



WITSA

World Innovation, Technology
and Services Alliance



OCTOBER, 2023

**BUILDING TRUST AND
DELIVERING ON THE PROMISE
OF ARTIFICIAL INTELLIGENCE**

CONTENTS

WITSA Chairman's Statement	4
WITSA AI Taskforce Chairman's Statement	5
WITSA Global Policy Action Committee Chairman's Statement	6
WITSA CEO's Statement	7
Executive Summary	8
About WITSA	11
AI Policy Principles for Governments	13
Responsible Best Practices in AI	22
Fostering Trustworthy AI through Standards and International Cooperation	31
History of Artificial Intelligence	34
What is AI?	39
Benefits of AI	47
Risks & Threats of AI	52
Conclusion	58



WITSA CHAIRMAN'S STATEMENT

Founded in 1978, the World Innovation, Technology and Services Alliance (WITSA) is a leading consortium of tech association members from over 80 countries/economies around the world.

As the leading recognized voice of the global tech industry, we aim to drive transformation and grow the industry given that tech is the key driver of the global economy. We advocate international public policies that advance the industry's growth and development, facilitate international trade and investment in tech products and services through our global network of contacts, and promote industry cooperation and strengthening our national associations through the sharing of knowledge, experience, and critical information.

As the challenges facing the tech industry are undisputedly global in nature, our members work together to achieve a shared vision on important issues of common interest. We make it possible for our members throughout the world to identify common issues and priorities, exchange valuable information, and present a united position on industry issues.

Aiming to fulfill the Promise of the Digital Age for everyone, we work closely with all stakeholders, including internationally recognized organizations to promote and facilitate growth to the benefit of all.

We strive to position digital technology and the digital economy as the new strategic pillar for our global community, to become a global digital ambassador and accelerator in order to promote the value of digital transformation as the key catalyst for economic growth and social prosperity, to promote greater international cooperation by supporting multilateral win-win trade agreements, and to drive awareness of the competitive advantage of the digital technology industry.

With this paper, WITSA calls for a sensible approach to build trust and deliver on the promise of artificial intelligence. It is my hope that this publication can serve as a helpful and constructive contribution to the crucial AI-related policy making processes taking place around the world by helping our members set ethical standards, educate decision-makers, engage stakeholders, and foster transparency and accountability, ultimately contributing to the beneficial and responsible use of AI in society.

Dr. Sean Seah
WITSA Chairman

WITSA AI TASKFORCE CHAIRMAN'S STATEMENT

New AI models stand to transform humans' relationship with computers, knowledge and even with themselves. Everyone and everything now seem to be pursuing such fine-tuned models as ways of providing access to knowledge. AI has the potential to solve some of humankind's biggest problems by developing new drugs, designing new materials to help fight climate change, and even untangling the complexities of fusion power.

Fears about AI have reached new levels because of the emergence of generative AI, which promises to democratize the creative sector and enable entirely new forms of innovation. This novelty has impressed many technology enthusiasts but alarmed others—especially those who believe AI is encroaching on creativity, which many people believe to be an essential difference that separates humans from machines.

However, we should not rush to impose regulations, or even a pause. As explained in this paper, existing AI systems will need to address such issues as bias, privacy and intellectual-property rights, and as the technology continues to advance, other challenges could still become apparent. The focal point at this time is to maintain a permanent market driven oversight and balance the benefits of AI with a sound judgment of the actual risks, and to be ready to adapt accordingly.

As Chairman of WITSA's AI Taskforce, I am very proud and pleased to have been working with the WITSA team in developing these sets of guidelines. When policymakers decide that regulation is necessary, then to avoid slowing AI innovation and adoption, we hope they can follow the policy principles suggested in this paper.

Robert Janssen

WITSA AI Taskforce Chairman

WITSA Deputy Chairman





WITSA GLOBAL POLICY ACTION COMMITTEE CHAIRMAN'S STATEMENT

As Chairman of WITSA's Global Policy Action Committee, I am excited to present our new report, "Building Trust and Delivering on the Promise of Artificial Intelligence".

As a global tech-driven organization, WITSA is well-positioned to contribute to the development of international public policy that supports a robust global tech infrastructure. WITSA aims to strengthen the industry at large by promoting an innovative, pro-competitive and fair trade environment; help governments, institutions and multilateral organizations understand future technology trends; and voice the concerns of the tech industry regarding policies that affect industry interests.

To facilitate these objectives, WITSA works closely with organizations including the World Trade Organization (WTO), the United Nations (UN), the Organization for Economic Cooperation and Development (OECD), the World Bank, Asia-Pacific Economic Cooperation (APEC), the International Telecommunications Union (ITU), the Internet Corporation for Assigned Names and Numbers (ICANN), and other international forums.

WITSA's key policy objectives are to foster an inclusive and innovative industry, as well as to increase competition through open markets and regulatory reform; protect intellectual property; and encourage cross-industry and government cooperation to enhance information security, bridge the education and skills gap, reduce tariff and non-tariff trade barriers to tech goods and services, and safeguard the viability and growth of the Internet and digital commerce.

Whether the promise of AI delivers on its potential depends on how well the AI ecosystem, governments, NGOs and other stakeholders manage perceived risks while fostering a regulatory environment that encourages innovation, best practices, consensus standards and international collaboration.

This paper will serve as an important benchmark for WITSA as it advances its comprehensive AI Action Plan and works with members to facilitate the transformative power of AI in driving economic growth, societal progress, and sustainable development.

Mr. Douglas Johnson
Chairman
WITSA Global Policy Action Committee

WITSA CEO'S STATEMENT

The world stands on the cusp of a technological revolution that not only promises hope and opportunity but also raises suspicion and dread. The dawn of the age of Artificial Intelligence is changing the world in more ways than one can imagine. AI has already impacted practically every human endeavor, from the way we work and communicate to the products and services we use every day. From healthcare and art to transportation and beyond, the reach of AI appears limitless. Ergo the intense global attention showered upon the latest renditions of generative AI.

Recognizing the fact that AI has the potential to transform everything, it is critically important to develop relevant policies that protect humanity without stifling innovation; strengthen enablers whilst providing safeguards; encourage transparency without sacrificing data privacy; promote workforce retraining whilst harnessing the AI multiplier; and foster international collaboration without smothering competition. It is up to us to ensure that we use this technology responsibly and ethically, to create a better world for everyone.

This paper recommends a careful and common-sense approach to harnessing the opportunities and addressing the risks arising from the rapid innovations in artificial intelligence, putting the spotlight on best practices, standards and regulation to help ensure that AI systems deployed are ethical, fair, transparent and accountable.

The paper emphasizes that innovation and trust in new technologies, including AI, are best supported when policy objectives and regulatory requirements make use of voluntary consensus-driven standardization to support implementation and compliance. To avoid slowing AI innovation and beneficial adoption, regulations should adhere to a specific list of policy principles identified in the paper. WITSA members are encouraged to engage their key stakeholders and policymakers in discussions to adopt these principles.

Dato' Dan E. Khoo
WITSA CEO



EXECUTIVE SUMMARY

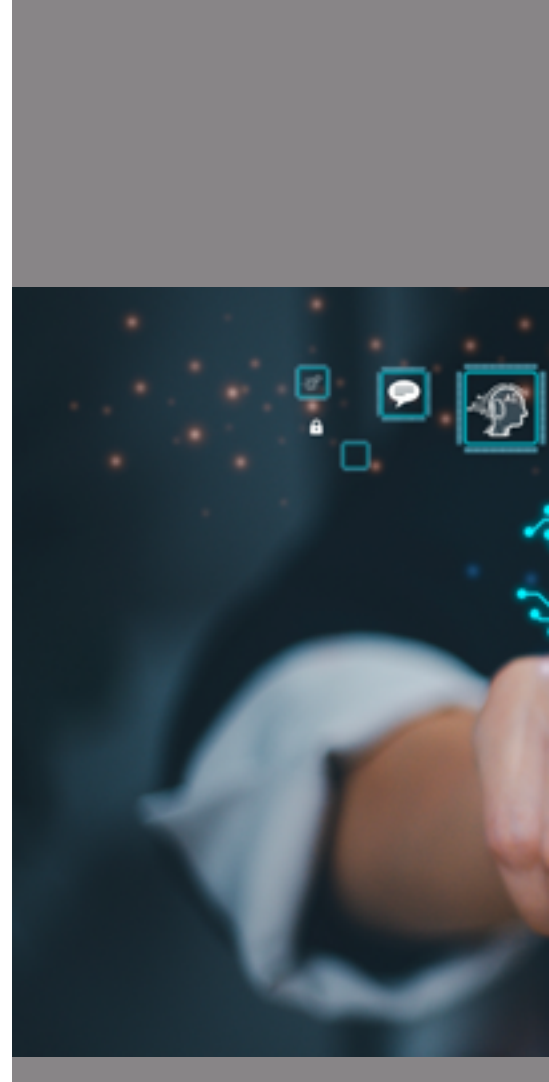
The time may have finally come for AI after periods of hype followed by several “AI winters” over the past 60 years. Today, it is impossible to read, listen to or watch any news outlet without seeing something about artificial intelligence (AI). In particular, new “large language models” (LLMs)—the sort that powers ChatGPT, the chatbot made by OpenAI, have surprised even their creators with their unexpected talents as they have been scaled up. Such “emergent” abilities include everything from solving logic puzzles and writing computer code to identifying films from plot summaries written in emoji. These models stand to transform humans’ relationship with computers, knowledge and even with themselves. AI has the potential to solve big problems such as developing new drugs, designing new materials to help fight climate change, or even untangling the complexities of fusion power.

Only a year ago, the mood in Silicon Valley was dour. Big Tech stocks were falling, the cryptocurrency bubble had popped¹, and a wave of layoffs was beginning to sweep through the industry. Since then, venture capitalists have been pouring money at AI start-ups, investing over \$11 billion in May alone, according to data firm PitchBook², an increase of 86 percent over the same month last year. The AI market is projected to reach \$407 billion by 2027³ with an annual growth rate of above 30 percent from 2023 to 2030. Organizations continue to find innovative ways to leverage AI for increased productivity and efficiency.

AI now powers so many real-world applications, ranging from facial recognition to language translators and assistants like Siri and Alexa, that we barely notice it. Along with these consumer applications, companies across sectors are increasingly harnessing AI’s power in their operations. Embracing AI promises considerable benefits for businesses and economies through its contributions to productivity growth and innovation.

At the same time, the world is struggling to figure out if and how to regulate these powerful tools. Like many things in life, the key is finding the right balance, and there are several trade-offs to keep in mind. The recent acceleration in both the power and visibility of AI systems, and growing awareness of their abilities and defects, have raised fears that the technology is now advancing so quickly that it cannot be safely controlled. Many academic and tech luminaries have been vocal about existential threats posted by AI.

To address those risks, WITSA believes that AI best practices are essential for AI developers, deployers and implementers as they help ensure that AI systems are fair, transparent, and accountable. Implementing AI is an essential step towards creating optimized operational efficiencies that increase longevity. AI can be used to improve business performance in areas including predictive maintenance, lets organizations handle tasks at a volume and velocity that’s simply not possible for humans to match, automates routine tasks, freeing up employees to focus on more creative and strategic work, and enables better decisions by providing accurate and timely data analysis. Ensuring AI systems are safe, reliable, and usable is crucial; industry-wide collaboration is vital to achieving this. This collaborative



1. <https://www.washingtonpost.com/business/2022/12/18/crypto-winter-ftx-collapse-bitcoin-prices/>

2. PitchBook: <https://pitchbook.com/news/articles/Amazon-Bedrock-generative-ai-q1-2023-vc-deals>

3. Global Newswire: <https://www.globenewswire.com/news-release/2022/08/19/2501600/0/en/Artificial-Intelligence-Market-Worth-407-0-Billion-By-2027-Exclusive-Report-by-MarketsandMarkets.html>



AI now powers so many real-world applications, ranging from facial recognition to language translators and assistants like Siri and Alexa, that we barely notice it.

Ensuring AI systems are safe, reliable, and usable is crucial; industry-wide collaboration is vital to achieving this. This collaborative approach provides a common language and framework for identifying risks and sharing solutions.

approach provides a common language and framework for identifying risks and sharing solutions.

As AI is a major opportunity for innovation and growth all around the world, many papers and recommendations have been issued by businesses, academia and organizations calling for standardization to raise trust and thereby promote the adoption of AI solutions. Indeed, trust in new technologies, such as AI, and innovation around these tools is best supported when policy objectives and regulatory requirements make use of voluntary consensus-driven standardization to support implementation. This way, compliance with policies and regulatory requirements as well as interoperability between different implementations can be achieved without limiting the potential for innovation by mandating specific technology choices.

In recent times, policymakers have proposed a variety of regulations to address concerns that this coming wave of AI systems may cause harm. Minimizing potential harm from AI systems is an important goal, but so too is maximizing the potential benefits of AI systems. Implementing many of these proposals, especially in their current form, is likely to have serious consequences because many of AI's potential benefits—including opportunities to use the technology both to save lives and to improve living standards—may be delayed or denied with poorly crafted regulations.



Policymakers should also take care to ensure that the benefits of regulation outweigh costs and harm to innovation, and work with stakeholders to optimize regulations and boost the uptake of trustworthy AI.

When policymakers decide that regulation is necessary, then to avoid slowing AI innovation and adoption, they should follow the policy principles identified in this paper. Governments should avoid pro-human biases – allowing businesses to use AI systems performing tasks traditionally fulfilled by humans. They should be regulating the performance of AI systems broadly and avoid prescriptive rules that address specific processes and methods that businesses must comply with.

Furthermore, policymakers should only regulate sectors and not technologies or models – making narrow rules for specific AI applications in distinct sectors, such as education, transportation or health care. They should avoid AI myopia as focusing too narrowly on AI as the culprit of a perceived problem (e.g., in hiring practices) diverts attention away from the opportunities that AI may provide to mitigate social harms. AI should be defined narrowly and precisely to reduce the risk of setting policies that affect other software and systems in unintended ways. Finally, governments should enforce existing rules capable of tackling harms that may arise from AI deployment before considering a need for AI specific legislation.

Policymakers should also take care to ensure that the benefits of regulation outweigh costs and harm to innovation, and work with stakeholders to optimize regulations and boost the uptake of trustworthy AI. Regulators should apply non-discrimination principles

to ensure that rules are not impacting businesses differently based on their size or based on their location and seek expert advice.

Policymakers should develop regulatory sandboxes for AI as an essential tool to address AI harms without compromising on innovation, saving policymakers considerable time and resources and aiding businesses by reducing the time and capital required to enter the market. The controlled environment of the sandbox approach should be a first immediate and mandatory step for countries to test and experiment with new technologies, business models, and regulatory approaches.

Furthermore, governments should lead by example by ensuring in-house use of responsible AI. AI regulations should take a risk-based approach, be tailored to different AI applications and services, always seek to boost the overall digital transformation as well as protecting the foundation of AI systems, such as *source code*, *proprietary algorithms*, and other *intellectual property*, and consider ways to support AI research and development.

Finally, governments should encourage reliance on international consensus standards and seek to align domestic requirements with international approaches to AI regulation. ■



ABOUT WITSA

WITSA is a global consortium of leading digital tech industry association members from over 80 countries/economies.

As the leading recognized voice of the global digital tech industry, WITSA aims to drive transformation and expand the use of tech globally; given that tech is the key driver of the global economy.

WITSA's members and stakeholders comprise national associations, multinational corporations, institutions and organizations, researchers, developers, manufacturers, software developers, telecommunication companies, suppliers, trainers and integrators of digital tech goods and services. As such, they represent a large and obviously vital constituent group for whom the effective balancing of concerns and rights affecting the security, privacy and information capability provided by digital tech products and services underpins business development and economic activity.

WITSA is a founding partner of the Digital Trade Network (DTN), a new initiative providing a permanent private sector resource for digital trade policy makers in Geneva. Through DTN, WITSA works with several other partner organizations to build an impartial, broad base of international supporters to work with the WTO, the UN Conference for Trade and Development (UNCTAD), the International Telecommunications Union (ITU), PeaceTech Lab, and related economic policy agencies in Geneva with a focus on the networked economy. ■



Policymakers should also take care to ensure that the benefits of regulation outweigh costs and harm to innovation, and work with stakeholders to optimize regulations and boost the uptake of trustworthy AI.



