. These build on and complement our work to protect digital safety, security and privacy, based on a broad range of guardrails such as classifiers, metaprompts, content filtering and operational monitoring and abuse detection, including that which potentially infringes third-party content. Our new Customer Copyright Commitment requires that customers use these technologies, creating incentives for everyone to better respect copyright concerns.

Setting the standard for responsible AI innovation – expanding our Customer Copyright Commitment to Azure OpenAI

Microsoft has set the standard with services and tools like Azure AI Content Safety, the Responsible AI Dashboard, model monitoring, and our industry-leading commitment to defend and indemnify commercial customers from lawsuits for copyright infringement.

We have recently expanded our Customer Copyright Commitment (CCC), to customers using Azure OpenAI Service. As more customers build with generative AI inside their organisations, they are inspired by the potential of this technology and are eager to commercialise it externally.

By extending the CCC to Azure OpenAI Service, Microsoft is broadening our commitment to defend our commercial customers and pay for any adverse judgments if they are sued for copyright infringement for using the outputs generated by Azure OpenAI Service.

Our shared AI journey

This announcement is a first step. Like all new technologies, AI raises legal questions that our industry will need to work through with a wide array of stakeholders. This step represents a pledge to our customers that the copyright liability of our products is ours to shoulder, not theirs. Microsoft is bullish on the benefits of AI, but, as with any powerful technology, we're clear-eyed about the challenges and risks associated with it, including protecting creative works. It is our responsibility to help manage these risks by listening to and working with others in the tech sector, authors and artists and their representatives, government officials, the academic community, and civil society.

We look forward to building on such announcements with new initiatives that help ensure that AI advances the spread of knowledge while protecting the rights and needs of creators.

A Copilot for every Microsoft Cloud experience

Modern work
Bing Chat Enterprise
Microsoft 365 Copilot
Windows Copilot

Business applications
Dynamics 365 Copilot
Copilot in Power Platforms
Microsoft Sales Copilot

Digital and app innovation
GitHub Copilot

Data & artificial intelligence
Copilot in Power BI

Security
Microsoft Security Copilot

Our commitments to advance safe, secure and trustworthy AI

In 2023, Microsoft published voluntary commitments to advance safe, secure and trustworthy AI. Microsoft not only endorsed all of the voluntary commitments defined by the White House but also independently committed to several others that support these critical goals.

Microsoft summarised the progress made on our AI safety policies ahead of the UK AI Safety Summit in November 2023 and we continue to support the efforts of the UK's AI Safety Institute. From supporting a pilot of the National AI Research Resource and continuing to advocate for the establishment of a national registry of high-risk AI systems, we believe that these measures will help advance transparency and accountability. We have also committed to broad-scale implementation of the NIST AI Risk Management Framework and adoption of cybersecurity practices that are attuned to unique AI risks. We know that this will lead to more trustworthy AI systems that benefit not only our customers, but the whole of society.

Alignment of our efforts with the White House Voluntary AI Commitments

Safe

White House Voluntary Commitments:
Companies choose to conduct red-teaming, share trust and safety information and help people identify AI-generated content

Microsoft Commitments:

- Test our systems using red-teaming and systematic measurements
- Contribute to industry efforts to develop evaluation standards for emerging safety and security issues
- Implement provenance tools to help people identify AI-generated audio or visual content
- Implement the NIST AI Risk Management Framework
- Implement robust reliability and safety practices for high-risk models & applications

Secure

White House Voluntary Commitments:
Companies choose to make investments to protect unreleased model weights, and incentivise the responsible disclosure of AI system vulnerabilities

Microsoft Commitments:

- Ensure that the cybersecurity risks of our AI products and services are identified and mitigated
- Participate in an approved multistakeholder exchange of threat information
- Support the development of a licensing regime for highly capable models
- Support the development of an expanded 'know-your-customer' concept for AI services

Trustworthy

White House Voluntary Commitments:
Companies choose to be transparent about system capabilities and limitations, prioritise research on societal risks and develop and deploy AI systems for the public good

Microsoft Commitments:

