

CHAPTER ONE

Counting the Cost

“AI is more dangerous than, say, mismanaged aircraft design or production maintenance or bad car production...it has the potential – however small one may regard that probability, but it is non-trivial – it has the potential of civilisation destruction.” Elon Musk

“I believe there are more reasons than not to be optimistic that we can manage the risks of AI while maximising their benefits but we need to move fast.” Bill Gates

It's straight out of George Orwell: machines that do your thinking and prompting, creating a new and controversial technology – artificial intelligence. The coming of age of AI has resulted in the most powerful technologies ever created. We already live in a world where our smartphone is our constant companion, always tracking our whereabouts. Many countries are using Pegasus, a spyware developed by the Israeli cyber-arms company NSO Group that is designed to be covertly and remotely installed on mobile phones running iOS and Android. Orwell wrote about the telescreen that constantly watched the people who watched it. Now AI has raised serious doubts about the benefits of technology. AI has the power to eliminate hundreds of thousands of jobs. There have already been massive layoffs in Big Tech because of the overbearing focus on AI. Thinking machines, which is what AI is, can do almost anything a human can, and more. A majority of humans have no understanding of what the Tech Giants in one compact area of Silicon Valley who already rule our digital world, are doing with AI, and we are largely left to the whims of their creation. It's like creating conscious machines, without most of the human population it serves having a say on the ethical and moral outcomes that AI will bring in its wake. It's a throwback to the iconic science fiction novel *The Hitchhiker's Guide to the Galaxy* which featured a supercomputer designed to figure out the answer to the ultimate question of “Life, the Universe and Everything”. That machine was called Deep Thought and it is no coincidence that Google's AI platform is called Deepmind. In fact, so revolutionary is the promise of AI that some experts believe that its projected power – over individuals, societies, battlefields, the creative arts, existing laws and systems of governance – will mean that a few companies, and countries, will dominate us all, exercising a monopoly that transcends current boundaries: physical, financial, ethical, moral, psychological and geographical.

November 1, 2023, witnessed the first ever summit on AI held, symbolically, at Bletchley Park, the same site where Alan Turing and his team created the world's first

programmable computer. Serving British Prime Minister Rishi Sunak had earlier shared plans to create the world's first "AI Safety Institute" and propel Britain into taking the lead on AI and to regulate its possible global ramifications. Delegates from 28 nations, including the US, India, Japan, France, the EU and China, agreed to work together to contain the potentially "catastrophic" risks posed by galloping advances in artificial intelligence. The summit focussed on cutting-edge "frontier" AI that has created that some scientists warn could pose a risk to humanity's very existence. Sunak said the declaration was "a landmark achievement that sees the world's greatest AI powers agree on the urgency behind understanding the risks of AI – helping ensure the long-term future of our children and grandchildren." US Vice President Kamala Harris urged Britain and other countries to go further and faster, stressing the transformations AI is already bringing and the need to hold tech companies accountable, including through legislation. She pointed to President Joe Biden's executive order in November, 2023, setting out AI safeguards as evidence the US is leading by example in developing rules for artificial intelligence that work in the public interest. Harris also encouraged other countries to sign up to a US-backed pledge to stick to "responsible and ethical" use of AI for military aims. "President Biden and I believe that all leaders ... have a moral, ethical and social duty to make sure that AI is adopted and advanced in a way that protects the public from potential harm", she said.

Getting the nations to sign the agreement, dubbed the Bletchley Declaration, was an achievement even if it was light on details and does not propose a way to regulate the development of AI. Some delegates at the Summit felt regulation was not possible given how fast the AI industry is evolving. The countries represented did pledge to work toward "shared agreement and responsibility" about AI risks, and hold a series of further meetings. South Korea and France are also planning to host virtual AI summits. Crucially, the summits will also include tech entrepreneurs and many other CEO's, intellectuals and prime movers in the technology field. The Bletchley Declaration, which was also later endorsed by Brazil, Ireland, Kenya, Saudi Arabia, Nigeria, and the United Arab Emirates, incorporates an acknowledgment of the substantial risks from potential intentional misuse or unintended issues of control of frontier AI – especially cybersecurity, biotechnology, and disinformation risks, according to the UK government, the summit host. Tech trillionaire Elon Musk issued a dire warning, saying that AI represents "one of the biggest threats to humanity" given its potential to become far more intelligent than people. "So, you know, we're not stronger or faster than other creatures, but we are more intelligent. And here we are, for the first time really in human history, with something that's going to be far more intelligent than us", he said at the summit. In essence the Declaration says: "There is potential for serious, even catastrophic, harm, either deliberate or unintentional, stemming from the most significant capabilities of these A.I. models. Many risks arising from A.I. are inherently international in nature, and so are best addressed through international cooperation. We resolve to work together in an inclusive manner to ensure human-centric, trustworthy and responsible A.I."