AI POLICY PRINCIPLES FOR GOVERNMENTS

As AI continues to improve, opportunities to use the technology to increase productivity and quality of life will flourish across many sectors of the economy, including health care, education, transportation, and more. In response, policymakers have proposed a variety of regulations to address concerns that this coming wave of AI systems may cause harm. Minimizing potential harm from AI systems is an important goal, but so too is maximizing the potential benefits of AI systems. Implementing many of these proposals, especially in their current form, is likely to have serious consequences because many of AI's potential benefits—including opportunities to use the technology both to save lives and to improve living standards—may be delayed or denied with poorly crafted regulations.

Fears about AI have reached new levels because of the emergence of generative AI. Generative AI—a novel tool that can produce complex text, images, and videos from simple inputs—promises to democratize the creative sector and enable entirely new forms of creativity. This novelty has impressed technology enthusiasts but alarmed many others—especially those who believe AI is encroaching on creativity, which many people believe to be an essential difference that separates humans from machines.

Yet, technology and human creativity have long been intertwined, and fears about the negative impact of new innovations have been overstated in the past. For example, prior innovations in the music sector led to fears that record albums would make live shows redundant or that radio would destroy the record industry or that sampling and other means of digital editing would undermine musical artistry. But these concerns never arrived. Over time, this and other tech panics fizzled out as the public embraced the new technology, markets adapted, and initial concerns turned out to be clearly overblown or never arrived.

The fears around new technologies follow a predictable trajectory called “the Tech Panic Cycle⁴.” Fears increase, peak, then decline over time as the public becomes familiar with the technology and its benefits. Indeed, other previous “generative” technologies in the creative sector such as the printing press, the phonograph, and the Cinématographe followed this same course. But unlike today, policymakers were unlikely to do much to

4. <https://datainnovation.org/2023/05/tech-panics-generative-ai-and-regulatory-caution/>

Another such area is occupational licensing, such as in the legal and medical fields. Whenever AI-enabled services can perform as well or better than a licensed professional at a specific task, those systems should be allowed.

regulate and restrict these technologies. As the panic over new AI innovations, such as generative AI, enters its most volatile stage, policymakers should recognize the predictable cycle we are in, and proceed cautiously regarding any regulatory efforts so as not to harm AI innovation.

If policymakers decide that regulation is necessary, then to avoid slowing AI innovation and adoption, they should follow the following policy principles:

Governments should avoid pro-human biases.

They should not discriminate against AI and permit AI systems to do what is legal for humans and prohibit what is illegal in the human realm. Businesses should be allowed to use AI systems performing tasks traditionally fulfilled by a human, e.g., a security guard or AI system verifying the identity of someone entering an office building. Another such area is occupational licensing, such as in the legal and medical fields. Whenever AI-enabled services can perform as well or better than a licensed professional at a specific task, those systems should be allowed.

This principle also applies to copyright and intellectual property rights. AI, such as generative AI systems should be held to the same high standards and laws pertaining to IP, e.g., while copyright law typically provides copyright owners certain exclusive rights, these rights are subject to exceptions and limitations, including those under fair use doctrine. Online piracy is clearly theft, as it is in the physical world, but seeking inspiration and learning from others is not, whether by humans or AI systems.



Similarly, while most legal systems confer certain rights to copyright holders, such as whether to display or perform their works publicly, both people and AI systems are free to observe these works, such as paintings in a museum, and use what they learn to create future content without the explicit permission of the copyright owners. Just as budding musicians and amateur painters do not pay copyright owners to study their techniques, styles or artistry, generative systems should be permitted to do the same.

AI systems should be permitted, as artists are, to create images and other content in the style of other artists because copyright does not give someone an exclusive right to a style⁵. Just like humans, generative AI systems do not produce remixes of existing content. They use massive amounts of training data to create new realistic content based on specific prompts based on statistical patterns and are not merely searching through existing data to find the closest match. Similarly, people who use AI to create content should be able to secure copyright

5. Greg Kanaan, "You Can't Copyright Style," The Legal Artist, February 1, 2016, <https://www.thelegalartist.com/blog/you-cant-copyright-style>.



“AI systems with human-competitive intelligence can pose profound risks to society and humanity [and that] powerful AI systems should be developed only once we are confident that their effects will be positive, and their risks will be manageable.”

protection for their works based on their substantial human input in the same way e.g., a photographer uses a camera to take a photograph.

Governments should regulate performance, not technical processes.

While addressing important concerns related to AI’s impact on safety, bias and efficacy, governments should focus on regulating the performance of AI systems broadly and avoid prescriptive rules that address specific processes and methods that businesses must comply with. Governments, businesses and consumers are better positioned to compare the performance of different systems and set minimum performance requirements when performance-based metrics are deployed.

For example, rather than requiring AI systems operators to use specific or diverse datasets to train their models, governments should establish performance-based regulations, such as, for example, requiring lenders to show that their credit scoring models assess risk fairly across all protected classes of customers. By enabling businesses to identify the best methods for complying with government set performance metrics, it grants them the flexibility necessary to comply the most efficiently, thereby optimizing innovation, product output and customer value.

Regulate sectors, not technologies.

Concerns about the rapid innovations in machine learning have been mounting, leading some

to claim that AI can pose profound risks to society and humanity. When OpenAI released ChatGPT to the public in late 2022, it took only two months to reach 100 million users, making it the fastest-growing consumer application in history. In late March 2023, a document was released online by the Future of Life Institute titled “Pause Giant AI Experiments: An Open Letter⁶”, which stated that “AI systems with human-competitive intelligence can pose profound risks to society and humanity [and that] powerful AI systems should be developed only once we are confident that their effects will be positive, and their risks will be manageable.”

In response, and since policymakers cannot foresee all future applications of AI, calls have been increasing to regulate such systems and their underlying metrics rather than specific uses. However, AI systems may be general-purpose technology that could be applied to many different applications. The risks and benefits of such systems depend entirely on how they are being used.

Therefore, when or if a need for regulation is warranted, governments should make narrow rules for specific AI applications in distinct sectors, such as education, transportation or health care. As such, regulators should differentiate between use cases, such as pertaining to medical diagnosis, autonomous vehicle navigation or stock trading, even when the underlying technology is similar. Moreover, in many cases existing legislation pertaining to these sectors, while not AI-specific, already enables authorities to conduct appropriate oversight of AI-enabled systems and their implementations.

6. Future of Life Institute: Pause Giant AI Experiments: An Open Letter (<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>)

Avoid AI myopia.

Many of the perceived societal harms from AI, such as negative impact on consumer privacy, bias in hiring practices, college admissions, credit scoring and the lack of redress when unfavorable outcomes occur, are concerns which transcend AI and have been manifest all through society for a very long time. The objective of policymakers should be to address and fix the broader problems, not just those aspects involving AI. Focusing too narrowly on AI as the culprit of a problem (e.g., in hiring practices) diverts attention away from the opportunities that AI may provide to mitigate discrimination, for example by bringing about more objectivity into hiring practices and reduce human bias.

As with any technology, AI can be used for good, or it can be abused by those applying it either through malevolence or incompetence. Fixating on AI while ignoring the bigger picture can serve to sideline innovators who work on addressing broader societal problems, such as facial recognition technology vis-à-vis broader sets of police reform that can decrease wrongful arrests.

Define AI precisely and narrowly.

Government authorities and legislators should take great care when defining AI to reduce the

risk of setting policies that affect other software and systems in unintended ways. An essential factor is to properly identify the component parts of AI systems beyond algorithms (such as datasets and computing power), as well as to define related key terms such as machine learning. Some algorithms have been applied for decades but do not constitute “artificial intelligence” or “machine learning” systems. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context.

Policymakers should not use expansive definitions of AI in rulemaking if the intended purpose is to manage deep learning or uninterpretable machine learning systems more narrowly. For example, there is a difference between the latest wave of AI systems that learn from data and experience, and traditional software and control systems that operate according to predictable rules, which have long been embedded in a wide variety of high-risk systems, from flight control to pacemakers.

One pertinent example is the EU’s AI Act, which in an early iteration featured a broad enough definition of AI that even simple spreadsheet software such as Google Sheets and Microsoft Excel likely would have been covered⁷. Such unintended errors incur significant impediments and costs on those developers and

Some algorithms have been applied for decades but do not constitute “artificial intelligence” or “machine learning” systems. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context.

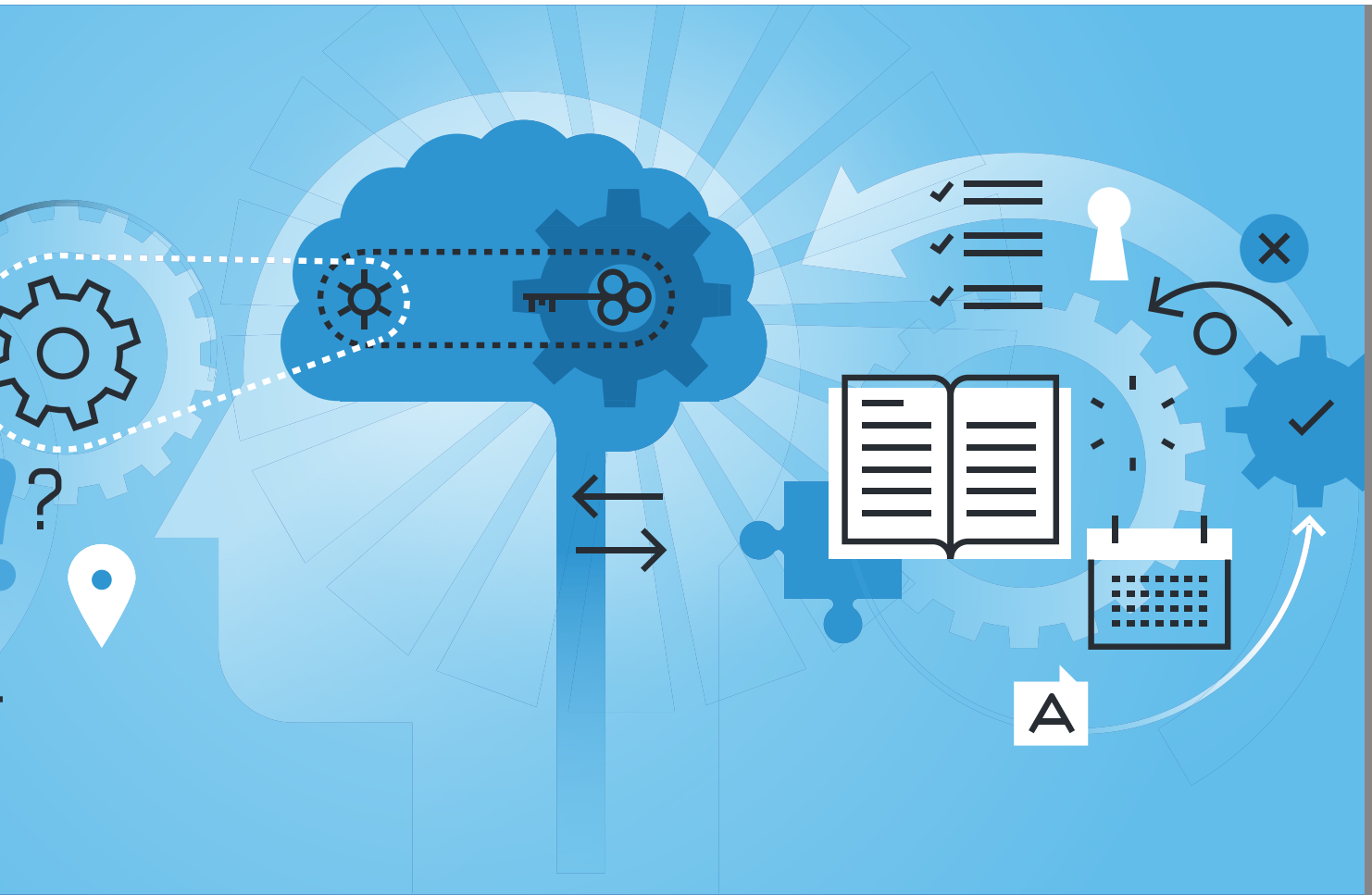


deployers of basic integration or automation products and services.

Recognize and enforce existing rules first.

It is important to note that many laws and regulations around the world addressing important societal issues such as discrimination, worker safety and product liability, already exist on a broader scale beyond the limitations of AI. In these cases, regulators are already equipped to tackle harms that may arise from AI deployment, often without a need for AI specific legislation. For example, businesses

7. Mikolaj Barczentewicz and Benjamin Mueller, “More than Meets the AI: The Hidden Costs of a European Software Law,” Center for Data Innovation, December 1, 2022, <https://www2.datainnovation.org/2021-more-than-meets-the-ai.pdf>; Insurance Europe, “Response to EC proposal for a Regulation on AI,”



In these cases, regulators are already equipped to tackle harms that may arise from AI deployment, often without a need for AI specific legislation.

are not allowed to discriminate in employment decisions regardless of whether an AI system is used. This was pointed out by the White House Office of Science and Technology Policy (OSTP) in connection with its October 4, 2022, Blueprint for an AI Bill of Rights⁸:

- *Some of these protections are already required by the U.S. Constitution or implemented under existing U.S. laws. For example, government surveillance and data search and seizure are subject to legal requirements and judicial oversight. There are Constitutional requirements for human review of criminal investigative matters and statutory requirements for judicial review. Civil rights laws protect the American people against discrimination. There are regulatory safety requirements for medical devices, as well as sector-, population-, or technology-specific privacy and security protection⁹.*

Regulators should consult with businesses and other stakeholders about how these existing regulations should be applied to emerging AI products and services.

8. <https://www.whitehouse.gov/ostp/news-updates/2022/10/04/fact-sheet-biden-harris-administration-announces-key-actions-to-advance-tech-accountability-and-protect-the-rights-of-the-american-public/>

9. White House OSTP: "Relationship to Existing Law and Policy" <https://www.whitehouse.gov/ostp/ai-bill-of-rights/relationship-to-existing-law-and-policy/>

The solution instead should be to reduce the overall challenges in compliance so that they become manageable for businesses regardless of size.

Ensure benefits of regulation outweigh costs and harm to innovation.

This is a consideration that applies to all kinds of governance. Policymakers should consider the impact of regulatory intervention on compliance costs, indirect productivity, innovation, as well as on competitiveness costs (such as impediments to the adoption of cost-effective emerging technologies that provide social and economic benefits). Unfortunately, regulators may pay insufficient attention to the costs of tech regulations and may also overestimate the stated positive impacts such rules have on innovation. Costs also frequently become a secondary concern when addressing perceived “fundamental rights” (e.g., privacy, security, non-discrimination). Regulators should consider that AI systems can just as easily address societal harms, and that the net impact of a regulation – especially when poorly drafted – can be negative.

Governments must work with stakeholders to optimize regulations and boost the uptake of trustworthy AI.

To create a competitive ecosystem for AI, stakeholders, such as industry, civil society, academia and governments need to work together in the shaping of national, regional and international rules and standards. Even when a regulation is deemed necessary, rule makers should explore all options to maximize benefits and reduce costs. Finding the most efficient ways to achieve regulatory objectives is imperative as overly burdensome and unnecessary rules will reduce innovation, adoption and force businesses to divert resources away from productive business activities towards compliance at the expense of the end user. Once enacted, policymakers must make regular impact assessments and amend and adjust the regulatory framework to ensure the cost-benefit implications remain as intended over time.

Non-discrimination: Treat firms equally.

AI policies should not be applied to businesses differently based on their size or based on their location. Policymakers may be tempted to exempt smaller entities from new laws as a way to reduce compliance burdens on such companies with fewer resources, and as a way to more easily reach legislative consensus on new rules. The solution instead should be to reduce the overall challenges in compliance so that they become manageable for businesses regardless of size. If the objective of legislation is to protect consumers from AI-related products and services, then all businesses should comply with

the new rules. Likewise, ensuring consumers are appropriately protected from perceived harms, businesses should be held accountable regardless of in which countries they are domiciled.

Seek expertise.

As AI technologies continue to advance at a rapid pace, having the right technical talent in place within as well as in consultation with government is crucial to ensure countries are prepared to address the potential risk and harms that AI systems can present. Governments and the private sector therefore urgently need to work together to advance tech innovation, enhance security and address other potential societal harms. Tech and industry expertise are essential in drafting effective legislation.