- Release an annual transparency report on the governance of our responsible AI program
- Design our AI systems so that people know when they are interacting with an AI system and be transparent about system capabilities and limitations
- Increase investment in our academic research programs
- Collaborate with the National Science Foundation to explore a pilot project to stand up the National AI Research Resource
- Support the development of a national registry of high-risk AI systems

Our investment in AI infrastructure, skills and security to boost the UK's AI potential

Microsoft announced a major AI infrastructure and skilling investment supported by a new partnership on security. It will help the UK seize the AI opportunity and ensure that AI innovation and safety progress together while creating jobs, improving services, and protecting public security. Microsoft is committing to more than doubling its datacentre footprint in the UK, training more than one million people for the AI economy and supporting the UK's growing AI safety and research efforts through partnerships with the government and leading universities.

The investment will cover three key areas:

1

Capacity: First, Microsoft will invest in the UK's AI infrastructure

Over the next three years Microsoft will spend £2.5 billion ($3.2bn) to expand its next generation AI datacentre infrastructure, bringing more than 20,000 of the most advanced GPUs to the UK by 2026. The single largest investment in its 40-year history in the country, Microsoft will grow its datacentre footprint across sites in London and Wales and potential expansion into northern England.

This infrastructure investment will help to meet the exploding demand for efficient, scalable and sustainable AI-specific compute power and the needs of the private and public sector waiting to take advantage of the latest cloud and AI breakthroughs.

To support research on AI, Microsoft will extend its Accelerating Foundation Models Research (AFMR) program to include prioritised access to GPUs for the UK's science and research community. AFMR drives interdisciplinary research on AI alignment and safety, beneficial applications of AI, and AI-driven scientific discovery in the natural and life sciences. This new UK effort will aim at harnessing the power of AI to accelerate scientific discovery via multiscale multimodal data generation through prioritised access to Microsoft's AI tools. This program includes researchers from the UK's world leading participating universities including Cambridge, Oxford, Imperial College, UCL, Bath, and Nottingham.

2

Capability: Second, Microsoft will invest in broad-based AI talent and education programs.

To support UK workers across the AI economy, Microsoft will make a multi-million-pound investment to train one million people with the skills they need to build and work with AI. This will include expanded training for people looking to start, or move into, a

career in AI. Working in partnership with multiple learning and non-profit partners, the program will focus on building AI fluency, developing AI technical skills, supporting AI business transformation, and promoting safe and responsible AI development and use including the first Professional Certificate on Generative AI.

As part of this skilling commitment, Microsoft will also turn all the lessons it has learned in operationalising responsible AI principles for its own AI engineers and developers, into learning modules for UK customers and partners. This training will help the UK's AI developer ecosystem to embed safety and security measures into their own systems and processes and takes Microsoft's training support for responsible AI beyond principles and embeds sound practice into the way AI is developed and integrated across the IT industry.

> "Microsoft will make a multi-million-pound investment to train one million people with the skills they need to build and work with AI"

Finally, to help ensure Microsoft technical trainers are adhering to the ethics and principles of developing AI solutions responsibly they will all complete and attest to Microsoft's "Responsible Generative AI" training.

3

Security: Third, Microsoft will invest in strong AI safety and security measures

These will cover both Microsoft's own infrastructure and support for AI developers and customers deploying and using AI applications. Microsoft will operate its AI services and infrastructure in accordance with industry-leading responsible AI practices. It will integrate the adoption and use of responsible AI principles into its Partner Pledge for its 25,000 UK partners and it will collaborate with the UK Government and AI Safety Institute on the ongoing development of refinements and improvements in this field.

AI governance

As Microsoft's Vice Chair and President, Brad Smith is responsible for spearheading the company's work and representing it publicly on a wide variety of critical issues involving the intersection of technology and society, including artificial intelligence, cybersecurity, privacy, environmental sustainability, human rights, digital safety, immigration, philanthropy, and products and business for non-profit customers. He leads a team of 2,000 business, legal, and corporate affairs professionals located in 54 countries and operating in more than 120 nations.

Advancing AI governance in the UK and internationally – Brad Smith

As this paper highlights, there are enormous opportunities to harness the power of AI to drive UK growth and improve our public services. The examples highlighted earlier in this Blueprint offer an insight into the potential. But another dimension is equally clear. It's not enough to focus only on the many opportunities to use AI to improve people's lives. We need to focus with equal determination on the challenges and risks that AI can create, and we need to manage them effectively.