The European Union on December 11, 2024, declared that it would become "the

first continent to set clear rules for the use of AI.” The EU’s first-of-a-kind AI Act, which was proposed by the European Commission two years ago, could serve as the benchmark for other countries. The deal, still to be ratified, would be, the EU stated, the first asset of comprehensive laws to regulate the use of artificial intelligence. The proposed legislation includes safeguards on the use of AI in the 24 European member states, setting clear guardrails on its adoption by law enforcement agencies, including strong restrictions on facial recognition technology, and using AI to manipulate human behaviour, among other restrictions on the increasing usage of AI in other areas, basically anything that impacts people’s safety and rights. The EU laws are seen as an attempt by governments and geopolitical blocs to take the lead on governance of AI, another arms race if you will, but restricted to the digital domain involving Artificial Intelligence. The EU proposals are crucial in terms of the rapid and largely unchecked growth of AI.

On March 21, 2004, the UN General Assembly adopted a landmark resolution on the promotion of “safe, secure and trustworthy” artificial intelligence (AI) systems. Adopting a United States-led draft resolution without a vote, the Assembly also highlighted the respect, protection and promotion of human rights in the design, development, deployment and the use of AI. The text was “co-sponsored” or backed by more than 120 other Member States. It represents the first time the Assembly has adopted a resolution on regulating the emerging field. The US National Security Advisor reportedly said that the adoption would represent an “historic step forward” for the safe use of AI.

Right now, as we are seeing the battle for AI enter a turbulent phase with OpenAI and its ChatGPT-4 becoming the focus of the AI technology race amid the shock sacking and equally surprising re-instatement of its founder and CEO, Sam Altman. He co-founded the company with Musk who later pulled out. OpenAI’s biggest investor to the tune of \$10 billion is actually a rival, Microsoft, which has its own AI platforms. The day after Altman’s sacking, Microsoft CEO Satya Nadella announced that Altman and OpenAI President Greg Brockman, who had also resigned, would be joining Microsoft to lead a new team for advanced research in Artificial Intelligence. In a head-spinning development, four days after their sacking, Altman and Brockman were back at OpenAI after a new board was appointed. The company’s original leadership wanted slow and steady growth with safeguards over how it is used. Altman and the company’s software whiz kids wanted rapid development and sale of the software. The Altman issue signals a churn in the AI space with the competition with other Tech giants becoming increasingly hostile and alarming in terms of what they are working on and whether it comes with existential risks. Altman’s sacking was evidently because he wanted the company to move faster on AI applications and their commercial use without keeping the board in the loop. Altman has been the poster boy of AI in public and private discussions about the social implications of the software. His sacking was clouded with rumours but it clearly had to do with a breakdown in communications with the then board of directors. Cited by *The New York Times*, the board accused Altman with lack of transparency. The board’s memo said,

“Put simply, Sam’s behaviour and lack of transparency in his interactions with the board undermined the board’s ability to effectively supervise the company in the manner it was mandated to do.” The board, mindful of its stated values at startup, wanted safeguards for the public good, investing in development with less concern about market valuations and stock options.

Months before OpenAI board member Ilya Sutskever would gain notoriety for his key role in the ouster of Altman, he co-authored a little-noticed but apocalyptic warning about the threat posed by artificial intelligence. Superintelligent AI, Sutskever co-wrote on a company blog, could lead to “the disempowerment of humanity or even human extinction”, since engineers are unable to prevent AI from “going rogue”. He quit the company later citing OpenAI’s charter which calls for avoiding AI uses if they “harm humanity”. The cry for caution from Sutskever, however, arrived at a period of breakneck growth for OpenAI. A \$10 billion investment from Microsoft helped fuel the development of ChatGPT-4, a viral conversation bot that the company says now boasts 100 million weekly users. GPT-5 is due for release shortly, with major implication for the future – and fears – of the technology. Expressing his concern over the matter, Musk recently posted on X, formerly Twitter, “Very important for the public to know why the board felt so strongly about their actions. If it was a matter of AI safety that would affect all of Earth.” He added, “OpenAI is not publicly traded and all the money in the world won’t matter if AI goes wrong.” Ahead of Altman’s four-day exile, staff researchers, according to two who spoke to *Reuters*, said that they had sent the board a letter warning of a powerful AI discovery called Q* that could pose a risk to humanity. That, apparently, was the issue that prompted the board to fire Altman, who, in an act of public retribution, fired them. What the Altman episode also made clear is that a handful of tech entrepreneurs now represent the interests and future of the remaining 7.9 billion inhabitants of the planet. That is a scary reality. Since the explosive launch of ChatGPT, several firms such as Google, Meta, Microsoft and Anthropic have launched their own models such as Gemini, Llama, Co-pilot and Claude, which are seen as being better than ChatGPT in numerous areas.