Given the complexity and rapid speed of AI innovation, regulators also often lack the resources necessary to effectively supervise and mitigate the risks emerging from the AI systems which may affect safety and fundamental rights. Throughout the regulatory process, policymakers should therefore consult with AI experts in the sectors that they seek to regulate.

Getting technical talent into the public workforce is the single biggest obstacle for effective regulation. Government cannot govern AI if it does not understand AI. Regulators should therefore ensure that they employ experts with the necessary AI and data literacy skills to fully understand the new and emerging AI technologies they are overseeing. This can include creating training programs for supervisors and management officials, form public-

private partnerships and academic-agency partnerships to attract AI talent to public service and build cross-functional teams

Adopt regulatory sandboxes.

Regulators should also consider a more innovative approach to regulation. Regulatory sandboxes for AI (AI sandbox) is an essential tool to regulate AI without compromising on innovation, and setting up regulatory sandboxes are currently viewed as the best solution for handling AI. A regulatory sandbox is a controlled environment that allows innovators and businesses to test and develop new AI technologies in an environment with reduced regulatory constraints. The idea behind the sandbox is to provide a safe space for businesses and regulators to work together to understand how new technologies can be developed and regulated in a responsible and ethical way.

AI sandboxes promise several advantages, such as promoting innovation by allowing for the development of new AI technologies in a controlled environment to reduce the risk by allowing temporary exemptions from specific rules and inviting detailed feedback on their operations to regulators, which then informs new permanent rules that take into account successful business models. This has proven to reduce the so-called 'time to market' for innovations, giving new businesses increased legal certainty and thereby leading to more innovation. AI sandboxes can therefore be a critical bridge between AI policy and regulation and large-scale deployment of AI solutions as they make it possible to test new technologies under a regulator's supervision and



A regulatory sandbox is a controlled environment that allows innovators and businesses to test and develop new AI technologies in an environment with reduced regulatory constraints.



To be successful in setting up AI sandboxes, governments should develop a national policy or strategy on AI, including review, and update where necessary, the existing consumer protection and data privacy law and frameworks.

frameworks, evaluating feasibility, demand, potential outcomes, and collateral effects. Inadequate specifications can be harmful to competition, consumers, data protection, and regulation.

contribute to evidence-based policymaking.

Regulating AI is particularly challenging for developing and underdeveloped countries due to limited capacity, resources, and exposure to the development and actual implementation of AI solutions in their jurisdiction. Regulatory sandboxes for AI therefore can be a valuable policy tool that helps inform and enable effective regulation of AI in emerging markets. They will help both developed and emerging markets enhance the capacity to craft

effective policy and regulatory frameworks on AI, will strengthen the national AI ecosystem and develop a reliable pipeline of impactful and safe AI solutions.

To be successful in setting up AI sandboxes, governments should develop a national policy or strategy on AI, including review, and update where necessary, the existing consumer protection and data privacy law and frameworks. AI sandboxes should adhere to thorough designs and testing with robust methodological and assessment

Regulatory, economic, and technical assessments from the use of AI sandboxes can inform decisions about whether to change or re-interpret legal statutes and can save government agencies considerable time, such as when they become aware that existing legal structures can address new technologies adequately. At the same time, businesses entering a sandbox benefit from a license exemption or other waiver, or from specific regulatory provisions, reducing the time and capital required to enter the market. As such, regulatory sandboxes can foster investment in companies participating in testbeds.



Adopt in-government use of responsible AI.

Governments should maximize their procurement and deployment of relevant AI solutions to help agencies deliver their mission and optimize constituent services. However, governments face numerous barriers—including lack of specialized talent, limited investments in AI research and innovation and often-unclear regulations designed to ensure that AI is applied in an ethical, secured, transparent, and human-centric manner across all sectors—that could prevent them¹⁰ from adopting AI use cases and capturing the value of AI. Government agencies should therefore consider hiring and resource a chief AI

officer, direct the creation of an interagency Chief AI Officers Council, which would be responsible for reviewing the agency AI use case inventories and identify dozens of key agency processes that could be transformed with AI in a manner consistent with privacy, civil rights, and civil liberties.

Regulations should take a risk-based approach.

AI regulation should be tailored to the type of application and service. Different use cases of AI have different impacts and risk profiles, and consequently require different methods to ensure risk mitigation, transparency, accountability and fairness.

Governments should encourage reliance on international consensus standards and seek to align regulatory requirements with international approaches to AI regulation (see *standards chapter, below*). Moreover, countries should seek to develop common principles to design and implement AI sandboxes in their respective jurisdictions (e.g., the EU's proposed 'Artificial Intelligence Act' calls for¹¹ the establishment of common rules to implement AI sandboxes in EU member countries).

Governments should seek to boost the overall digital transformation.

Effective measures to improve and make the most out of AI can be taken today, by increasing investment in research and innovation, strengthening digital education and skills training as well as encouraging data access and infrastructure development.

Funding for AI research and development

Governments should encourage multi-stakeholder partnerships and lab-to-market initiatives, such as centers for excellence, innovation hubs, or research centers that underpin industry's crucial role in developing and deploying AI solutions.

Protecting the foundation of AI systems

Policymakers should support provisions within legislation that protect the foundation of AI systems, including source code, proprietary algorithms, and other intellectual property. Governments should avoid requirements that force companies to transfer or provide access to technology, source code, algorithms, or encryption keys as conditions for doing business with the federal government or as a general practice for business-to-business operations. ■

10. <https://www.mckinsey.com/industries/public-sector/our-insights/the-potential-value-of-ai-and-how-governments-could-look-to-capture-it>

11. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI\(2022\)733544_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf)

RESPONSIBLE BEST PRACTICES IN AI

AI best practices are important for AI developers because they help ensure that AI systems are fair, transparent, and accountable. Implementing AI is an essential step towards creating optimized operational efficiencies that increase longevity. AI can be used to improve business performance in areas including predictive maintenance.

AI developers must move quickly to develop and deploy systems that address algorithmic bias. The need for diverse representation in data sets and user research is essential to ensure fair and unbiased AI systems. Best practices should also provide guidelines on how to make AI systems transparent, understandable, and accountable while protecting individual privacy. This necessitates the need for cross-sector collaboration so that the tech industry can develop robust and safe AI systems that benefit everyone.

One of the fundamental questions in AI ethics is ensuring that AI systems are developed and deployed without reinforcing existing social biases or creating new ones. To achieve this, industry must consider the data sets being used and ensure they represent the broadest possible set of voices. Inclusivity in the development process and identifying potential harms through user research is also essential.

Social bias can be infused into AI systems through the data sets used to train them. Unrepresentative data sets containing biases, such as image data sets with predominantly one race or lacking cultural differentiation, can result in biased AI systems. Furthermore, applying AI systems unevenly in society can perpetuate existing stereotypes. Today's AI boom will amplify social problems if industry doesn't act early in product development processes. To make AI systems transparent and understandable to the average person, prioritizing explainability during the development process is key. Techniques such as "chain of thought prompts" can help AI systems show their work and make their decision-making process more understandable. User research is also vital to ensure that explanations are clear and users can identify uncertainties in AI-generated content.

Protecting consumers' privacy and ensuring responsible AI use require transparency and consent. Tech industry should follow guidelines for responsible generative AI, which include respecting data provenance and only using customer data with consent. Allowing users to opt in, opt-out, or have control over their data use is critical for privacy. As the competition for innovation in generative AI intensifies, maintaining human control and autonomy over increasingly autonomous AI systems is more important than ever. Empowering users to make informed decisions about the use of AI-





Protecting consumers' privacy and ensuring responsible AI use require transparency and consent. Tech industry should follow guidelines for responsible generative AI, which include respecting data provenance and only using customer data with consent.



Investing in safeguards and focusing on the here and now, rather than solely on potential future harms, can help mitigate these issues and ensure the responsible development and use of AI systems.

generated content and keeping a human in the loop can help maintain control.

Ensuring AI systems are safe, reliable, and usable is crucial; industry-wide collaboration is vital to achieving this. This collaborative approach provides a common language and framework for identifying risks and sharing solutions. Failing to address these ethical AI issues can have severe consequences, as seen in cases of wrongful arrests due to facial recognition errors or the generation of harmful images. Investing in safeguards and focusing on the here and now, rather than solely on potential future harms, can help mitigate these issues and ensure the responsible development and use of AI systems.

WITSA endorses the OECD AI Principles¹², which promote use of AI that is innovative and trustworthy and that respects human rights and democratic values. Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time:

Inclusive growth, sustainable development and well-being

Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.


Human-centered values and fairness

AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of the art.

Transparency and explainability

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of the art:

12. <https://oecd.ai/en/ai-principles>



In 2021, 193 countries adopted a Recommendation on the Ethics of Artificial Intelligence¹³, created by the United Nations Educational, Scientific, and Cultural Organization (UNESCO). The recommendations focus on data protection, mass surveillance, and resource efficiency (but not computer superintelligence).

- to foster a general understanding of AI systems,
- to make stakeholders aware of their interactions with AI systems, including in the workplace,
- to enable those affected by an AI system to understand the outcome, and,
- to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

Robustness, security and safety

- AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.
- To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of the art.
- AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

Accountability

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of the art.

WITSA also supports the UNESCO Recommendation on the ethics of artificial intelligence.

In 2021, 193 countries adopted a Recommendation on the Ethics of Artificial Intelligence¹³, created by the United Nations Educational, Scientific, and Cultural Organization (UNESCO). The recommendations focus on data protection, mass surveillance, and resource efficiency (but not computer superintelligence).

1. **Human rights and human dignity:** Respect, protection and promotion of human rights and fundamental freedoms and human dignity
 - **Proportionality and Do No Harm:** The use of AI systems must not go beyond what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses.
 - **Safety and Security:** Unwanted harms (safety risks) as well as vulnerabilities to attack (security risks) should be avoided and addressed by AI actors.

13. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>

- **Right to Privacy and Data Protection:** Privacy must be protected and promoted throughout the AI lifecycle. Adequate data protection frameworks should also be established.
 - **Multi-stakeholder and Adaptive Governance & Collaboration:** International law & national sovereignty must be respected in the use of data. Additionally, participation of diverse stakeholders is necessary for inclusive approaches to AI governance.
 - **Responsibility and Accountability:** AI systems should be auditable and traceable. There should be oversight, impact assessment, audit and due diligence mechanisms in place to avoid conflicts with human rights norms and threats to environmental wellbeing.
 - **Transparency and Explainability:** The ethical deployment of AI systems depends on their transparency & explainability (T&E). The level of T&E should be appropriate to the context, as there may be tensions between T&E and other principles such as privacy, safety and security.
 - **Human Oversight and Determination:** Countries should ensure that AI systems do not displace ultimate human responsibility and accountability.
 - **Sustainability:** AI technologies should be assessed against their impacts on 'sustainability', understood as a set of constantly evolving goals including those set out in the UN's Sustainable Development Goals.
 - **Awareness & Literacy:** Public understanding of AI and data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy.
 - **Fairness and Non-Discrimination:** AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all.
2. **Living in peaceful, just and interconnected societies**
- AI actors should play a participative and enabling role to ensure peaceful and just societies, which is



based on an interconnected future for the benefit of all, consistent with human rights and fundamental freedoms. The value of living in peaceful and just societies points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.

3. Ensuring diversity and inclusiveness

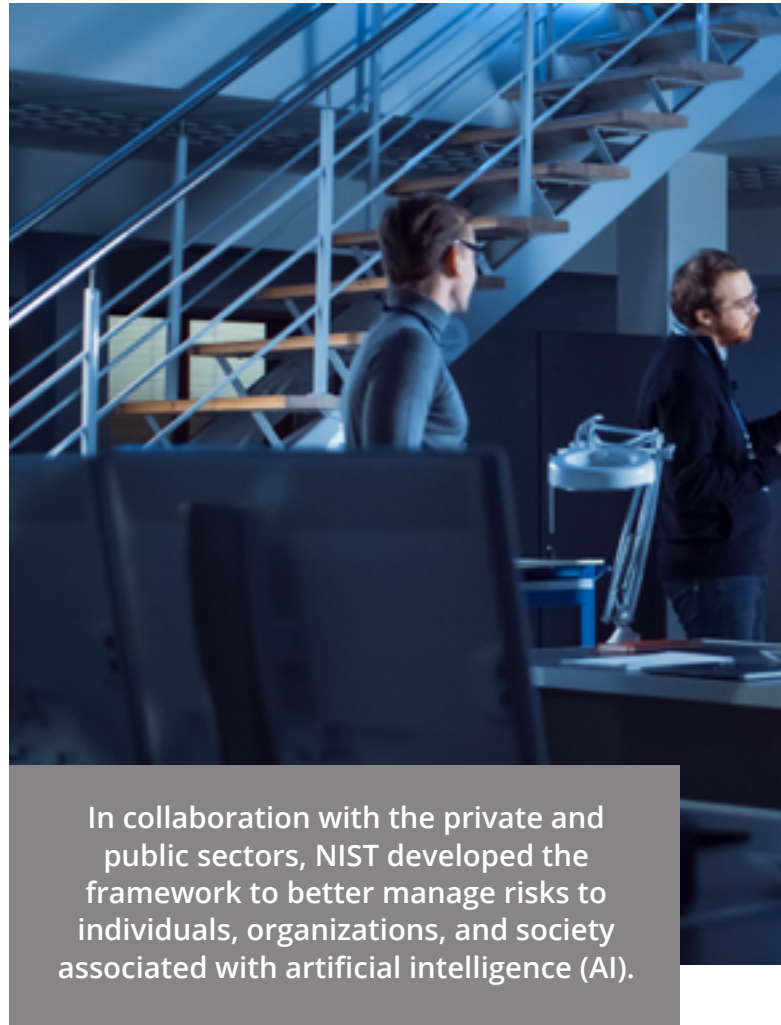
Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, consistent with international law, including human rights law. This may be done by promoting active participation of all individuals or groups regardless of race, color, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds. The scope of lifestyle choices, beliefs, opinions, expressions or personal experiences, including the optional use of AI systems and the co-design of these architectures should not be restricted during any phase of the life cycle of AI systems. Furthermore, efforts, including international cooperation, should be made to overcome, and never take advantage of, the lack of necessary technological infrastructure,

synthetic media production; Support additional research to shape future data-sharing initiatives and determine what types of data would be most appropriate and beneficial to collect and report, while balancing considerations such as transparency and privacy preservation; Take steps to research, develop, and deploy technologies that are as forensically detectable as possible for manipulation, without stifling innovation in photorealism; and that retain durable disclosure of synthesis, such as watermarks or cryptographically bound provenance that are discoverable, preserve privacy, and are made readily available to the broader community and provided open source; and Provide a published, accessible policy outlining the ethical use of your technologies and use restrictions that users will be expected to adhere to and providers seek to enforce

- **Practices for Creators:** Be transparent to content consumers about: How you received informed consent from the subject(s) of a piece of manipulated content, appropriate to product and context, except for when used toward reasonable artistic, satirical, or expressive ends; how you think about the ethical use of technology and use restrictions (e.g., through a published, accessible policy, on your website, or in posts about your work) and consult these guidelines before creating synthetic media; the capabilities, limitations, and potential risks of synthetic content. Disclose when the media you have created or introduced includes synthetic elements especially when failure to know about synthesis changes the way the content is perceived. Take advantage of any disclosure tools provided by those building technology and infrastructure for synthetic media.
- **Practices for Distributors and Publishers:** Disclose when you confidently detect third-party/user-generated synthetic content. Provide a published, accessible policy outlining the organization's approach to synthetic media that you will adhere to and seek to enforce.

NIST Artificial Intelligence Risk Management Framework

WITSA furthermore recognized the importance work conducted by the National Institute of Standards and



In collaboration with the private and public sectors, NIST developed the framework to better manage risks to individuals, organizations, and society associated with artificial intelligence (AI).

Technology (NIST), with the January 26, 2023 publication of AI Risk Management Framework (AI RMF 1.0)¹⁶.

In collaboration with the private and public sectors, NIST developed the framework to better manage risks to individuals, organizations, and society associated with artificial intelligence (AI). The NIST AI Risk Management Framework (AI RMF) is intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. The Framework was developed through a consensus-driven, open, transparent, and collaborative process. It is designed to equip organizations and individuals with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, deployment, and use of AI systems over time.

16. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>



For AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks.

For AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks. This Framework articulates the following characteristics of trustworthy AI and offers guidance for addressing them. Characteristics of trustworthy AI systems include: valid and reliable; safe, secure and resilient; accountable and transparent; explainable and interpretable; privacy-enhancing, and fair with harmful bias managed. Creating trustworthy AI requires balancing each of these characteristics based on the AI system’s context of use. While all characteristics are socio-technical system attributes, accountability and transparency also relate to the processes and activities internal to an AI system and its external setting. Neglecting these characteristics can increase the probability and magnitude of negative consequences. ■



Safe



Secure & Resilient



Explainable & Interpretable



Privacy-Enhanced



Fair - With Harmful Bias Managed



Accountable & Transparent



Valid & Reliable



Indeed, trust in new technologies, such as AI, and innovation around these technologies are best supported when policy objectives and regulatory requirements make use of voluntary industry-driven standardization to support implementation.



FOSTERING TRUSTWORTHY AI THROUGH CONSENSUS STANDARDS AND INTERNATIONAL COOPERATION

Even more than many domains of science and engineering in the 21st century, the international AI landscape is deeply collaborative, especially when it comes to research, innovation, and standardization. As countries move from developing frameworks and policies to more concrete efforts to regulate AI, demand for AI standards will grow¹⁷. These include standards for risk management, data governance, and technical documentation that can establish compliance with emerging legal requirements. International AI standards will also be needed to develop commonly accepted labeling practices that can facilitate business-to-business (B2B) contracting and to demonstrate conformity with AI regulations; address the ethics of AI systems (transparency, neutrality/lack of bias, etc.); and maximize the harmonization and interoperability for AI systems globally.

As AI is a major opportunity for innovation and growth all around the world, many papers and recommendations have been issued by businesses, academia and organizations calling for standardization¹⁸ to raise trust and thereby promote the adoption of AI technologies and solutions. Indeed, trust in new technologies, such as AI, and innovation around these technologies are best supported when policy objectives and regulatory requirements make use of voluntary industry-driven standardization to support implementation. This way, compliance

17. <https://www.brookings.edu/articles/strengthening-international-cooperation-on-ai/>

18. <https://www.digitaleurope.org/resources/digitaleurope-recommendations-on-standardisation-in-the-field-of-artificial-intelligence/>

with policies and regulatory requirements as well as interoperability between different implementations can be achieved without limiting the potential for innovation by mandating specific technology choices.

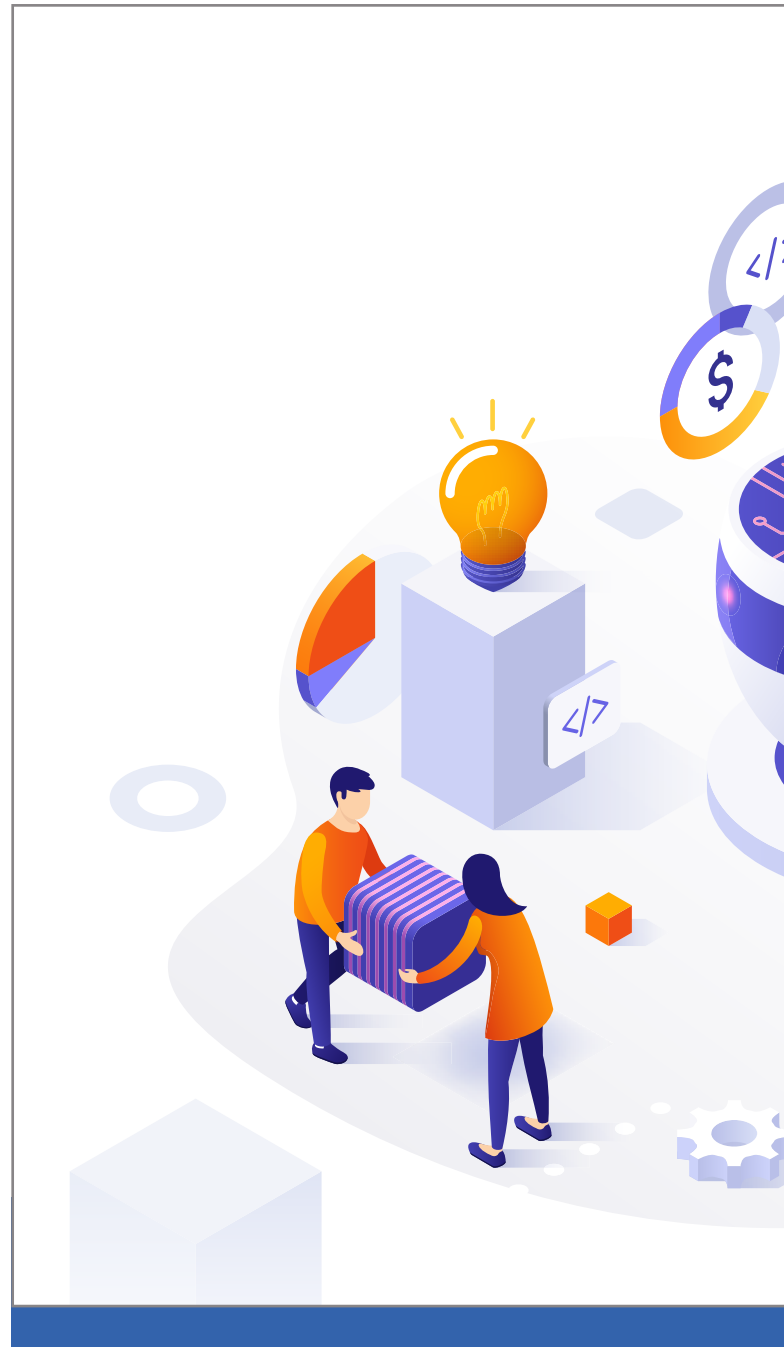
At the same time, we caution against inappropriate use of standardization. AI is global. While nations and regions may compete for AI innovation power, their industries need a global scale to effectively compete and develop their activities. That is best achieved through global standards rather than regional and national ones. For the same reason, standards are not an appropriate tool for codifying cultural norms or values: AI standards can and should support adherence to ethical principles, but they need to be adaptable to a wide variety of value systems.

Sound AI standards can support international trade and investment in AI, expanding AI opportunity globally and increasing returns to investment in AI R&D. However, AI standards should not be used in a way that establishes barriers to trade (TBT). Instead, governments should encourage engagement in global WTO TBT-compliant standardization initiatives (though, the WTO TBT Agreement's¹⁹ relevance to AI standards is limited by its application only to goods, whereas many AI standards will apply to services). Recent trade agreements have started to address AI issues, including support for AI standards, but more is needed. An effective international AI standards development process is also needed to avoid bifurcated AI standards.

The immediate challenges that standardization should address are:

- Establishing consensus around AI foundational concepts, management and governance practices.
- Framing concepts and best practices to establish trustworthiness of AI, including in areas such as privacy, cybersecurity, safety, reliability, and transparency.

Global standardization efforts in these areas are ongoing, for instance within ISO/IEC JTC 1²⁰ Information Technology and the IEEE Standards Association. Such efforts can help ensure that global AI systems are ethically sound, robust, and trustworthy, that opportunities from AI are widely distributed, and that standards are technically sound and research-driven regardless of sector or application. The EU and the U.S., through the U.S.-E.U. Trade and Technology Council²¹, are also



An effective international AI standards development process is also needed to avoid bifurcated AI standards.

19. https://www.wto.org/english/tratop_e/tbt_e/tbt_e.htm

20. <https://jtc1.info.org/>

21. <https://www.state.gov/u-s-eu-trade-and-technology-council-ttc/>



Standards offer a first step to help erect and support guardrails in the international and market competition for AI.

for²³ the development and adoption of international technical standards for trustworthy AI.

Today, we are early in the development of the technology and even earlier in its governance. Standards offer a first step to help erect and support guardrails in the international and market competition for AI. It is therefore crucial that governments promote the awareness of this work so that the international standards developed take into consideration any country-specific policy objectives relating to AI, aiming to ensure that AI systems are accurate, reliable, safe and non-discriminatory, regardless of their origin.

Key AI Standards Principles

- Standards can play an important role in complementing emerging policies, laws and regulation around AI by connecting objectives and requirements with practical implementation. Standards also contribute to fostering innovation and enabling interoperability.
- Governments should promote engagement in AI global and regional standardization initiatives and leverage global standards development to support national AI policies and regulation.
- Initial AI standardization work should focus on finding consensus around AI foundational concepts, management and governance practices.
- AI standards need to be applicable to a variety of contexts and should not constitute barriers to trade.
- Compliance with policies and regulatory requirements should be achieved without mandating specific technology choices. ■

developing²² a voluntary code of conduct on artificial intelligence, hoping to develop common standards for applying AI to bridge the gap, ahead of legislation being passed to regulate uses of the AI in respective countries and regions around the world. Furthermore, leaders of the Group of Seven (G7) nations on May 20, 2023, called

22. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/31/u-s-eu-joint-statement-of-the-trade-and-technology-council-2/>

23. https://www.reuters.com/world/g7-calls-adoption-international-technical-standards-ai-2023-05-20/?utm_source=Sailthru&utm_medium=Newsletter&utm_campaign=E2%80%A6

HISTORY OF ARTIFICIAL INTELLIGENCE

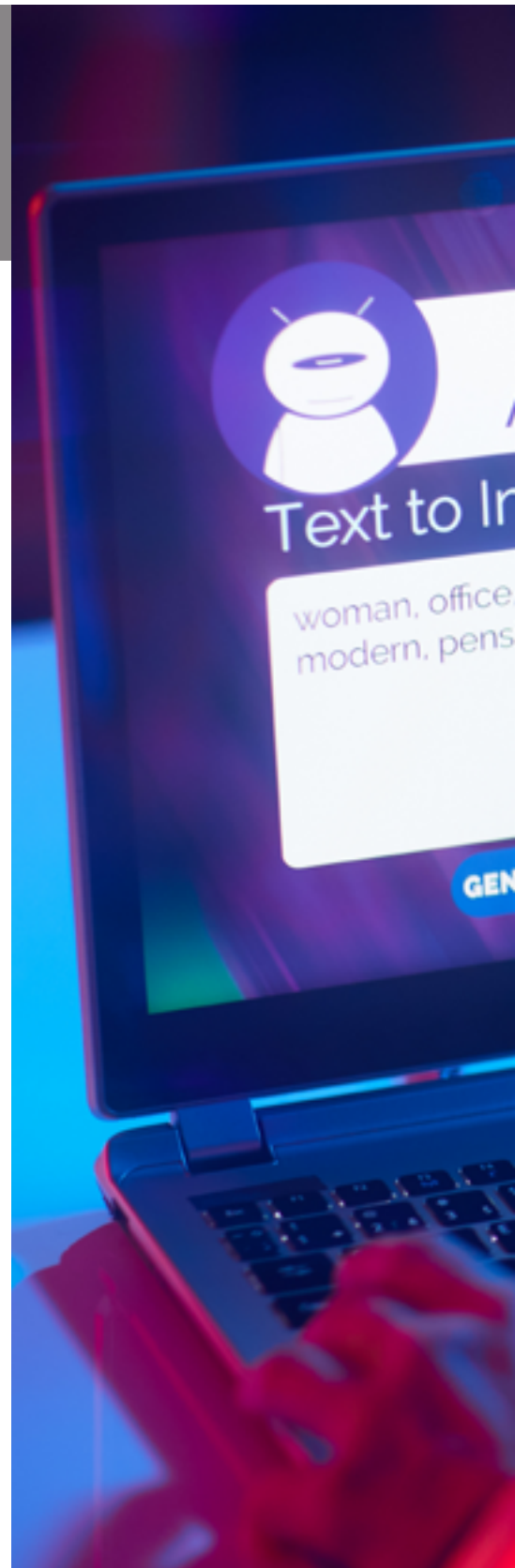
AI is a very broad field encompassing research into many different types of problems, from ad targeting to weather prediction, autonomous vehicles to photo tagging, chess playing to speech recognition. While the field of AI research as a whole has always included work on many different topics in parallel, the seeming center of gravity involving the most exciting progress has shifted over the years.

Human beings wouldn't have gotten very far without our machines. From the wheel that revolutionized agriculture to the screw that held together increasingly complex construction projects to the robot-enabled assembly lines of today, machines have made life as we know it possible.

The origin of AI dates back to antiquity, with myths, stories and rumors of artificial beings endowed with intelligence or consciousness by master craftsmen. Early concepts of AI were imagined by ancient philosophers who sought to describe the process of human thinking as the mechanical manipulation of symbols. This work culminated in the invention of the programmable digital computer²⁴ in 1945, a machine based on the abstract essence of mathematical reasoning. This device and the ideas behind it inspired a handful of scientists to begin seriously discussing the possibility of building an electronic brain.

Modern AI began at Dartmouth College with a workshop²⁵ in 1956, where researchers estimated that computers "as intelligent as a human being" would emerge within a single generation. Despite millions of dollars in subsequent funding, innovation and investments in AI stalled in the early

The origin of AI dates back to antiquity, with myths, stories and rumors of artificial beings endowed with intelligence or consciousness by master craftsmen. Early concepts of AI were imagined by ancient philosophers who sought to describe the process of human thinking as the mechanical manipulation of symbols.



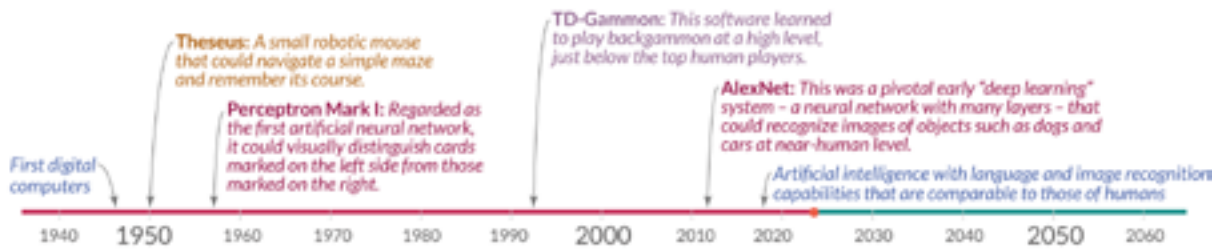
24. <https://en.wikipedia.org/wiki/Computer>

25. https://en.wikipedia.org/wiki/Dartmouth_workshop

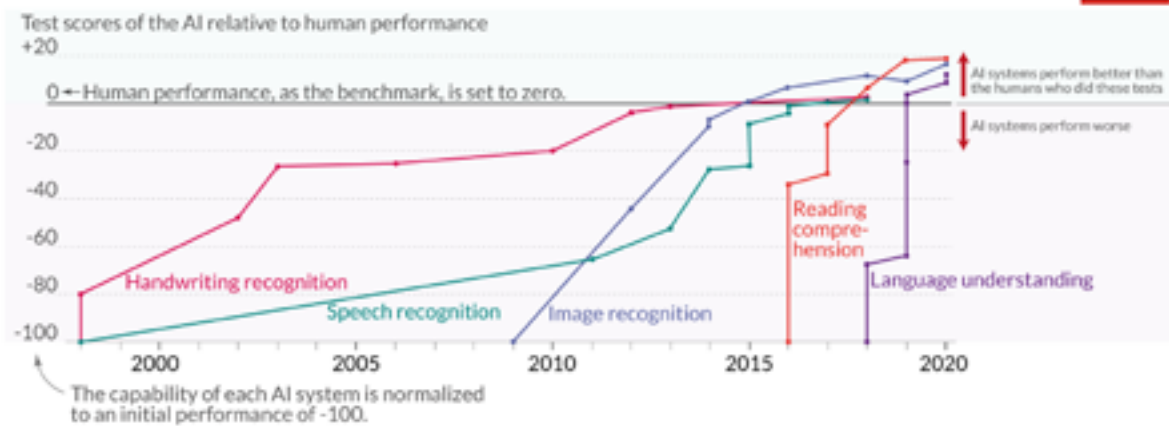


This device and the ideas behind it inspired a handful of scientists to begin seriously discussing the possibility of building an electronic brain.

A timeline of notable artificial intelligence systems



Language and image recognition capabilities of AI systems have improved rapidly



Data source: Kiela et al. (2021) - Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org - Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

Machine learning is now responsible for some of the most significant advancements in technology. It is being used for the new industry of self-driving vehicles, and for exploring the galaxy as it helps in identifying exoplanets.