This is perhaps one of the most important lessons from the role of social media. Little more than a decade ago, technologists and political commentators alike gushed about the role of social media in spreading democracy during the Arab Spring. Yet, five years later, we learned that social media, like so many other technologies before it, would become both a weapon and a tool – in this case aimed at democracy itself.

> "We are committed and determined as a company to develop and deploy AI in a safe and responsible way"

Today, we are 10 years older and wiser, and we need to put that wisdom to work. We need to think early on and in a clear-eyed way about the problems that could lie ahead. As technology moves forward, it's just as important to ensure proper control over AI as it is to pursue its benefits. We are committed and determined as a company to develop and deploy AI in a safe and responsible way. We also recognise, however, that the guardrails needed for AI require a broadly shared sense of responsibility and should not be left to technology companies alone. In short, tech companies will need to step up, and governments will need to move faster.

When we at Microsoft adopted our six ethical principles for AI in 2018, we noted that one principle was the bedrock for everything else – accountability. This is the fundamental need to ensure that machines remain subject to effective oversight by people, and the people who design and operate machines remain

accountable to everyone else. In short, we must always ensure that AI remains under human control. This must be a first-order priority for technology companies and governments alike.

This connects directly with another essential concept. In a democratic society, one of our foundational principles is that no person is above the law. No government is above the law. No company is above the law, and no product or technology should be above the law. This leads to a critical conclusion: People who design and operate AI systems cannot be accountable unless their decisions and actions are subject to the rule of law. In May 2023, Microsoft released a whitepaper, Governing AI: A Blueprint for the Future, which sought to address the question of how do we best govern AI.

At Microsoft, we build on lessons learned from many years of work, investment, and input. Here we draw some of that thinking together with a specific focus on the UK, its leadership on AI governance and the potential of AI in improving public services, and what a viable path to advancing AI governance internationally may look like. The UK, with its leading universities and research institutions, is making great strides in the safe and responsible advance of AI. With the first ever global summit on AI Safety and ongoing discussions, the UK is at the forefront of international efforts to govern the development and deployment of AI technologies.

We applaud the steps taken by the UK to contribute to the debate on AI governance globally, through publishing guidelines for secure AI development, producing a principles-based white paper on AI governance, and holding a global summit on AI Safety. The challenge facing policymakers around the world is how to establish a governance framework that will encourage the use of safe, secure and trustworthy AI for years to come.

At Microsoft, we're investing in the cloud technologies, the foundational infrastructure that AI is built on, to drive advances and accelerate adoption. Throughout our 40 year history in the UK, we've been supportive of a regulatory framework that guarantees safety and upholds fundamental rights while continuing to enable innovations that will ensure that the UK remains a globally competitive country. Our intention is to use our expertise to offer constructive contributions to help inform the work ahead. Collaboration with leaders and policymakers in the UK, across Europe and around the world is not only important but essential.

In this spirit, here we want to expand upon our five point blueprint, highlight how it aligns with the UK approach to date, and provide some thoughts on the opportunities to build on this governance framework.

A five-point blueprint for governing AI

1. Implement and build upon new government-led AI safety frameworks
2. Require safety brakes for AI systems that control critical infrastructure
3. Develop a broader legal and regulatory framework based on the technology architecture for AI
4. Promote transparency and ensure academic and public access to AI
5. Pursue new public-private partnerships to use AI as an effective tool to address the inevitable societal challenges that come with new technology

1

Implement and build upon new government-led AI safety frameworks

One of the most effective ways to accelerate government action is to build on existing or emerging governmental frameworks to advance AI safety.

A key element to ensuring the safer use of this technology is a risk-based approach, with defined processes around risk identification and mitigation as well as testing systems before deployment. Across the world, institutions are advancing on the development of such risk-based approaches, such as the AI Risk Management Framework developed by the US National Institute of Standards and Technology, or NIST, and the new international standard ISO/IEC 42001 on AI Management Systems.