Artificial Intelligence is a game-changer in many areas of business and life. Some experts believe there will be a competitive AI landscape as companies seek ways to optimise business operations with machine learning, for example. Altman’s pursuit of profits and his growing ambitions to build a world-spanning consumer business triggered concerns among some OpenAI board members that the company had abandoned its founding principles to be a counterweight to Big Tech. *The Washington Post* reported that, “Many in the technology world credited him with infusing new energy and a sense of possibility into a beleaguered sector that has been dominated by tech giants like Google and Amazon for more than a decade.” Altman was clearly hoping to enter the league of tech titans such as billionaire Elon Musk, Meta CEO Mark Zuckerberg and even the late Apple CEO Steve Jobs (Altman has made it to the latest *Forbes* billionaire list). OpenAI’s ChatGPT was a pioneer in the AI space and had a spectacular launch. OpenAI’s dilemma was moving the technology forward to stay ahead of a horde

of competition while not disrupting humanity in the process. As British author Aldous Huxley reminded us way back in 1932 in his dystopian novel *Brave New World*: “People will come to love their oppression, to adore the technologies that undo their capacity to think.” AI is the most advanced technology since the birth of the internet, and potentially the most contentious. A few days after *Time* magazine asked on its cover if AI represented the death of humanity, leading tech experts and CEOs from Silicon Valley released this startling statement in May, 2023: “Mitigating the risks of extinction from AI should be a global priority, alongside other societal-scale risks such as pandemics and nuclear war.” Among the signatories were Altman and Dennis Hassabis, the CEO of Google’s AI platform, DeepMind. That reflects the most memorable prediction by the late globally renowned physicist Stephen Hawking who said in 2014 that: “full artificial intelligence could spell the end of the human race.”

The troubling issue facing the world is whether AI’s incredible technology will allow machines to replace humans and lead to mass unemployment and all the risks that go with such an eventuality. Elon Musk, among the world’s richest men, co-founded OpenAI in 2015 but then pulled out and is now focussed on his own AI-based cars, robots and neural implants. Microsoft has a majority stake in OpenAI and is now doubling down on its AI platforms. As its CEO Satya Nadella says: “The age of AI is upon us and Microsoft is powering it.” Its AI platforms include Bing, Edge, Co-Pilot and Microsoft Build. Microsoft is currently one of the leaders in the AI race. The Washington-based tech giant is actively involved in numerous AI projects, including its investments in OpenAI research and deployment companies. However, its AI initiatives are demanding a significant amount of power and water for cooling data centres. The data centres that support AI operations rely heavily on electricity, and the enormous power consumption required for AI platforms is in conflict with Microsoft’s goal of transitioning to renewable energy sources. To address this, the company appears to be considering a shift towards nuclear energy to meet the demands of its AI operations. According to a job listing, the company is currently seeking a principal program manager who will be responsible for developing and implementing a global Small Modular Reactor (SMR) and microreactor energy strategy. “This senior position is tasked with leading the technical assessment for the integration of SMR and microreactors to power the data centres that the Microsoft Cloud and AI reside on”, reads the job description. That suggests that Microsoft is planning to adopt nuclear power as a solution to feed its power-hungry AI projects and is seeking someone to develop a plan for SMRs, which are next-generation nuclear reactors. They are also expected to be simpler and more cost-effective to construct than older, much larger reactors. However, nuclear projects bring along a host of other concerns which are not in line with the company’s environmental goals. Reportedly, Microsoft has not yet clarified how it plans to address these challenges, such as sourcing highly enriched uranium fuel to operate nuclear reactors or managing the significant amount of nuclear waste generated by SMRs. The race to become the leader in AI technology is clearly fraught with risks and reversals of earlier positions.

