# Artificial General Intelligence

Introduction

Artificial intelligence has emerged as a strategic domain shaping the future of global politics, economics, and security. As countries vie for AI dominance, they recognize that AI technologies, such as GPTs and AI agents, have the potential to revolutionize not only industries and markets but also military capabilities and strategies. In this high-stakes contest, the United States and China are leading the pack, with Europe and Russia striving to catch up.

This briefing provides an in-depth examination of the current landscape of the AI arms race, highlighting the key players, their objectives, and the challenges they face in harnessing the power of AI technologies. We will explore the impact of AI on defense systems and military strategies, with a particular focus on AI agents and the development of lethal autonomous weapons systems (LAWS). Additionally, we will discuss the pressing ethical and security concerns arising from the integration of AI into military operations and the role of international cooperation in addressing these complex issues.

The journey towards Artificial Intelligence: A retrospective

The discovery of Artificial Intelligence began in the 1950s, with visionaries like Alan Turing laying the groundwork for intelligent machines. Turing's 1950 paper, "Computing Machinery and Intelligence," introduced the concept of machine intelligence and proposed the Turing Test as a criterion for evaluating whether a machine could exhibit human-like intelligence.

In 1957, Frank Rosenblatt introduced perceptrons, marking a significant milestone in AI history. Perceptrons, the earliest form of artificial neural networks, showed promise in simple pattern recognition tasks. However, they had limitations in solving certain problems, such as the XOR problem. The XOR problem is a binary classification task where inputs are either both true or both false, and perceptrons could not solve it due to their linear nature. Marvin Minsky and Seymour Papert exposed these limitations in their 1969 book "Perceptrons," contributing to a shift in AI research focus.
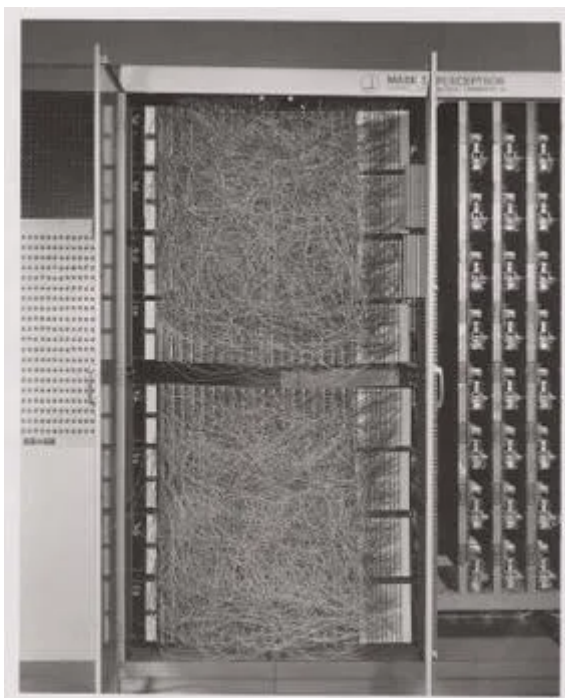
Mark I Perceptron machine, the first implementation of the perceptron algorithm