Microsoft has committed to implementing the NIST AI risk management framework, and we will implement future relevant international standards. Opportunities to align such frameworks internationally should continue to be an important part of the ongoing dialogue between the UK and its partners, as highlighted in the International Technology Strategy, and as part of the upcoming Global AI Summit hosted by the UK.

The principles-based approach the UK Government's AI White Paper proposes affords a durable foundation, as technology continues to advance over time, society's expectations evolve, and new approaches to assessing risk and establishing safeguards are developed.

Given the broad and varied nature of how AI can be used, regulation should adopt a risk-based approach, focused on mitigating the risks posed by AI systems used in high-risk or consequential scenarios.

> "...a comprehensive approach to system safety that would keep effective human oversight..."

2

Require effective safety brakes for AI systems that control critical infrastructure

Increasingly, the public is debating questions around the control of AI as it becomes more powerful. Similarly, concerns exist regarding AI control of critical infrastructure like the electrical grid, water system, and traffic flows.

Our blueprint proposes new safety requirements that, in effect, would create safety brakes for AI systems that control the operation of designated critical infrastructure. These fail-safe systems would be part of a comprehensive approach to system safety that would keep effective human oversight, resilience, and robustness top of mind.

They would be akin to the braking systems engineers have long built into other tech such as school buses and high-speed trains, to safely manage not just everyday scenarios, but emergencies as well.

In this approach, the government would define the class of high-risk AI systems that control critical infrastructure and warrant such safety measures

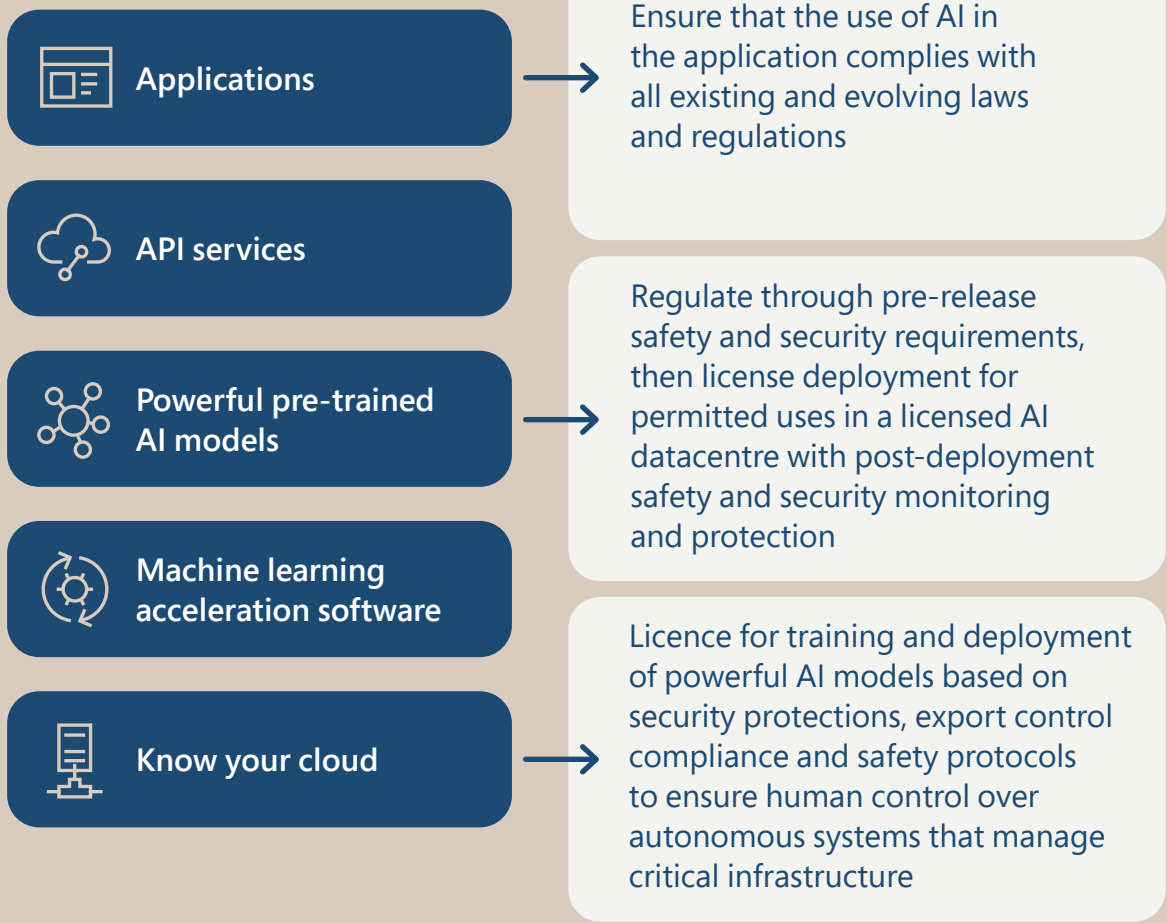
as part of a comprehensive approach to system management. New legislation would require operators of these systems to build safety brakes into high-risk AI systems by design. The government would then oblige operators to test high-risk systems regularly. And these systems would be deployed only in licensed AI datacentres that would provide a second layer of protection and ensure security.