Elon Musk, after pulling out of OpenAI, launched his AI startup, a chatbox called Grok tightly integrated with X (formerly twitter), to take on OpenAI, Google, Meta

and Microsoft. Others, however, fear that the new tech bubble is potentially harmful as commercialisation of the technology remains at an early stage. Google, in December 2023, introduced its new AI model, Gemini. The model is designed to behave in human-like ways, better than what other models can achieve. Gemini is a new artificial intelligence model that outperforms other models in tasks like understanding, summarising, reasoning, coding, and planning. It comes in three versions: Pro, Ultra, and Nano. The Pro version is already available, and the Ultra version will be released in 2024. What that does to existing jobs is hard to tell at this stage in AI's evolution. The founder of Nvidia, the sixth largest technology company in the world, Jensen Huang, has predicted that his vision is to unify the company's computer-graphics research with its generative AI to produce three dimensional, inhabitable worlds and populate them with realistic-seeming people, as he told *The New Yorker*. He calls it the Omniverse. It sounds like science fiction but the tech giants working on AI believe in its reality. The reality is that it could lead to mass unemployment as thinking machines replace humans. Both Gemini and Nvidia have come under fire in recent times, one for messing up its answers and showing obvious bias and Nvidia which makes a majority of the chips that power AI, being sued for using material that breaches copyright laws.

Global management consultants McKinsey & Company predicts that, depending upon various adoption scenarios, automation will displace between 400 and 800 million jobs by 2030, requiring as many as 375 million people to switch job categories entirely. How could such a shift not cause fear and concern, especially for the world's vulnerable countries and populations? The Brookings Institution writes of a "new" kind of automation with more advanced robotics and AI that can bring work displacement to college graduates and professionals as much as it has to vehicle drivers and retail workers. Blue-collar and white-collar jobs will be eliminated – basically, anything that requires middle skills (meaning that it requires some training, but not much). But will the people who lost their middle-skilled jobs be able to move into these high-skill roles instead? Certainly not without significant training and re-education. If the middle class loses jobs and stops spending money on food services, gardening, homes, health, etc., the transition could be very painful. It's no secret that rising unemployment has a negative impact on society; less volunteerism, higher crime, and drug abuse are all correlated. A period of high unemployment, in which tens of millions of people are incapable of getting a job because they simply don't have the necessary skills, could well be a reality. Already, these low skilled jobs are quickly being overtaken by AI. Even high-skilled jobs are under threat. US television and film writers went on strike since the start of May 2023. One of their key demands is that the studios and streaming giants agree to limits on the future use of AI-powered writing tools, such as ChatGPT. The writers and their union – the Writers Guild of America – wanted it in writing that AI can only be used for research purposes, and not to ever replace them. In the past, technology has tended to increase rather than reduce total employment, but economists acknowledge that "we're in uncharted territory" with AI. A survey of economists showed disagreement about whether the increasing use of robots and AI will cause a substantial increase in long-term unemployment. In one

early study, Michael Osborne and Carl Benedikt Frey estimated 47% of US jobs are at “high risk” of potential automation, while an (OECD) report classified only 9% of US jobs as “high risk”. The methodology of speculating about future employment levels has been criticised as lacking evidential foundation. Unlike previous waves of automation, many middle-class jobs may be eliminated by artificial intelligence; *The Economist* stated in 2015 that “the worry that AI could do to white-collar jobs what steam power did to blue-collar ones during the Industrial Revolution” is “worth taking seriously”. Jobs at extreme risk range from paralegals to radiologists and factory workers, while job demand is likely to increase for care-related professions. In April 2023, it was reported that 70% of the jobs for Chinese video game illustrators had been eliminated by generative artificial intelligence.