1970s, leading to what is known as the first “AI Winter²⁶”. Then, in the early 1980s, a new era of innovation and funding emerged, followed by a second “AI Winter” less than a decade later. The current boom in AI funding and innovation started in beginning of the 21st Century, when machine learning had reached an advanced level of development and was successfully applied to many problems in industry and academia because of new methods as well as the availability of powerful and specialized computer hardware, and the emergence of large data sets.

Machine learning is now responsible for some of the most significant advancements in technology. It is

being used for the new industry of self-driving vehicles, and for exploring the galaxy as it helps in identifying exoplanets²⁷.

In the early 2010s²⁸ there was significant progress in image classification and speech recognition; in the mid-2010s²⁹, the focus shifted to reinforcement learning (especially for games such as Go and StarCraft); and since the late 2010s³⁰ and early 2020s³¹, there has been an explosive innovation in language and image generation. This chronological breakdown is very approximate, and it is important to note that work on all these areas—and many more—has been ongoing throughout these periods as well as much earlier.

26. https://en.wikipedia.org/wiki/AI_winter
 27. <https://learningenglish.voanews.com/a/machine-learning-helps-nasa-confirm-301-new-exoplanets/6326342.html>
 28. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
 29. <https://www.nature.com/articles/nature16961>
 30. <https://arxiv.org/abs/1706.03762>
 31. <https://arxiv.org/abs/2103.00020>



Smart machines are getting faster and more complex. Some computers have now crossed the exascale computing threshold, meaning that they can perform as many calculations in a single second as an individual could in 31,688,765,000 years. But it's not just about computation. Computers and other devices are now acquiring skills and perception that have previously been the sole purview of human beings.

As outlined by Our World in Data³², these rapid advances in AI capabilities have made it possible to use machines in a wide range of new domains. When you book a flight, it is often an artificial intelligence, and no longer a human, that decides³³ what you pay. When you get to the airport, it is an AI system that monitors³⁴ what you do at the airport. And once you are on the plane, an AI system assists the pilot in flying³⁵ you to your destination. AI systems also increasingly determine whether you get a loan³⁶, are eligible³⁷ for welfare, or get hired³⁸ for a particular job. Increasingly they help determine who gets released from jail³⁹. Many governments are purchasing autonomous weapons systems⁴⁰ for warfare, and others are using AI systems for surveillance and oppression⁴¹. AI systems help to program⁴² the software you use and translate⁴³ the texts you read. Virtual assistants⁴⁴, operated by speech recognition, have entered many

AI systems help to program the software you use and translate the texts you read. Virtual assistants, operated by speech recognition, have entered many households over the last decade.

households over the last decade. Now self-driving cars are becoming a reality. In the last few years, AI systems have helped to make progress⁴⁵ on some of the hardest problems in science.

Large AIs named recommender systems⁴⁶ determine what you see on social media, which products are shown to you in online shops, and what gets recommended to you on YouTube. Increasingly they are not just recommending the media we consume but based on their capacity to generate images and texts, they are also creating⁴⁷ the media we consume. Artificial intelligence is no longer the technology of the future; AI is here, and much of what is reality now would have looked like science-fiction just a few years ago. It is a technology that already impacts all peoples, in all corners of the world. ■

32. <https://ourworldindata.org/>

33. <https://www.bloomberg.com/news/articles/2022-10-20/artificial-intelligence-helps-airlines-find-the-right-prices-for-flight-tickets#xj4y7vzkg>

34. <https://www.sourcesecurity.com/news/co-2166-ga.132.html>

35. <https://www.airbus.com/en/innovation/industry-4-0/artificial-intelligence>

36. <https://www.brookings.edu/articles/reducing-bias-in-ai-based-financial-services/>

37. <https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625>

38. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

39. <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/>

40. https://en.wikipedia.org/wiki/Lethal_autonomous_weapon

41. <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>

42. https://en.wikipedia.org/wiki/GitHub_Copilot

43. https://en.wikipedia.org/wiki/Machine_translation

44. https://en.wikipedia.org/wiki/Virtual_assistant

45. <https://www.nature.com/articles/s42254-022-00518-3>

46. https://en.wikipedia.org/wiki/Recommender_system

47. <https://www.nature.com/articles/d41586-021-00530-0>



AI has become a catchall term for applications that perform complex tasks that once required human input, such as communicating with customers online or playing chess.

WHAT IS AI?

According to McKinsey⁴⁸, “artificial intelligence is a machine’s ability to perform the cognitive functions we usually associate with human minds”. AI has become a catchall term for applications that perform complex tasks that once required human input, such as communicating with customers online or playing chess. The term is often used interchangeably with its subfields, which include machine learning (ML) and deep learning.

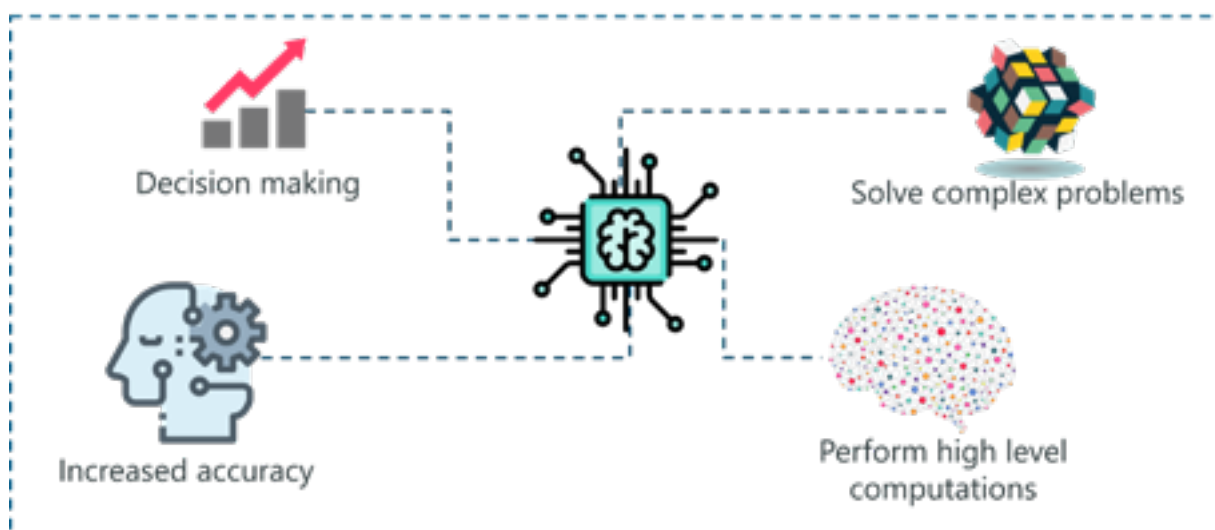
In 1955, the term Artificial Intelligence was defined by emeritus Stanford Professor John McCarthy⁴⁹. He defined AI as: ‘The science and engineering of making intelligent machines.’

Artificial Intelligence can also be defined as the development of computer systems that are capable of performing tasks that require human intelligence, such as decision making, object detection, solving complex problems and so on.

There are **three main stages**, through which AI can evolve:

1. Artificial Narrow Intelligence
2. Artificial General Intelligence
3. Artificial Super Intelligence

Artificial Intelligence can also be defined as the development of computer systems that are capable of performing tasks that require human intelligence, such as decision making, object detection, solving complex problems and so on.



What Is AI – Types Of Artificial Intelligence – Edureka

48. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai>

49. <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>



There are currently no existing examples of Strong AI, however, it is believed that we will soon be able to create machines that are as smart as humans.

Artificial Narrow Intelligence (ANI): Also known as Weak AI, ANI is the stage of Artificial Intelligence involving machines that can perform only a narrowly defined set of specific tasks. At this stage, the machine does not possess any thinking ability, it just performs a set of pre-defined functions. Examples of Weak AI include Siri, Alexa, Self-driving cars, Alpha-Go, Sophia the humanoid and so on. Almost all the AI-based systems built till this date fall under the category of Weak AI.

Artificial General Intelligence (AGI): Also known as Strong AI, AGI is the stage in the evolution of Artificial Intelligence wherein machines will possess the ability to think and make decisions just like us humans. There are currently no existing examples of Strong AI, however, it is believed that we will soon be able to create machines that are as smart as humans.

Artificial Super Intelligence (ASI): Artificial Super Intelligence is the stage of Artificial Intelligence when the capability of computers will surpass human beings. ASI is currently a hypothetical situation as depicted in movies and science fiction books, where machines have taken over the world.

There are **four main types** of AI:

1. Reactive Machines AI
2. Limited Memory AI
3. Theory Of Mind AI
4. Self-aware AI

Reactive Machine AI: This type of AI includes machines that operate solely based on the present data, taking into account only the current situation. Reactive AI machines cannot form inferences from the data to evaluate their future actions. They can perform a narrowed range of pre-defined tasks. An example of Reactive AI is the famous IBM Deep Blue Chess computer⁵⁰ that beat the world champion, Garry Kasparov in 1997.

50. <https://www.chess.com/terms/deep-blue-chess-computer#:~:text=Deep%20Blue%20was%20a%20chess,victory%20for%20machine%20versus%20man>

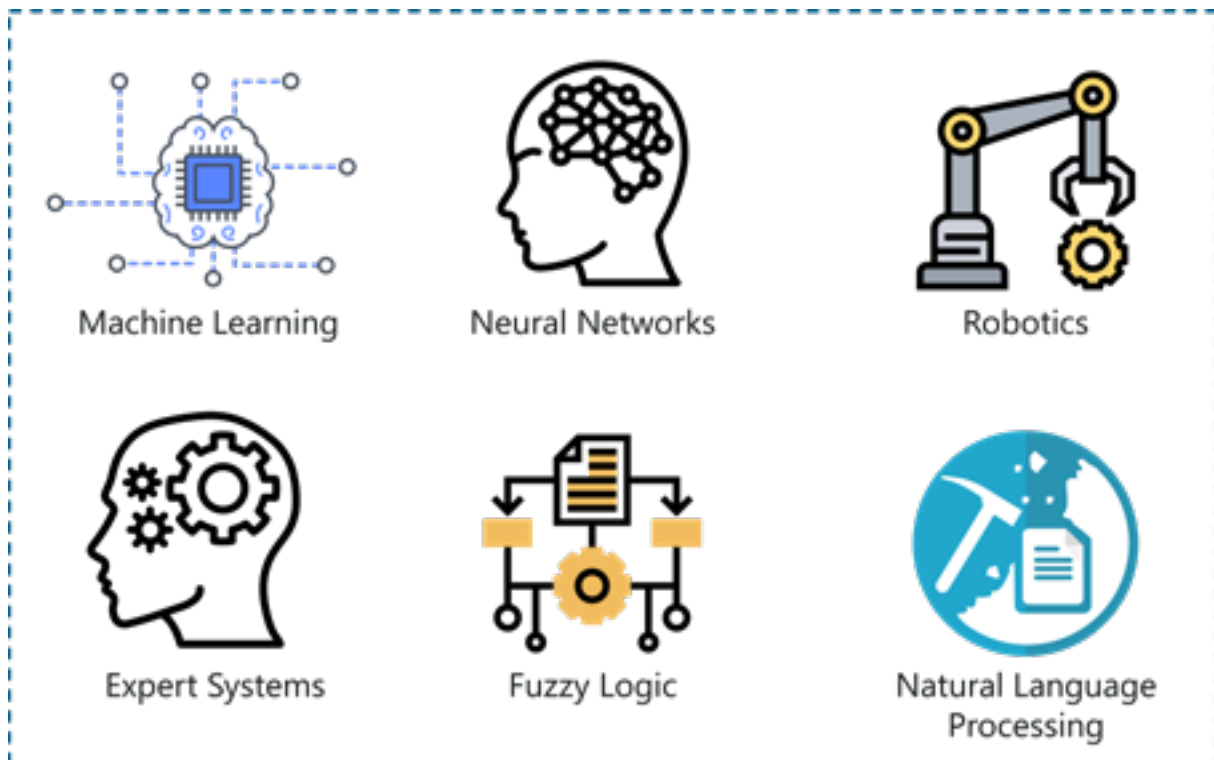
Limited Memory AI: Like the name suggests Limited Memory AI can make informed and improved decisions by studying the past data from its memory. Such an AI has a short-lived or a temporary memory that can be used to store past experiences and hence evaluate future actions. Self-driving cars are Limited Memory AI, that uses the data collected in the recent past to make immediate decisions. For example, self-driving cars use sensors to identify civilians crossing the road, steep roads, traffic signals and so on to make better driving decisions. This helps to prevent any future accidents.

Theory Of Mind AI: The Theory of Mind AI (ToM) is a more advanced

type of Artificial Intelligence. ToM, or the ability to impute unobservable mental states to others, is central to human social interactions, communication, empathy, self-consciousness, and morality. This type of AI focuses mainly on emotional intelligence so that human beliefs and thoughts can be better comprehended. ToM has not yet been fully developed but rigorous research is happening in this area. A February 2023 study⁵¹ conducted by Michal Kosinski, a

computational psychologist from Stanford University, used several iterations of OpenAI's GPT neural network—from GPT-1 to the latest GPT-3.5—to perform ToM tests, a series of experiments first developed in 1978⁵² to measure the complexity of a chimpanzee's mind to predict the behavior of others. Results show that GPT's ToM ability arrived spontaneously in the last two years and the latest iteration delivered results comparable to a 9-year-old human.

This type of AI focuses mainly on emotional intelligence so that human beliefs and thoughts can be better comprehended.



Domains Of AI – Types Of Artificial Intelligence – Edureka

51. <https://arxiv.org/abs/2302.02083>

52. <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/does-the-chimpanzee-have-a-theory-of-mind/1E96B02CD9850016B7C93BC6D2FEF1D0> <https://>

Self-Aware AI: The final type of AI is self-aware AI. This will be when machines are not only aware of emotions and mental states of others, but also their own. In theory, when self-aware AI is achieved, we would have AI that has human-level consciousness and equals human intelligence with the same needs, desires and emotions. At the moment, this AI has not been developed successfully yet because the hardware or algorithms that will support it are not yet available.

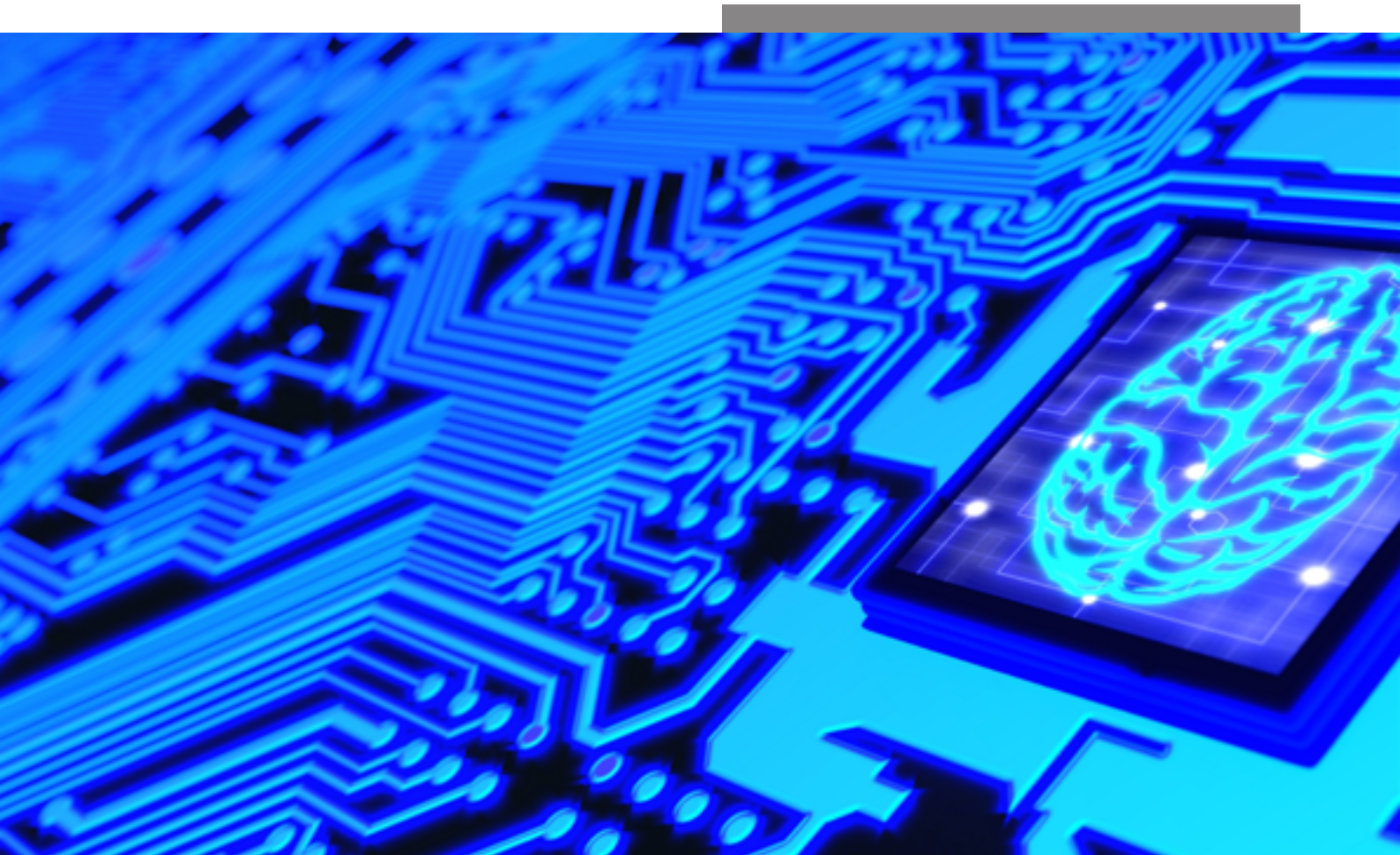
Artificial Intelligence can be used to solve real-world problems by implementing the following **processes/techniques**, or “**branches**” of AI:

1. Machine Learning
2. Deep Learning
3. Natural Language Processing
4. Robotics
5. Expert Systems
6. Fuzzy Logic

Machine Learning: Machine Learning is the science of getting machines to interpret, process and analyze data to solve real-world problems. Under Machine Learning there are three categories: (1) Supervised Learning; (2) Unsupervised Learning; and (3) Reinforcement Learning

Machine learning is, in part, based on a model⁵³ of brain cell interaction. The model was created in 1949 by Donald Hebb and until the late 1970s⁵⁴, was a part of

In theory, when self-aware AI is achieved, we would have AI that has human-level consciousness and equals human intelligence with the same needs, desires and emotions.