3

Develop a broad legal and regulatory framework based on the technology architecture for AI

As we've worked the past year with AI models at the frontier of this new technology, we've concluded that it's critical to develop a legal and regulatory architecture for AI that reflects the technology architecture for AI itself. Regulatory responsibilities need to be placed upon different actors based on their role in managing different aspects of AI technology. Those closest to relevant decisions on design, deployment and use are best placed to comply with corresponding responsibilities and mitigate the respective risks, as they understand best the specific context and use-case.

A proposed AI regulatory architecture



The UK's AI White Paper's risk-based approach acknowledges the challenges to regulating the complex AI architecture. Our blueprint proposes that an AI regulatory architecture should mirror the AI technology stack. At the top of this stack is the application layer, where information and services are delivered to users. This is the layer where the safety and rights of people will most be impacted, especially because the impact of AI can vary markedly in different settings.

In the UK context, we welcome the UK Government's proposal to empower sectoral regulators to apply existing laws and advance a framework for responsible development and use of AI applications that is effective and coherent across sectors. Such application-level regulation is critical to ensuring that people have protection under the law and the context-specific nature of AI risks is properly accounted for. We also believe that attention is required at the other layers of the technology stack: for the most highly capable pre-trained AI models, and for the datacentre infrastructure that makes them possible.

It's also important to make sure that obligations are attached to powerful AI models, with a focus on a defined class of highly capable foundation models and calibrated to model-level risk. This will impact two layers of the technology stack. The first will require new regulations for these models themselves. And the second will involve obligations for the AI infrastructure operators on which these models are developed and deployed. The blueprint we developed offers suggested goals for each of these layers. The different roles and responsibilities require joint support. We are committed to helping our customers apply "Know Your Customer" (KYC) principles through our recently announced AI Assessment Program. Financial institutions use this framework to verify customer identities, establish risk profiles and monitor transactions to help detect suspicious activity.

We believe this approach can apply to AI in what we are calling "KY3C": know one's cloud, one's customers and one's content.

> "Even before the AI Act is implemented, we will test all our AI systems prior to release"

Promote transparency and ensure academic and nonprofit access to AI

It's also critical to advance the transparency of AI systems and broaden access to AI resources. While there are some inherent tensions between transparency and the need for security, there exist many opportunities to make AI systems more transparent. That's why Microsoft has committed to an annual AI transparency report and other steps to expand transparency for our AI services. Deceptive, AI-generated content or "deepfakes" – especially audiovisual content impersonating political figures – are of particular concern in terms of their potential harm to society and the democratic process.

We believe that there is benefit in requiring AI-generated audio-visual content to be labelled so that the public "knows the content" it is receiving and when it has been produced by an AI model rather than a human being.

In tackling this issue, we can start with building blocks that exist already. One of these is the Coalition for Content Provenance Authenticity, or C2PA, a

global standards body with more than 60 members including Adobe, the BBC, Intel, Microsoft, Publicis Groupe Sony, and Truepic. The group is dedicated to bolstering trust and transparency of online information including releasing the world's first technical specification for certifying digital content in 2022, which now includes support for Generative AI.