AI basically means thinking machines that are almost exclusively being built in a 50 kilometre area in Northern California that is the global centre for technology and innovation and is located in the southern part of the San Francisco Bay. The radical, life-changing new technology is being powered by a few hundred men and women who write in coded language only they and computers can understand. The rest of the world does not have a say in the ethics behind their innovations. They don't have a say over whether it should even exist in the first place. “We're recreating God”, one AI engineer working on large language models (LLMs) recently said in a media interview, “We're creating conscious machines.” That can mean many things to many people. Brain-computer interface technology, like that of Elon Musk's Neuralink, is capable of implanting monitoring smart devices into our brains. What if our thoughts could be read, or worse, influenced by business or government agencies? Unlike earlier technological advances, AI comes with many ethical challenges: Lack of transparency, decisions that are not always intelligible to humans. Further, AI is not neutral. Its decisions are vulnerable to inaccuracies, discriminatory outcomes or bias. The big question of course is whether artificial intelligence can be kept under control? Jimmy Wales, the founder of Wikipedia, says that believing it can be is akin to “magical thinking”. “In many cases, politicians and their aides have a weak understanding of how the internet works, and what it is possible to achieve”, says Wales, who has spent many hours explaining the technology and its role in free speech to politicians around the globe.

The issue of whether AI should be regulated, and to what extent, heated up this summer when UN Secretary General António Guterres convened the first ever UN Security Council meeting to specifically discuss its potential danger. Speaking in regard to everything from AI-powered cyberattacks, to the risk of malfunctioning AI, how it can spread misinformation, and even the interaction between AI and nuclear weapons, Guterres said: “Without action to address these risks, we are derelict in our responsibilities to present and future generations.” Guterres has since moved forward with the establishment of a UN panel to investigate what global regulation might be needed. Called the High-Level Advisory Body for Artificial Intelligence, this will comprise “present and former government experts, as well as experts from industry, civil society, and academia”.

However, some AI insiders are sceptical that global regulation can be successful. One such person is Pierre Haren, who has been researching AI for many years. His experience includes seven years at computer giant IBM, where he led the team that launched the Watson super computer in 2010. Watson can answer user's questions, and was one of the pioneers of AI. Despite Haren's background, he says he was "flabbergasted" by the emergence and capability of ChatGPT and other "generative AI" programs in the recent past. Generative AI is, put simply, technology that can quickly, often almost immediately, create new content, be it words, images, music or videos. And it can take an idea from one example, and apply it to an entirely different situation. Haren said in an interview that such an ability is human-like. "This thing is not like a parrot, repeating what we feed into it", he said. "It's making high-level analogies." Can there be a set of rules to stop AI getting out of control? "We live in a world with non-cooperative nations like North Korea and Iran", he says. "They won't recognise regulations around AI." In other words, AI can only be regulated by individual countries. In his written testimony before a Senate Judiciary sub-committee, Microsoft President Brad Smith wrote: "Congress should ensure that AI is used in a manner that complies with longstanding legal protections for consumers and citizens. This should include the protection of privacy, civil rights, and the needs of children, as well as safeguards against dangerous deepfakes and election interference."

The key issue remains the invisible risks. In his sci-fi novel *Do Androids Dream of Electric Sheep?* Philip K. Dick considers the idea that our understanding of human subjectivity is altered by technology created with artificial intelligence. In a recent *Time* magazine article on the most influential people in AI, Lila Ibrahim, chief Operating Officer at Google's DeepMind project, recalls being struck by the thought of just how severe the risks could be. "I have twin daughters, I kept thinking, can I tuck them in at night?" She has obviously overcome those fears but others are not so sanguine. In the same article, Alondra Wilson, former director of the White House Office of Science and Technology, says: "AI is already being used for deepfake videos that target politicians and could be weaponised by lobbyists to disseminate falsehoods on a much larger scale. Disinformation and misinformation are going to get a lot worse." Machine learning applications will be biased if they learn from biased data. The developers may not be aware that the bias exists. Bias can be introduced by the way training data is selected and by the way a model is deployed. If a biased algorithm is used to make decisions that can seriously harm people (as it can in medicine, finance, the military or policing) then the algorithm may cause discrimination. It has become a serious area of academic study within AI. Researchers have discovered it is not always possible to define "fairness" in a way that satisfies all stakeholders. On June 28, 2015, Google Photos new image-labelling feature mistakenly identified Jacky Alcine and a friend as "gorillas" because they were black. The system was trained on a dataset that contained very few images of black people; a problem called "sample size disparity". Google says it "fixed" this problem by preventing the system from labelling *anything* as a "gorilla". Eight years later, in 2023, Google Photos still could not identify a gorilla, and neither could similar products from Apple, Facebook, Microsoft and Amazon.