53. www.dataversity.net/what-is-machine-learning/

54. <https://www.dataversity.net/a-brief-history-of-machine-learning/>

AI's general evolution. Then, it branched off to evolve on its own. Machine learning has become a very important response tool for cloud computing and e-commerce and is being used in a variety of cutting-edge technologies. Machine learning is a necessary aspect of modern business and research for many organizations today. It uses algorithms and neural network models to assist computer systems in progressively improving their performance. Machine learning algorithms automatically build a mathematical model using sample data – also known as “training data” – to make decisions without being specifically programmed to make those decisions.

Deep Learning: Deep Learning is the process of implementing Neural Networks on high dimensional data to gain insights and form solutions. Deep Learning is an advanced field of Machine Learning that can be used to solve more complicated problems. Deep Learning is the logic behind the face verification algorithm on Facebook, self-driving cars, virtual assistants like Siri, Alexa and

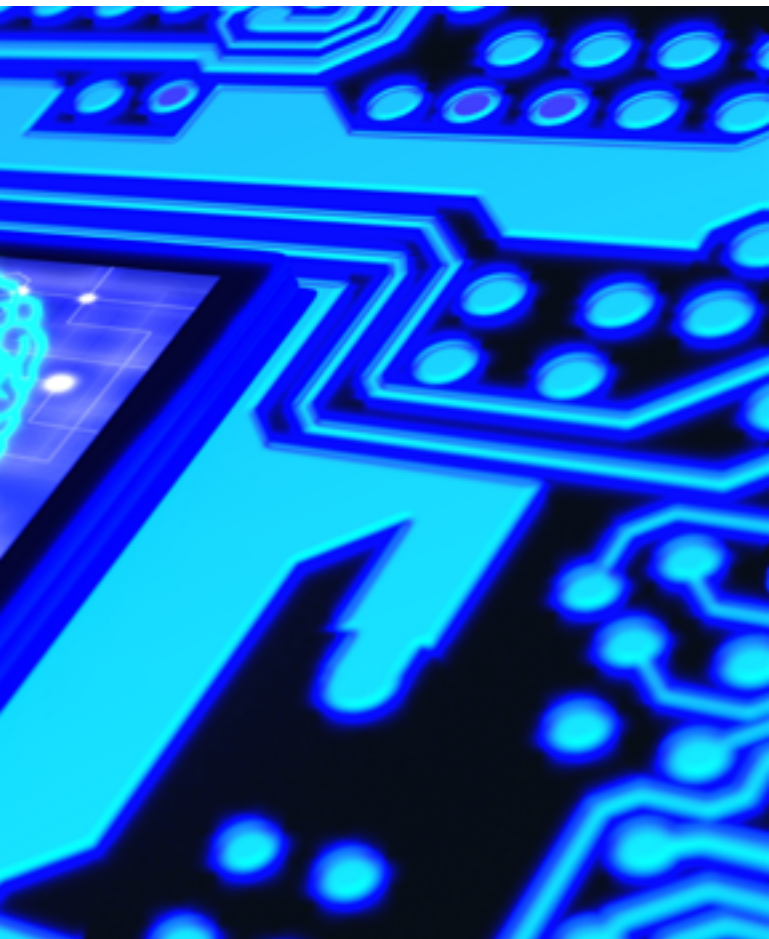
Machine learning algorithms automatically build a mathematical model using sample data – also known as “training data” – to make decisions without being specifically programmed to make those decisions.

more. The history of deep learning can be traced back to 1943⁵⁵, when Walter Pitts and Warren McCulloch created a computer model based on the neural networks⁵⁶ of the human brain. Deep learning has provided image-based product searches – eBay, Etsy– and efficient ways to inspect products on the assembly line. The first supports consumer convenience, while the second is an example of business productivity. The future evolution of artificial intelligence is dependent on deep learning, which is still evolving and subject to new and creative ideas.

Natural Language Processing (NLP): NLP refers to the science of drawing insights from natural human language to communicate with machines and grow businesses. Twitter uses NLP to filter out terroristic language in their tweets, Amazon uses NLP to understand customer reviews and improve user experience. NLP is a discipline with a long history. It was born in the 1950s as a sub-area of Artificial Intelligence and Linguistics, with the aim of studying the problems derived from the automatic generation and understanding of natural language. Although works from earlier periods can be found, it was in 1950 that Alan Turing⁵⁷ published an article entitled “Intelligence” that proposed what is now called the Turing test⁵⁸ as a criterion of intelligence.

Robotics: Robotics is a branch of Artificial Intelligence which focuses on different branches and applications of robots. AI Robots are artificial agents acting in a real-world environment to produce results by taking accountable actions. Sophia the humanoid⁵⁹ is a good example of AI in robotics.

Fuzzy Logic: Fuzzy logic is a computing approach based on the principles of “degrees of truth” instead of the usual modern computer logic i.e., Boolean⁶⁰ in nature. Generally, fuzzy logic AI systems are used for both commercial and practical purposes such as controlling machines and consumer products. If not accurate reasoning, it at least provides acceptable reasoning, which helps in dealing with the uncertainty in engineering.



55. <https://www.historyofinformation.com/detail.php?entryid=782>
 56. <https://www.dataversity.net/artificial-neural-networks-overview/>
 57. https://en.wikipedia.org/wiki/Alan_Turing
 58. https://en.wikipedia.org/wiki/Turing_test
 59. [https://en.wikipedia.org/wiki/Sophia_\(robot\)](https://en.wikipedia.org/wiki/Sophia_(robot))
 60. <https://www.techtarget.com/whatis/definition/Boolean>



Fuzzy logic is used in the medical field to solve complex problems that involve decision making. They are also used in automatic gearboxes, vehicle environment control and so on.

Expert Systems: An expert system is an AI-based computer system that learns and reciprocates the decision-making ability of a human expert. It uses AI technologies to simulate the judgment and behavior of a human or an organization that has expertise and experience in a particular field. Modern expert knowledge systems use machine learning and artificial intelligence to simulate the behavior or judgment of domain experts. These systems can improve their performance over time as they gain more experience, just as humans do. The concept of expert systems was developed in the 1970s by computer scientist Edward Feigenbaum⁶¹, a computer science professor at Stanford University and founder of Stanford's Knowledge Systems Laboratory. In the first decade of the 2000s, there was a "resurrection" for the technology, while using the term Rule Based

Modern expert knowledge systems use machine learning and artificial intelligence to simulate the behavior or judgment of domain experts.

Systems⁶², with significant success stories and adoption. Many of the leading major business application suite vendors (such as SAP, Siebel, and Oracle) integrated expert system abilities into their suite of products as a way of specifying business logic – rule engines are no longer simply for defining the rules an expert would use but for any type of complex, volatile, and critical business logic; they often go hand in hand with business process automation and integration environments.

Other terms: Generative AI, Large Language Models, and Foundation Models

These three terms suddenly seem to be everywhere and are often used interchangeably.

Generative AI is a broad term that can be used for any AI system whose primary function is to generate content. This is in contrast to AI systems that perform other functions, such as classifying data (e.g., assigning labels to images), grouping data (e.g., identifying customer segments with similar purchasing behavior), or choosing actions (e.g., steering an autonomous vehicle). Typical examples of generative AI systems include image generators (such as Midjourney⁶³ or Stable Diffusion⁶⁴), large language models (such as GPT-4⁶⁵, PaLM⁶⁶, or Claude⁶⁷), code generation tools (such as Copilot⁶⁸), or audio generation tools (such as VALL-E⁶⁹ or resemble.ai⁷⁰). *Using the term "generative AI" emphasizes the content-creating function of these systems. It is a relatively intuitive term that covers a range of types of AI that*

61. https://en.wikipedia.org/wiki/Edward_Feigenbaum

62. https://en.wikipedia.org/wiki/Rule-based_system

63. <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>

64. <https://stablediffusionweb.com/>

65. <https://openai.com/research/gpt-4>

66. <https://ai.google/discover/palm2/>

67. <https://www.anthropic.com/index/introducing-claude>

68. <https://github.com/features/copilot>

69. <https://vall-e.pro/>

70. <https://www.resemble.ai/>

The original model provides a base (“foundation”) on which other things can be built. This is in contrast to many other AI systems, which are specifically trained and then used for a *particular purpose*.

LLMs themselves, or to say that they are powered by underlying LLMs.

Foundation model is a term popularized⁸¹ by Stanford Institute for Human-Centered Artificial Intelligence (HAI⁸²). It refers to AI systems with broad capabilities that can be adapted to a range of different, more specific purposes. The original model provides a base (“foundation”) on which other things can be built. This is in contrast to many other AI systems, which are specifically trained and then used for a *particular purpose*. “Foundation model” is often used roughly synonymously with “large language model” as language models are currently the clearest example of systems with broad capabilities that can be adapted for specific purposes. The relevant distinction between the terms is that “large language models” specifically refers to language-focused systems, while “foundation model” is attempting to stake out a broader function-based concept, which could stretch to accommodate new types of systems in the future. Common examples of foundation models include many of the same systems referenced as LLMs. To illustrate what it means to build something more specific on top of a broader base, consider ChatGPT. For the original ChatGPT, an LLM called GPT-3.5⁸³ served as the foundation model. Essentially, OpenAI used chat-specific data to create a tweaked version of GPT-3.5 that was optimized for a chatbot setting and subsequently built into ChatGPT. ■

have progressed rapidly in recent years.

Generative AI became much more powerful with the introduction of generative adversarial networks (GAN⁷¹) in 2014, which produced the first practical deep neural networks capable of learning generative, rather than discriminative, models of complex data such as images. These deep generative models were the first able to output not only class labels for images, but to output entire images. Then, in 2017, Transformers⁷² were introduced by a team of Google researchers who were looking to build a more efficient translator. In a paper entitled “Attention Is All You Need⁷³,” the researchers laid out a new technique to discern the meaning of words based on how they characterized other words in phrases, sentences and essays. Transformers are the current state-of-the-art foundational technology underpinning many advances in large language models, such as generative pre-trained transformers (GPTs⁷⁴). They’re now expanding into multimodal AI⁷⁵ applications capable of correlating content as diverse as text, images, audio and

robot instructions across numerous media types more efficiently than techniques like GANs.

Large language models (LLMs)

are a type of AI system that works with language. In the same way that an aeronautical engineer might use software to model an airplane wing, a researcher creating an LLM aims to model language, i.e., to create a simplified—but useful—digital representation. The “large” part of the term describes the trend towards training language models with more parameters.¹ A key finding of the past several years of language model research has been that using more data and computational power to train models with more parameters consistently results in better performance. Accordingly, cutting-edge language models trained today might have thousands or even millions of times as many parameters as language models trained ten years ago, hence the description “large.” Typical examples of LLMs include OpenAI’s GPT-4⁷⁶, Google’s PaLM⁷⁷, and Meta’s LLaMA⁷⁸. There is some ambiguity about whether to refer to specific products (such as OpenAI’s ChatGPT⁷⁹ or Google’s Bard⁸⁰) as

71. <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>
72. <https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/>
73. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
74. https://en.wikipedia.org/wiki/Generative_pre-trained_transformer
75. https://www.techtarget.com/searchenterpriseai/definition/multimodal-AI?Offer=abMeterCharCount_var2
76. <https://openai.com/research/gpt-4>
77. <https://ai.google/discover/palm2/>
78. <https://ai.meta.com/blog/large-language-model-llama-meta-ai/>
79. <https://openai.com/blog/chatgpt>
80. <https://bard.google.com/chat>
81. <https://arxiv.org/abs/2108.07258>
82. <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>
83. <https://platform.openai.com/docs/models/gpt-3-5>

BENEFITS OF AI

The time may have finally come for artificial intelligence (AI) after periods of hype followed by several “AI winters” over the past 60 years. AI now powers so many real-world applications, ranging from facial recognition to language translators and assistants like Siri and Alexa, that we barely notice it. Along with these consumer applications, companies across sectors are increasingly harnessing AI’s power in their operations. Embracing AI promises considerable benefits for businesses and economies through its contributions to productivity growth and innovation.

Industrial operations: According to McKinsey, AI can be used to improve business performance⁸⁴ in areas including predictive maintenance, where deep learning’s ability to analyze large amounts of high-dimensional data from audio and images can effectively detect anomalies in factory assembly lines or aircraft engines. A 2018 McKinsey Global Institute analysis⁸⁵ of more than 400 use cases across 19 industries and nine business functions found that AI improved on traditional analytics techniques in 69 percent of potential use cases⁸⁶. It estimated that deep learning techniques based on artificial neural networks could generate as much as 40 percent of the total potential value that all analytics techniques could provide by 2030 and that several of the deep learning techniques could enable up to *\$6 trillion* in value annually.

AI can help prevent breakdowns before they happen through machine learning by monitoring machine-component performance through networked sensors in real time to detect early warning signs that a machine might be breaking down, prompting preventative measures. Other AI benefits include:

- intelligent manufacturing to automate factories and allowing human employees to focus on monitoring the factory floor and operating computer system;
- maximize oil-well performance by installing networked sensors in subsea oil wells and use machine-learning software to monitor oil-well performance in real time, resulting in maximized production and minimized downtime;
- improve dairy supply chains through machine-learning tools to forecast sales estimates and production output;
- create smarter industrial design through an AI software based on a designer’s specific criteria, such as function, cost, and material.



AI can help prevent breakdowns before they happen through machine learning by monitoring machine-component performance through network sensors in real time to detect early warning signs that a machine might be breaking down, prompting preventative measures.

84. <https://www.mckinsey.com/featured-insights/artificial-intelligence/the-real-world-potential-and-limitations-of-artificial-intelligence>

85. <https://www.mckinsey.com/featured-insights/artificial-intelligence/the-promise-and-challenge-of-the-age-of-artificial-intelligence>

86. <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>



By automating routine or unsafe activities and those prone to human error, AI could allow humans to be more productive and to work and live more safely.

Economies also stand to benefit from AI, through increased productivity and innovation. Deployment of AI and automation technologies can do much to lift the global economy and increase global prosperity. At a time of aging and falling birth rates, productivity growth becomes critical for long-term economic growth.