As Microsoft's Chief Scientific Officer, Eric Horvitz, said in 2022, "I believe that content provenance will have an important role to play in fostering transparency and fortifying trust in what we see and hear online." In May 2023, we announced new media provenance capabilities and plans to use C2PA to mark and sign AI-generated videos and images with metadata about their origin, enabling users to verify that a piece of content is AI-generated for images produced by Microsoft Designer and Bing Image Creator. Bing Image Creator now discloses content as AI-generated automatically.

There will be opportunities in the coming months to take important steps together on both sides of the Atlantic and globally to advance these objectives.

We also believe it is critical to expand access to AI resources for academic research and the nonprofit community. Unless academic researchers can obtain access to substantially more computing resources, there is a real risk that scientific and technological inquiry will suffer, including that relating to AI itself.

Our blueprint calls for new steps, including those we will take across Microsoft, to address these priorities.

Pursue new public-private partnerships to use AI as an effective tool to address the inevitable societal challenges that come with new technology

One lesson from recent years is that democratic societies often can accomplish the most when they harness the power of technology and bring the public and private sectors together. It's a lesson we need to build upon to address the impact of AI on society.

AI is an extraordinary tool. But, like other technologies, it too can become a powerful weapon, and there will be some around the world who will seek to use it that way.

We need to work together to develop defensive AI technologies that will create a shield that can withstand and defeat the actions of any bad actor on the planet. Important work is needed now to use AI to protect democracy and fundamental rights, provide broad access to the AI skills that will promote inclusive growth and use the power of AI to advance the planet's sustainability needs.

> "Democratic societies often can accomplish the most when they harness the power of technology"

Perhaps more than anything, a wave of new AI technology provides an occasion for thinking big and acting boldly. In each area, the key to success will be to develop concrete initiatives and bring governments, companies, and NGOs together to advance them. Microsoft will do its part in each area.

Protecting elections

Over the coming years, billions of people around the world will have the opportunity to vote in nationwide elections. From India to the European Union, to the United Kingdom and United States, the world's democracies will be shaped by citizens exercising one of their most fundamental rights. But while voters exercise this right, another force is also at work to influence and possibly interfere with the outcomes of these consequential contests.

As detailed in a new threat intelligence assessment published by Microsoft's Threat Analysis Center (MTAC), "Protecting Election 2024 from Foreign Malign Influence," this year may bring unprecedented challenges for the protection of elections. As described in the report, the world in 2024 may see multiple authoritarian nation states seek to interfere in electoral processes. And they may combine traditional techniques with AI and other new technologies to threaten the integrity of electoral systems.

We are grounding Microsoft's Election Protection Commitments in a set of principles to help safeguard voters, candidates and campaigns, and election authorities worldwide. These principles are:

Voters have a right to transparent and authoritative information regarding elections.

Candidates should be able to assert when content originates from their campaign and have recourse when their likeness or content is distorted by AI for the purpose of deceiving the public during the course of an election.

Political campaigns should protect themselves from cyber threats and be able to navigate AI with access to affordable and easily deployed tools, trainings, and support.

Election authorities should be able to ensure a secure and resilient election process and have access to tools and services that enable this process.

Staying ahead and responding to threats against voters, candidates, political campaigns, and election authorities will require a combination of steps, including a range of tools and tactics.

1 Microsoft will help candidates and campaigns maintain greater control over their content and likeness by launching Content Credentials as a Service. This new tool enables users to digitally sign and authenticate media using the Coalition for Content Provenance and Authenticity's (C2PA) digital watermarking credentials, a set of metadata that encode details about the content's provenance using cryptography. Users can attach Content Credentials to their images or videos to show how, when, and by whom the content was created or edited, including if it was generated by AI.

2 Microsoft will help political campaigns navigate cybersecurity challenges and the new world of AI by deploying a newly formed "Campaign Success Team" within Microsoft Philanthropies' Tech for Social Impact organisation. This team will advise and support campaigns as they navigate the world of AI, combat the spread of cyber influence campaigns, and protect the authenticity of their own content and images. The Campaign Success Team will also continue to promote existing cyber protection programs such as M365 for Campaigns and AccountGuard.