AI has created many new phrases and applications, one being "Transhumanism". Robot designer

Hans Moravec, cyberneticist Kevin Warwick, and inventor Ray Kurzweil have predicted that humans and machines will merge in the future into cyborgs that are more capable and powerful than either. This idea, called transhumanism, has roots in Aldous Huxley and science fiction writer Robert Ettinger. The late Edward Fredkin, the American computer scientist, physicist and businessman who was an early pioneer of digital physics, argued that “artificial intelligence is the next stage in evolution”, an idea first proposed by Samuel Butler’s essay “Darwin among the Machines” as far back as 1863, and expanded upon by George Dyson in his book of the same name in 1998. Dyson’s book is a story of humankind’s journey into a digital wilderness. As Greg Brockman, President and co-founder of OpenAI said in a recent interview: “Imagine you actually built an AGI (Artificial General Intelligence) a system that can match human performance on all cognitive tasks and it’s your first-time deploying AGI. Are you going to get that right?” In end-April, 2024, Amnesty International’s annual report warned about the threat of new technologies if left unchecked, saying the rapid advancement in artificial intelligence and mass surveillance tools could be deployed to fuel conflict, encroach on rights and freedoms and sow discord in a landmark election year (2024) and that unregulated tech advances “can be weaponised to discriminate, disinform and divide.” The latest edition of *Time* 100 ranks the most influential people in the world, among them Yoshua Bengio, known as the godfather of AI. The profile quotes world renowned computer scientist Geoffrey Hinton as saying Bengio “has been concerned for many years with the social impact of artificial intelligence...and worried by the potential for catastrophic outcomes from future AI” and lauds him for his contribution to the landmark Montreal Declaration for the Responsible of AI and his focus on safety protocols.

It’s clearly a brave new world, or a grave new world, depending on the positives of AI and the negatives. Right now, with GAFAM (Google, Apple, Facebook, Amazon, and Microsoft) and OpenAI in frenetic competition to turn AI into the Next Big Thing, it will change the world as we know it, but in what direction and with what consequences? That is the allure and the danger of revolutionary new technology. In *The Economist* issue of October 28, 2023, it asked: “Will Artificial intelligence kill us all? In one nightmarish scenario, AI outsmarts humanity and goes rogue, taking over computers and factories and filling the sky with killer drones. In another, Large Language Models (LLMs) of the sort that power generative AI like ChatGPT give the bad guys the know-how to create devastating cyberweapons and deadly pathogens.” Now that Ukraine, Russia, Israel, Iran and the Iran-backed Houthis are flooding the skies with drones there is the real possibility that drones using AI could be turned into autonomous weapons systems, able to observe the battlefield from above and take combat decisions independent of human guidance. No technology is perfect and autonomous drones are in danger of malfunctioning or making an inaccurate decision which could endanger innocent lives.

In a recent simulation exercise, a US autonomous weapon system turned against its operator during combat operations, raising serious issues and questions about the increasing use of AI in warfare. The *Warzone* reported that during the Royal Aeronautical Society’s Future Combat Air and Space Capabilities Summit in London in May 2024, Colonel Tucker Hamilton, US Air Force Chief of Artificial Intelligence AI Test and Operations, described a simulation wherein an AI-enabled drone was tasked with

suppression of an air defence mission against surface-to-air missile (SAM) sites, with the final engagement order to be given by its human operator. Although Hamilton noted that the AI-enabled drone was trained not to go against its human operator, the drone attacked the communications tower used by the latter to communicate with the former, then went on to destroy the SAM site. While Hamilton stressed the hypothetical nature of the experimental simulation, he said that scenario illustrates what can happen if fail-safes such as geo-fencing, remote kill switches, self-destruct and selective disabling of weapons are rendered moot.

Before the AI summit near London, Washington was setting the stage for defining an AI regulation rulebook by starting public consultations on how to regulate artificial intelligence tools. This follows a move by the White House Office of Science and Technology Policy to unveil a Blueprint for an AI Bill of Rights. Serving President Joe Biden has issued a series of orders that requires AI companies to share the results of tests of their newer products with the federal government before making the new capabilities available to consumers. The safety tests undertaken by developers, known as “red teaming”, are aimed at ensuring that new products do not pose a threat to users or the public at large. Following the order, the federal government is empowered to force a developer to tweak or abandon a product or initiative. “These measures will ensure AI systems are safe, secure, and trustworthy before companies make them public”, the White House said. Thus, a new rule seeks to codify the use of watermarks that alert consumers to a product enabled by AI, which could potentially limit the threat posed by content such as deepfakes. Another standard asks biotechnology firms to take appropriate precautions when using AI to create or manipulate biological material. Overall, 52% of Americans said they felt more concerned than excited about the increased use of artificial intelligence. Just 10% say they are more excited than concerned, while 36% say they feel an equal mix of these emotions. The share of Americans who are mostly concerned about AI in daily life is up 14 percentage points since December 2022, when 38% expressed this view. Concern about AI outweighs excitement across all major demographic groups. Still, there are some notable differences, particularly by age. About six-in-ten adults ages 65 and older (61%) are mostly concerned about the growing use of AI in daily life, while 4% are mostly excited. That gap is much smaller among those ages 18 to 29: 42% are more concerned and 17% are more excited.