Alongside the economic benefits and challenges, AI will impact society in a positive way, as it helps tackle societal challenges ranging from health and nutrition to equality and inclusion. AI can help tackle some of society's most pressing challenges. By automating routine or unsafe activities and those prone to human error, AI could allow humans to be more productive and to work and live more safely.

Public safety: AI can also reduce the need for humans to work in unsafe environments such as offshore oil rigs and coal mines. DARPA, in the DARPA Robotics Challenge⁸⁷, for example, is testing small robots that can be deployed in disaster areas to reduce the need for humans to be

put in harm's way. Other examples include pinpointing gunshots to fight crime by implementing gunshot-detecting systems that uses networked audio sensors dispersed through city blocks and machine learning to automatically identify the audio signatures of gunshots and report their location to police with high degrees of accuracy; predicting crime hotspots in major cities using natural language processing and machine learning to analyze hundreds of data points, such as social-media activity, emergency-call locations, geotagged posts' proximity to schools, and other variables to create heat maps of areas likely to have elevated levels of criminal activity; autonomously disposing of car bombs by using autonomous robotic platforms that coordinate with each other to position themselves under a vehicle suspected of having an explosive device and move it to a safe location; predicting buildings' fire risk with machine-learning models that can analyze data to recommend inspection for buildings that pose the highest fire risks, to help fire and rescue departments better prioritize inspections; and making security screening less invasive

87. <https://www.darpa.mil/program/darpa-robotics-challenge>

through machine learning to quickly analyze the contents of bags at security checkpoints and identify dangerous objects without humans having to sort through them manually.

Social good: AI has the potential to bring about numerous positive changes in society, including enhanced productivity, improved healthcare, and increased access to education. AI-powered technologies can also help solve complex problems and make our daily lives easier and more convenient. AI can help by mapping poverty with satellite data utilizing machine-learning algorithms that can analyze satellite imagery to map impoverished areas, which can help governments and development organizations improve aid efforts; measuring literacy rates with machine-learning algorithms that can determine literacy rates in developing countries by analyzing mobile-phone data; cracking down on human trafficking with AI tools that scan pages on the deep web— websites that are not indexed on search engines—and analyze their contents for signs of illegal solicitations of sex, which are often linked with human trafficking, to aid investigations; stopping abusive internet trolls by utilizing AI tools to promote freedom of expression and combat extremism online, designed to detect and filter out abusive language online to combat online harassment (such as Jigsaw⁸⁸ trained Conversation AI); and supporting refugees' mental health using natural language processing to have conversations with users in their native languages via text message and analyze their emotional states to provide recommendations that can help improve their mental health.

Transportation & logistics: AI can optimize routing of delivery traffic, improving fuel efficiency and reducing delivery times. AI may also receive better information about accidents, weather events, and other disruptions than human drivers do. Moreover, it may also take control in dangerous situations. AI-controlled vehicles are also programmed to drive more efficiently than human drivers, which can help reduce traffic congestion. AI can help in making public transportation autonomous with self-driving public shuttle buses that use AI to help navigate public roads and drive among regular traffic; platooning autonomous trucks — monitoring a fleet of autonomous trucks' speed, proximity, and the road to drive close together to improve efficiency, much like bicyclists drafting off each other; hailing a self-driving taxi to pick up and drop off passengers almost entirely autonomously; teaching trains to drive themselves autonomously and more safely and efficiently, and with less downtime than human-operated trains; and making long-haul trucking easier by utilizing computer-vision systems that analyze truck parking lots along the highway to automatically detect when a spot is available and notify approaching truckers.

Healthcare: AI algorithms can monitor patients' health data over time and provide recommendations for lifestyle changes and treatment options that can help manage



their condition. This can lead to better patient outcomes, improved quality of life, and reduced health care costs. Image classification performed on photos of skin taken via a mobile phone app can evaluate whether moles are cancerous, facilitating early-stage diagnosis for individuals with limited access to dermatologists. Object detection can help visually impaired people navigate and interact with their environment by identifying obstacles such as cars and lamp posts. Natural language processing can be used to track disease outbreaks by monitoring and analyzing text messages in local languages.

Object detection can help visually impaired people navigate and interact with their environment by identifying obstacles such as cars and lamp posts. Natural language processing can be used to track disease outbreaks by monitoring and analyzing text messages in local languages.

88. <https://jigsaw.google.com/the-current/toxicity/>



AI can also foster greater treatment and monitoring by helping diabetes patients make smarter diet decisions through smartphone apps for patients with type 2 diabetes that uses AI to analyze medical research and users' behavior to make personalized recommendations about how they can alter their diet to better manage their disease

Other examples include preventing vision loss in diabetes patients through the usage of image-analysis algorithms that can analyze retinal scans of diabetes patients and learn to identify subtle signs of diabetes-linked retinal damage with high accuracy, faster than traditional human analysis and without needing to send scans to a lab; predicting schizophrenia by analyzing speech with machine-learning systems capable of predicting if a person at risk of developing psychosis caused by schizophrenia will develop the condition with 100 percent accuracy by analyzing his or her speech, which can exhibit telltale signs of the condition; figuring out how to prevent pancreatic cancer by using AI to analyze large amounts of oncological data to create a complete model of how pancreatic cancer functions; automating a microscope to diagnose malaria by using artificial neural networks to rapidly analyze blood samples in the field and diagnose malaria with near perfect accuracy; and diagnosing voice disorders by using systems that use wearable devices to collect data about the movement of a user's vocal cords and uses machine learning to detect subtle signs of abnormal speech that could indicate a person has voice disorders such as muscle tension dysphonia (MTD).

AI can also foster greater treatment and monitoring by helping diabetes patients make smarter diet decisions through smartphone apps for patients with type 2 diabetes that use AI to analyze medical research and

users' behavior to make personalized recommendations about how they can alter their diet to better manage their disease; streamlining drug discovery by leveraging machine-learning systems to prioritize which experiments they should conduct to test new drugs, reducing the number of unnecessary tests; making stitches safer with robotic surgeons that can administer stitches more precisely than human surgeons; using AI to speed radiotherapy with AI systems that can reduce the time needed to provide radiotherapy treatment to patients with head and neck cancers; and increasing participation in clinical trials by utilizing machine-learning systems to evaluate whether a patient is likely to participate in a clinical trial by analyzing objective and subjective factors about a patient, such as age, race, attitude toward medical research and health conditions.

Accessibility: Making the Internet more accessible for people with visual impairments, helping people understand sign language, interpreting emotions in facial expressions in order to allow parents of children with autism to help their kids improve the ability to recognize emotions, making it easier to get around in a wheelchair by using computer-vision algorithms to automatically help users navigate their environments, identifying dangerous sounds associated with dangerous situations through machine learning, such as sirens or squealing tires, and warn hard-of-hearing users about the noise.



AI can be used to predict area-specific weather implications and provide targeted weather analysis; aid in disaster prevention and response, such as predicting where earthquakes do the most damage, keeping emergency responders out of danger by using AI to monitor data to detect signs of danger

most damage, keeping emergency responders out of danger by using AI to monitor data to detect signs of danger; detecting disease outbreaks by using AI to analyze news stories in multiple languages; understanding a crisis with social media by using machine learning to monitor and analyze social media posts and automatically compile social media activity related to a particular crisis to aid humanitarian response; and avoiding dangerous solar flares using machine-learning systems that can predict M- and X-class solar flares, which produce dangerously high levels of radiation that could harm airline passengers, damage power grids, and disrupt communication satellites.

Agriculture: Farming indoors autonomously using networked sensors and machine learning to constantly monitor an indoor farm's environment and plant growth and adjust lighting, temperature, humidity, water, and soil nutrient levels to maximize a farm's productivity; learning as soon as plants get sick, using machine-learning algorithm to monitor crops and warn farmers; forecasting crop yields from space by adapting deep-learning image-analysis software to analyze satellite photos of farmland to forecast crop yields faster and more accurately than official government estimates; spot-treating crops utilizing robotic systems that can drive through a field and take photos of plants as well as using computer-vision algorithms to identify weeds and spray targeted bursts of herbicide directly on them, rather than the whole field; and making vegetable sorting easy with robotic systems powered by machine-learning software that can automatically sort vegetables based on their visual differences such as size and shape.

Weather & disaster prevention: AI can be used to predict area-specific weather implications and provide targeted weather analysis; aid in disaster prevention and response, such as predicting where earthquakes do the

Education: AI has the potential to solve several modern education challenges such as closing the technology gap between students and teachers, keeping the learning system ethical and transparent, allowing remote learning, and developing quality data and information solutions for the modern education process. AI can help personalize math classes by helping math teachers develop personalized lesson plans; predicting which students will drop out by using machine learning models that can analyze student data, such as demographics and

academic performance, and historical data to predict which students were at risk of dropping out and prompt early intervention; automating teacher assistants to help respond to student inquiries for online courses; making it easier to learn new languages by using machine learning to analyze users' activity and progression to develop personalized lesson plans, as well as regularly test new strategies for instruction to evaluate their effectiveness; and giving students feedback in real time by using machine learning to evaluate students' writing while they draft essays to provide feedback.

Legal: AI has the potential to transform⁸⁹ the legal profession. It could reduce big firms' manpower advantage. In large, complex lawsuits, these firms tell dozens of associates to read millions of pages of documents looking for answers to senior lawyers' questions and hunches. Now a single lawyer or small firm will be able to upload these documents into a litigation-prep AI system and begin querying them. AI could also change how firms make money. Firms profit by having armies of young lawyers to whom they pay less than they charge clients. If AI can do the work of those armies in seconds, firms may move to charging flat fees based on the service provided, rather than for the amount of time spent providing it. AI could make legal services cheaper, particularly for small and medium-sized businesses that currently struggle to afford them.

Energy: Some of the possible applications of Artificial Intelligence in energy include but are not limited to smart grids, data digitalization, forecasting, and more advanced resource management. AI can help transform energy companies⁹⁰ by automating grid data collection and implementing analysis frameworks. With the vast amount of data existing in the energy sector, converting it into reusable information for AI and Machine Learning algorithms is a go-to option. AI can also predict renewable energy availability with machine-learning

systems that analyzes data from weather stations, solar plants, wind farms, and weather satellites to generate weather forecasts more accurately than previously possible and predict renewable energy availability up to weeks in advance; model energy consumption for more efficient buildings by using machine learning to create highly granular models of a building's energy efficiency; teaching data centers to make themselves more efficient by utilizing software to automatically optimize energy efficiency while responding to factors such as increased usage and changing weather; learning how to manage home energy use by learning homeowner's preferences and schedules to optimize home heating and cooling and save consumers energy costs; and even picking the best spot for wind farms with machine-learning systems that can predict variations in wind speeds over time to help power companies more quickly evaluate potential locations for wind farms.

Environment: AI-enabled technologies have huge potential to support positive climate action. From digital twin technology that model the Earth, to algorithms to make data centers more efficient, AI applications already support the green transition. AI technologies can help stopping deforestation before it starts by using AI to analyze satellite imagery of forests over time to detect early warning signs of illegal logging that can prompt intervention before any trees are cut down; predicting dangerous air pollution levels by using machine learning systems that can analyze data about pollution levels in adversely affected cities to forecast changes to air quality days in advance and with more accurate than traditional forecasting; improving antipoaching efforts by using AI to identify poaching routes and optimize resource deployment to prevent poaching in affected areas; saving threatened birds by using systems of acoustic sensors and machine learning to improve efforts to save the threatened bird populations.

AI is teaching robots to recycle by taking advantage of autonomous recycling systems⁹¹ utilizing robotic arms, arrays of sensors, and algorithms to identify recyclable items in waste and separate them for recycling, removing the need for manual sorting. AI is also increasingly used to help consumers improve at recycling⁹² with consumer-oriented scanners placed in cafeterias and restaurants, smartphone apps, and QR code product labeling. ■

AI is teaching robots to recycle by taking advantage of autonomous recycling systems utilizing robotic arms, arrays of sensors, and algorithms to identify recyclable items in waste and separate them for recycling, removing the need for manual sorting.

89. https://www.economist.com/business/2023/06/06/generative-ai-could-radically-alter-the-practice-of-law?utm_content=article-image-5&etear=nl_today_5&utm_campaign=r.the-economist-today&utm_medium=email.internal-newsletter.np&utm_source=salesforce-marketing-cloud&utm_term=6/6/2023&utm_id=1626755

90. <https://www.n-ix.com/artificial-intelligence-in-energy/#:~:text=AI%20can%20help%20transform%20energy,Smart%20forecasting.>

91. <https://www.axios.com/2023/04/04/recycling-robots-ai-landfill>

92. https://www.axios.com/2023/06/05/ai-recycling-garbage-sorting-trash-oscar-intuitive?mc_cid=a7377f370a&mc_eid=949c1157eb

RISKS & THREATS OF AI

The recent acceleration in both the power and visibility of AI systems, and growing awareness of their abilities and defects, have raised fears that the technology is now advancing so quickly that it cannot be safely controlled. Many academic and tech luminaries have been vocal about existential threats posed by AI, including the famed theoretical physicist, cosmologist, and author Stephen Hawking, who once cautioned⁹³ that:

“The development of full artificial intelligence could spell the end of the human race... It would take off on its own, and re-design itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn’t compete, and would be superseded.” Stephen Hawking

Business magnate and investor Elon Musk has stated⁹⁴ that:

“The pace of progress in artificial intelligence (I’m not referring to narrow AI) is incredibly fast. Unless you have direct exposure to groups like Deepmind, you have no idea how fast—it is growing at a pace close to exponential. The risk of something seriously dangerous happening is in the five-year timeframe. 10 years at most.” Elon Musk

A March 22, 2023, letter⁹⁵ from the Future of Life Institute signed by more than 1,000 tech leaders, researchers and others poignantly asked:

“Should we automate away all the jobs, including the fulfilling ones? Should we develop non-human minds that might eventually outnumber, outsmart...and replace us? Should we risk loss of control of our civilization?” Future of Life Institute

On May 30, 2023, leaders from OpenAI, Google DeepMind, Anthropic and other AI labs signed a one-sentence statement⁹⁶ released by the Center for AI Safety, a nonprofit organization, warning that:

“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war.” Center for AI Safety

Whether or not such warnings about existential risks from AI are warranted, as the world witnesses unprecedented growth in AI technologies, it is important to contemplate the potential risks and challenges associated with their widespread adoption. It is indisputable that AI does present some very real threats— from job loss to security and privacy concerns — and promoting awareness of issues helps us engage in constructive conversations about AI's legal, ethical, and societal implications. A summary of some of the most discussed risks from AI is provided below:



93. <https://bernardmarr.com/28-best-quotes-about-artificial-intelligence/#:~:text=%E2%80%9CThere%20is%20no%20reason%20and,artificial%20intelligence%20machine%20by%202035.%E2%80%9D&text=%E2%80%9CIs%20artificial%20intelligence%20less%20than%20our%20intelligence%3F%E2%80%9D&text=%E2%80%9CBy%20far%2C%20the%20greatest%20danger,early%20that%20they%20understand%20it.%E2%80%9D>
94. <https://dataconomy.com/2014/11/19/tesla-motors-and-spacex-ceo-elon-musk-expresses-concern-about-rapidly-advancing-artificial-intelligence/#:~:text=%E2%80%9CThe%20pace%20of%20progress%20in,firm%20Musk%20recently%20invested%20in.>
95. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
96. <https://www.safe.ai/statement-on-ai-risk>

These models often rely on intricate algorithms that are not easily understandable to humans, leading to a lack of accountability and trust.

Transparency & the Black Box problem: A lack of transparency in AI systems, particularly in deep learning models that can be complex and difficult to interpret, is an urgent issue. The black box problem stems from the difficulty in understanding how AI systems and machine learning models process data and generate predictions or decisions. These models often rely on intricate algorithms that are not easily understandable to humans, leading to a lack of accountability and trust. This opaqueness obscures the decision-making

processes and underlying logic of these technologies. When people can't comprehend how an AI system arrives at its conclusions, it can lead to distrust and resistance to adopting these technologies.