"Microsoft will empower voters with authoritative election information"

3 Microsoft will create and provide access to a new "Election Communications Hub" to support democratic governments around the world as they build secure and resilient election processes. This hub will provide election authorities with access to Microsoft security and support teams in the days and weeks leading up to their election, allowing them to reach out and get swift support if they run into any major security challenges. This new offering builds on existing security programs such as the Azure for Elections offering available to state and local election agencies and their partners in the US.

4 We will use our voice as a company to support legislative and legal changes that will add to the protection of campaigns and electoral processes from deepfakes and other harmful uses of new technologies.

5 Microsoft will empower voters with authoritative election information on Bing. We will do this in partnership with organisations that provide information on authoritative sources, ensuring that queries about election administration will surface reputable sites. Microsoft will also publish regular reports on foreign malign influence researched and reported by the company's MTAC team.

International partnership to advance AI governance

The UK's early start towards AI governance offers an opportunity to establish an effective governance framework, grounded in the rule of law. But beyond legislative frameworks at the level of nation states, multilateral public-private partnership is needed to ensure AI governance can have an impact today, not just a few years from now, and at the international level.

This is important to serve as an interim solution before regulations come into effect, but, perhaps more importantly, it will help us work towards a common set of shared principles that can guide both nation states and companies alike.

In parallel to the UK's focus on AI governance, there is an opportunity for the UK, the European Union, the United States, the other members of the G7 as well as India and Indonesia as part of the AI Hiroshima Process and related next steps, to move forward together on a set of shared values and principles. If we can work with others on a voluntary basis, then we'll all move faster and with greater care and focus. That's not just good news for the world of technology, it is good news for the world as a whole.

Working towards a globally coherent approach is important, recognising that AI – like many technologies – is and will be developed and used across borders. And it will enable everyone, with the proper controls in place, to access the best tools and solutions for their needs.

We are very encouraged by recent international steps, including the annual G7 Summit in Hiroshima in May 2023, where leaders committed to "advance international discussions on inclusive artificial intelligence (AI) governance and interoperability to achieve our common vision and goal of trustworthy AI, in line with our shared democratic values."

> "Multilateral public-private partnership is needed to ensure AI governance can have an impact today, not just a few years from now"

To make the many different aspects of AI governance work on an international level, we will need a multilateral framework that connects various national rules and ensures that an AI system certified as safe in one jurisdiction can also qualify as safe in another. There are many effective precedents for this, such as common safety standards set by the International Civil Aviation Organisation, which means an airplane does not need to be refitted mid-flight from London to New York.

Before the formal AI regulations come into force, it is important that we take steps today to implement safety brakes for AI systems that control critical infrastructure. The concept of safety brakes, along with licensing for highly capable foundation models and AI infrastructure obligations, should be key elements of the voluntary, internationally coordinated G7 code that signatory nation states agree to incorporate into their national systems.

High-risk AI systems, relating to critical infrastructure (e.g., transport, electrical grids, water systems) or systems that can lead to serious fundamental rights

In our view, an international code should:

- Build on the work already done at the OECD to develop principles for trustworthy AI
- Provide a means for regulated AI developers to attest to the safety of these systems against internationally agreed standards
- Promote innovation and access by providing a means for mutual recognition of compliance and safety across borders

violations or other significant harms, could require additional international regulatory agencies, based on the model of the International Civil Aviation Organisation, for example.

Lastly, we must ensure academic researchers have access to study AI systems in depth. There are important open research questions around AI systems, including how one evaluates them properly across responsible AI dimensions, how one best makes them explainable, and how they best align with human values.

The work the OECD is doing on the evaluation of AI systems is seeing good progress. But there's an opportunity to go further and faster by fostering international research collaboration and boosting the efforts of the academic communities by feeding into that process. The UK is well placed to lead on this, partnering with the US and the EU.

AI governance is a journey, not a destination.

No one has all the answers, and it is important we listen, learn and collaborate. Strong, healthy dialogue between the tech industry, governments, businesses, academia and civil society organisations is vital to make sure governance keeps pace with the speed at which AI is developing. Together, we can help realise AI's potential to be a positive force for good.

"AI governance is a journey, not a destination. No one has all the answers, and it's important we listen, learn and collaborate"

Microsoft

AI

Created with sustainability in mind

This publication is printed on FSC-certified paper
produced from responsibly managed forests