China too has released its own set of measures to regulate AI. This also comes in the wake of calls by tech leaders Elon Musk, Steve Wozniak (Apple co-founder) and over 15,000 others for a six-month pause in AI development in April 2023, saying labs are in an “out-of-control race” to develop systems that no one can fully control. Geoffrey Hinton, a British-born computer scientist who some describe as one of the Godfathers of AI, quit his job at Google in 2023 so that he could voice his fears of AI more openly. He told *The Guardian Weekly*, “My confidence was shaken by the realisation that biological intelligence and digital intelligence are very different” and expressed the view that his main concern was that AI firms were trying to build intelligences with the potential “to outthink humanity”.

These developments come as policymakers across jurisdictions have stepped up regulatory scrutiny of generative AI tools, prompted by ChatGPT's game-changing launch. The concerns being flagged fall into three broad heads: privacy, system bias and violation of intellectual property rights. The policy response has been different too, across jurisdictions, with the European Union having taken a predictably tougher stance by proposing to bring in a new AI Act that segregates artificial intelligence according to use-case scenarios, based broadly on the degree of invasiveness and risk. At an event on September 29, 2023, UK Prime Minister Rishi Sunak said: "While it (AI) offers significant opportunities, it could also pose an existential threat". A paper from the UK Government Office for Science and Technology says: "It could be used to commit fraud and mount cyberattacks and, more alarmingly, it will enhance terrorist capabilities in developing weapons, including biological weapons." Simultaneously, Britain is adopting a "light-touch" approach that aims to foster, and not stifle, innovation in this nascent field.

The obvious danger is that AI provides a number of tools that are particularly useful for authoritarian governments: smart spyware, face recognition and voice recognition allow widespread surveillance which helps classify so-called "enemies of the state". More alarmingly, terrorist organisations, criminals and rogue states can use weaponised AI such as advanced digital warfare and lethal autonomous weapons. This clearly shows that AI is such a huge leap in technology that countries are still battling to impose some kind of guardrails, knowing the dangers that come with the alarming growth of artificial intelligence. To add to the alarm bells, tech industry watchers say that Microsoft has laid off an entire team dedicated to guiding AI innovation that leads to ethical, responsible and sustainable outcomes. The axing of the ethics and society team, as reported by *Platformer*, is part of a recent spate of layoffs that affected 10,000 employees across the company. This comes as Microsoft invests billions more dollars into its partnership with OpenAI, the startup behind art- and text-generating AI systems like ChatGPT and DALL-E 2, and revamps its Bing search engine and Edge web browser to be powered by a new, next-generation large language model that is "more powerful than ChatGPT and customised specifically for search." The move calls into question Microsoft's commitment to ensuring its product design and AI principles are closely intertwined at a time when the company is making its controversial AI tools available to the public. Already, we are seeing instances across the world, including in India, where extortionists are making calls telling terrified parents that their child has been kidnapped and demanding a ransom for their release, along with a video of the child that is so authentic in voice and mannerisms that most parents give in to the demands before finding out their offspring is alive and well. Generative AI can use an individual's social media posts and videos to create a clone of the person and use it as a weapon of blackmail. If parents cannot tell if a video of their offspring has been morphed, there is no limit to the damage that criminals can do with AI.

That explains why the jury is still out on whether AI constitutes a threat or a new phase in humankind's evolution. As an article in the *Scientific American* put it: "Artificial intelligence is everywhere and it poses a monumental problem for those who

should monitor and regulate it.” The Global Partnership on Artificial Intelligence (GPAI) was signed by only 15 countries and has made little headway since. In fact, most governments have struggled to launch a coordinated and concentrated global effort leading to fears that by the time the rapidly expanding AI field completely dominates our world, it may be too late.