Bias & Discrimination: AI systems can inadvertently perpetuate or amplify societal biases due to biased training data or algorithmic design. There is ample evidence of the discriminatory harm that AI tools can cause to already marginalized groups. After all, AI is built by humans and deployed in systems and institutions that have been marked by entrenched discrimination — from the criminal legal system, to housing, to the workplace, to our financial systems. Bias is often baked into the outcomes the AI is asked to predict. To minimize discrimination and ensure fairness, it is crucial to invest in the development of unbiased algorithms and diverse training data sets. These concerns are also troubling in the high-risk setting that is healthcare, and



As AI technologies become increasingly sophisticated, the security risks associated with their use and the potential for misuse also increase. Hackers and malicious actors can harness the power of AI to develop more advanced cyberattacks, bypass security measures, and exploit vulnerabilities in systems.



even more so because marginalized populations—those that already face discrimination from the health system from both structural factors and scientific factors may lose even more. Biases in these approaches can have literal life-and-death stakes.

Ethical dilemmas: Instilling moral and ethical values in AI systems, especially in decision-making contexts with significant consequences, presents a considerable challenge. AI decisions are not always intelligible to humans. AI is not neutral⁹⁷: AI-based decisions are susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias. Surveillance practices for data gathering and privacy of court users. Researchers and developers must prioritize the ethical implications of AI technologies to avoid negative societal impacts.

Privacy concerns: AI technologies often collect and analyze large amounts of personal data, raising issues related to data privacy and security. The main privacy concerns surrounding AI are the potential for data breaches and unauthorized access to personal information. With so much data being collected and processed, there is a risk that it could fall into the wrong hands, either through hacking or other security breaches. To mitigate privacy risks, strict data protection

and safe data handling practices are essential.

AI Security Risks: As AI technologies become increasingly sophisticated, the security risks associated with their use and the potential for misuse also increase. Hackers and malicious actors can harness the power of AI to develop more advanced cyberattacks, bypass security measures, and exploit vulnerabilities in systems. An April 2023 Stanford and Georgetown report⁹⁸ outlines how AI systems, especially those based on the techniques of machine learning, are remarkably vulnerable to a range of attacks and that evasion, data poisoning, and exploitation of traditional software flaws can deceive, manipulate, and compromise AI systems⁹⁹, to the point of rendering them ineffective. To mitigate these security risks, governments and organizations need to develop best practices for secure AI development and deployment and foster international cooperation to establish global norms and regulations that protect against AI security threats.

Job displacement: AI-driven automation has the potential to lead to job losses across many industries, particularly for low-skilled workers, although there is evidence that AI and other emerging technologies will create more jobs than it eliminates: A World Economic

Forum report¹⁰⁰ predicts that the "number of jobs destroyed will be surpassed by the number of 'jobs of tomorrow' created." According to the October 2020 report, by 2025, machines could displace about 85 million jobs — but create 97 million new roles "more adapted to the new division of labor between humans, machines and algorithms." As AI technologies continue to develop and become more efficient, the workforce must adapt and acquire new skills to remain relevant in the changing landscape.

Environmental impact: With increased data usage also comes an increased carbon footprint. Not surprisingly, the data sets required to train and run AI models are large and often complex, leading to higher energy consumption. According to an MIT study¹⁰¹, the cloud now has a larger carbon footprint than the entire airline industry and training a single AI model can emit more than 626,000 pounds of carbon dioxide

97. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases#:~:text=But%20there%20are%20many%20ethical,and%20privacy%20of%20court%20users.>
 98. https://fsi9-prod.s3.us-west-1.amazonaws.com/s3fs-public/2023-04/adversarial_machine_learning_and_cybersecurity_v7_pdf_1.pdf
 99. <https://www.lawfaremedia.org/article/managing-cybersecurity-vulnerabilities-artificial-intelligence>



Even more than other forms of machine learning, generative AI requires dizzying amounts of computational power and specialized computer chips, such as GPUs, that only the wealthiest of companies can afford.

equivalent [2]. As many organizations are monitoring their carbon footprints closely, this adds another layer of consideration to decisions made around data centers, machine learning and energy use.

In a December 2022 report¹⁰² on artificial intelligence, the U.S. Whitehouse pinpointed the computational costs of generative AI as a national concern. The White House wrote that the technology is expected to “dramatically increase computational demands and the associated environmental impacts,” and that there’s an “urgent need” to design more sustainable systems. Based on estimates of ChatGPT’s usage and computing needs, data scientist Kasper Groes Albin Ludvigsen estimated that it may have used as much electricity¹⁰³ in January 2023 as 175,000 people — the equivalent of a midsize city. While the data computations have implications for greenhouse gas emissions, it also diverts energy resources that could be used for other purposes¹⁰⁴ — such as AI computing tasks other than AI large language models, potentially slowing down the development and application of AI for other meaningful uses, such as in health care, drug discovery, and cancer detection.

Economic Inequality: AI has the potential to contribute to economic inequality by disproportionately benefiting affluent individuals and corporations. Job losses due to AI-driven automation are more likely to affect low-skilled workers, potentially leading to a growing income gap and reduced opportunities for social mobility. The concentration of AI development and ownership within a small number of large corporations and governments can exacerbate this inequality as they accumulate wealth and power while smaller businesses may struggle to compete. Even more than other forms of machine learning, generative AI requires dizzying amounts of computational power and specialized computer chips, such as GPUs, that only the wealthiest of companies can afford. Policies and initiatives such as reskilling programs, social safety nets, and inclusive AI development that promotes a broad distribution of opportunities can help offset some of these risks.

AI Copyright & Intellectual Property Right Risks: It is important to be aware of the potential risks of AI-generated content. For instance, AI-generated content could infringe on someone’s intellectual property (IP) rights by reproducing copyrighted material. If this occurs and the content is used, someone may be held liable, forced to remove the content, and potentially required to compensate the copyright or trademark holder. Generative AI platforms are trained on data lakes and question snippets — billions of parameters that are constructed by software processing huge archives of images and text. The AI platforms recover patterns and relationships, which they then use to create rules, and then make judgments and predictions, when responding to a


100. <https://www.weforum.org/reports/the-future-of-jobs-report-2020/in-full/executive-summary>

101. <https://news.mit.edu/2020/artificial-intelligence-ai-carbon-footprint-0423>

102. <https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf>

103. <https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4>

104. https://www.washingtonpost.com/technology/2023/06/05/chatgpt-hidden-cost-gpu-compute/?utm_campaign=wp_post_most&utm_medium=email&utm_source=newsletter&wpisrc=nl_most



The rise of AI-driven autonomous weaponry also raises concerns about the dangers of rogue states or non-state actors using this technology — especially when we consider the potential loss of human control in critical decision-making processes.

prompt. This process comes with legal risks, including intellectual property infringement. In many cases, it also poses legal questions that are still being resolved. For example, does copyright, patent, trademark infringement apply to AI creations? Is it clear who owns the content that generative AI platforms create for you, or your customers? Before businesses can embrace the benefits of generative AI, they need to understand the risks¹⁰⁵ — and how to protect themselves

AI Competition Arms Race: The risk of countries engaging in an AI arms race could lead to the rapid development of AI technologies with potentially harmful consequences. As referenced above, a March 22, 2023, letter¹⁰⁶ from the Future of Life Institute signed by more than 1,000 tech leaders, researchers and others, including Apple co-founder Steve Wozniak, have urged intelligence labs to pause the development of advanced AI systems. The letter states that AI tools present “profound risks to society and humanity.

Risks from Autonomous Weapon Systems and Military AI: The rise of AI-driven autonomous weaponry also raises concerns about the dangers of rogue states or non-state actors using this technology — especially when we consider the potential loss of human control in critical decision-making processes. As stated¹⁰⁷ by the Effective Altruism Forum, the use and proliferation of autonomous weapon systems appears likely in the near future, but the risks of AI-enabled warfare are under-studied and under-funded. Autonomous weapons and military applications of AI more broadly (such as early-warning and decision-support systems) have the potential to increase the risk factors for a variety of issues, including great power war, nuclear stability, and AI safety. Several of these issues are potential pathways towards existential and global catastrophic risks. Autonomy in weapon systems therefore affects both the long-term future of the world and the lives of billions of people today.

The Psychological Toll of AI: Erosion of Empathy and Trust: Increasing reliance on AI-driven communication and interactions could lead to

diminished empathy, social skills, and human connections. As AI becomes more sophisticated, it is increasingly difficult to differentiate between human and machine interactions. This can lead to a sense of distrust and cynicism¹⁰⁸ in our relationships with others. To preserve the essence of our social nature, we must strive to maintain a balance between technology and human interaction.

Disinformation and Threat to Democracy: AI-generated content, such as deepfakes, contributes to the spread of false information and the manipulation of public opinion. Efforts to detect and combat AI-generated misinformation are critical in preserving the integrity of information in the digital age. As outlined in Stanford University’s “One Hundred Year Study on Artificial Intelligence (AI100)¹⁰⁹, AI systems are being used in the service of disinformation on the internet, giving them the potential to become a threat to democracy and a tool for fascism. From deepfake videos to online bots manipulating public discourse by feigning consensus and spreading fake news, there is the

105. <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>

106. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

107. <https://forum.effectivealtruism.org/posts/RKMNZn7r6cT2Yaorf/risks-from-autonomous-weapon-systems-and-military-ai>

108. <https://www.finextra.com/blogposting/24174/the-psychological-toll-of-ai-erosion-of-empathy-and-trust-in-our-relationships>

109. <https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=277a52db2706>



danger of AI systems undermining social trust. The technology can be co-opted by criminals, rogue states, ideological extremists, or simply special interest groups, to manipulate people for economic gain or political advantage. Disinformation poses serious threats to society, as it effectively changes and manipulates evidence to create social feedback loops that undermine any sense of objective truth. The debates about what is real quickly evolve into debates about who gets to decide what is real, resulting in renegotiations of power structures that often serve entrenched interests.

AI Risks to Productivity and Attention: We are being told that AI is making coders and customer service representatives and writers more productive. However, there is a risk in measuring AI's potential

benefits without also considering its likely costs. One risk is that these systems will do more to distract and entertain than to focus, such as bringing about an information overload: Marcela Martin, BuzzFeed's president, encapsulated this concern when she told investors¹¹⁰ in May 2023 that, "Instead of generating 10 ideas in a minute, AI can generate hundreds of ideas in a second." On the surface, that can appear as a good thing, but applied across the economy, someone somewhere will have to process all that information. What will this do for productivity? One lesson of the digital age is that more is not always better. More emails and more reports and more Slacks and more tweets and more videos and more news articles and more slide decks and more Zoom calls have not always led to more great ideas. Gloria Mark, a professor of information science at the University of California, Irvine, and the author of "Attention Span¹¹¹", cautioned¹¹² that while we can produce more information, that means there's more information for humans to process. "Our processing capability is the bottleneck."

AI & Unintended Consequences: AI systems, due to their complexity and lack of human oversight, might exhibit unexpected behaviors or make decisions with unforeseen consequences. This unpredictability can result in outcomes that negatively impact individuals, businesses, or society as a whole. Robust testing, validation, and monitoring¹¹³ processes can help developers and researchers identify and fix these types of issues before they escalate.

Existential Risks: The development of artificial general intelligence (AGI)¹¹⁴ or Artificial Super Intelligence (ASI)¹¹⁵ that surpasses human intelligence raises long-term concerns for humanity. The prospect of AGI could lead to unintended and potentially catastrophic consequences, as these advanced AI systems may not be aligned with human values or priorities. According to a January 26, 2023, OECD report¹¹⁶, It is not unrealistic to expect recursively self-improving AI to arrive soon. Breakthroughs in AI have been coming in rapid succession with AlphaGo¹¹⁷, GPT3¹¹⁸, Gato¹¹⁹, Dall-E2¹²⁰, AlphaCode¹²¹, and others. And at least 72 projects¹²² with the explicit goal of creating AGI have been set up and funded, including DeepMind with 1300 staff. Despite rapid progress, some doubt¹²³ that we are headed for AGI or will ever be able to invent such a technology or that it will pose an existential threat¹²⁴. Experts often argue that we do not understand the human brain and the role of consciousness in intelligence adequately enough.

To mitigate these risks, the AI research community needs to actively engage in safety research, collaborate on ethical guidelines, and promote transparency in AGI development. Ensuring that AGI serves the best interests of humanity and does not pose a threat to our existence is paramount. ■

110. <https://futurism.com/buzzfeed-ai-replace-content>

111. <https://www.harpercollins.com/products/attention-span-gloria-mark?variant=40346590117922>

112. <https://www.nytimes.com/2023/05/28/opinion/artificial-intelligence-thinking-minds-concentration.html>

113. <https://hbr.org/2021/05/5-rules-to-manage-ais-unintended-consequences>

114. https://en.wikipedia.org/wiki/Artificial_general_intelligence

115. <https://www.geeksforgeeks.org/what-is-artificial-super-intelligence-asi/>

116. <https://oecd.ai/en/wonk/existential-threat>

117. <https://www.deepmind.com/research/highlighted-research/alphago>

118. <https://openai.com/product>

119. <https://www.deepmind.com/publications/a-generalist-agent>

120. <https://openai.com/dall-e-2>

121. <https://www.deepmind.com/blog/competitive-programming-with-alphacode>

122. https://gcrinstitute.org/papers/055_agi-2020.pdf

123. <https://scottaaronson.blog/?p=6524>

124. <https://aiimpacts.org/list-of-sources-arguing-against-existential-risk-from-ai/>

CONCLUSION

Printed books made it possible for scholars to broaden larger fields of knowledge than had ever before been possible. In that there is an obvious analogy for AI, and especially the new Large Language Models, which trained on a given corpus of knowledge can derive all manner of things from it. The new AI models stand to transform humans' relationship¹²⁵ with computers, knowledge and even with themselves. Everyone and everything now seem to be pursuing such fine-tuned models as ways of providing access to knowledge. AI has the potential to solve some of humankind's biggest problems by developing new drugs, designing new materials to help fight climate change, and even untangling¹²⁶ the complexities of fusion power.

Whether the promise of AI will deliver on its potential depends on how well businesses, AI developers, governments and other stakeholders manage perceived risks while fostering a regulatory environment that encourages innovation, best practices, voluntary standards and international collaboration.

The recent acceleration in both the power and visibility of AI systems, and growing awareness of their abilities and defects, have raised fears that the technology is now advancing so quickly that it cannot be safely controlled. However, exaggerated and misleading concerns about the AI tools' potential to cause harm has crowded out reasonable discussion about the technology, generating a familiar, yet unfortunate, "tech panic."

The fear that machines will steal jobs is centuries old. But so far new technology has created new jobs to replace the ones it has destroyed. Machines tend to be able to perform some tasks, not others, increasing demand for people who can do the jobs machines cannot. The degree of existential risk posed by AI is hotly debated, but all involve a huge amount of guesswork, and a leap from today's technology. Even experts tend to overstate the risks in their focus areas, compared with other forecasters.

Imposing heavy regulation, or indeed a pause, today is manifestly an overreaction; when regulation is warranted, it is for more earthly reasons than saving humanity. As explained above, existing AI systems raise real concerns about bias, privacy and intellectual-property rights, and as the technology continues to advance, other problems could still become apparent. Nevertheless, the key is to balance the promise of AI with a sound judgment of the actual risks, and to be ready to adapt accordingly.

The fears around new technologies follow a predictable trajectory known as "the Tech Panic Cycle¹²⁷." Fears increase, peak, then decline over time as the public becomes familiar with the technology and its benefits. Indeed, other previous "generative" technologies in the creative sector such as the printing press, the phonograph, and the Cinématographe followed a similar course. But unlike today, policymakers then were unlikely to do much to regulate and restrict these technologies. As the panic over new AI innovations, such as generative AI, enters its most volatile stage, policymakers should recognize the predictable cycle we are in, and proceed cautiously regarding any regulatory efforts so as not to harm AI innovation.

Policymakers should therefore consider a more innovative approach to regulation: Regulatory sandboxes for AI is an essential tool to regulate AI without compromising on innovation and can save policymakers considerable time and resources by informing decisions about whether to change or re-interpret legal frameworks, as well as aiding businesses by reducing the time and capital required to enter the market. The controlled environment of the sandbox approach should be a first immediate and mandatory step for countries to test and experiment with new technologies, business models, and regulatory approaches. ■

Anders Halvorsen | ahalvorsen@witsa.org

125. <https://www.economist.com/essay/2023/04/20/how-ai-could-change-computing-culture-and-the-course-of-history>

126. <https://www.wired.com/story/deepmind-ai-nuclear-fusion/>

127. <https://datainnovation.org/2023/05/tech-panics-generative-ai-and-regulatory-caution/>

WITSA

8300 Boone Boulevard
Suite 450
Vienna VA 22182
United States of America

witsa.org | admin@witsa.org