John McCarthy's development of the LISP programming language in 1958 provided a foundation for early rule-based AI systems. LISP became the dominant programming language for AI research and enabled the construction of early expert systems and other AI applications.
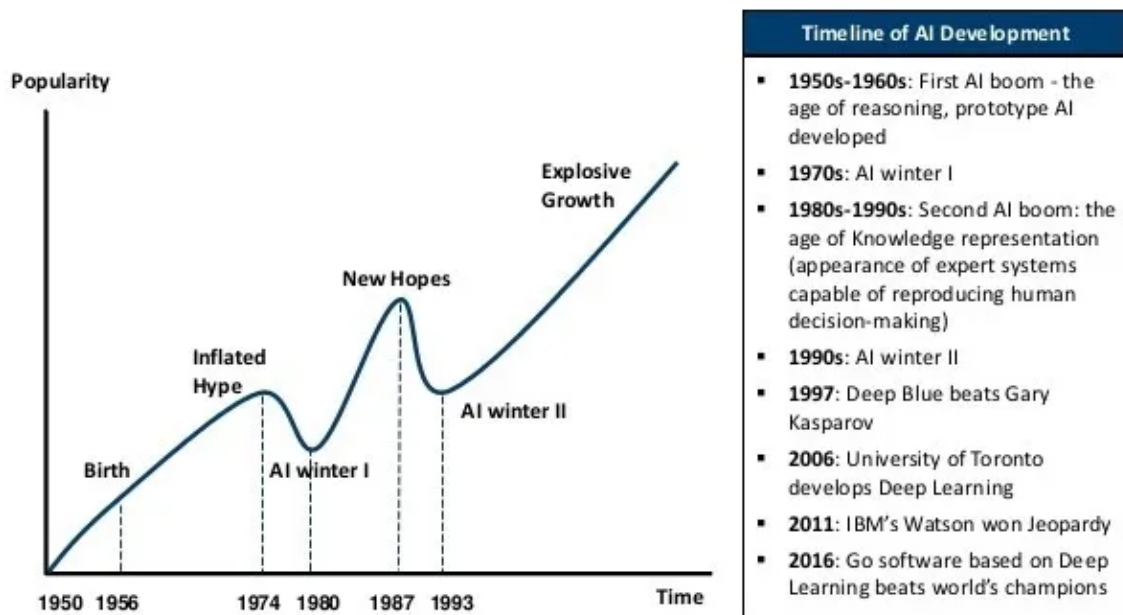
During the 1980s, decision trees and the Prolog programming language gained prominence in AI research. Decision trees, a type of algorithm that uses a tree-like structure for decision-making, found applications in areas such as medical diagnosis and credit risk assessment. Prolog, a logic programming language, facilitated the development of knowledge-based systems and expert systems, driving AI research in a new direction.

However, AI research experienced two notable downturns, known as AI winters, in 1974-1980 and 1987-1993. The first AI winter was triggered by a combination of overly optimistic expectations, limited computational power, and a lack of funding, primarily due to the critique of perceptrons by Minsky and Papert. The second AI winter was caused by the failure of expert systems to deliver on their promises, as well as the burgeoning interest in alternative technologies such as personal computing.

The resurgence of AI began with the rekindling of interest in neural networks and the development of the backpropagation algorithm, a learning technique that enabled more efficient training of neural networks. Geoff Hinton, a pivotal figure in AI research, made a significant breakthrough with his team's victory in the ImageNet Challenge in 2012. By leveraging deep learning techniques, Hinton's team achieved unprecedented accuracy in image recognition, setting the stage for the AI renaissance.

The surge in computational power, fueled by the increasing accessibility of Graphics Processing Units (GPUs), has enabled the development of AI agents with a wide range of applications. GPUs have proven instrumental in training deep learning

AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING"...

Timeline of AI Development

- **1950s-1960s:** First AI boom - the age of reasoning, prototype AI developed
- **1970s:** AI winter I
- **1980s-1990s:** Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s:** AI winter II
- **1997:** Deep Blue beats Gary Kasparov
- **2006:** University of Toronto develops Deep Learning
- **2011:** IBM's Watson won Jeopardy
- **2016:** Go software based on Deep Learning beats world's champions

The Boom and Bust Cycle of AI Research [Source]

The Emergence of Transformer Networks and Generative Pretrained Transformers: A New Era in Artificial Intelligence

A remarkable shift has occurred in the realm of artificial intelligence (AI), with the advent of transformer networks and the Generative Pre-trained Transformer (GPT) series. These sophisticated innovations have redefined the landscape of natural language processing (NLP) and opened up a plethora of applications in various domains.

The 2017 paper "Attention is All You Need" by Vaswani and colleagues heralded a new era in deep learning architectures. Transformers, designed to manage sequential data, employ self-attention mechanisms to discern the significance of different input elements. This approach has proven adept at identifyingransfor long-range dependencies and intricate relationships within data, paving the way for substantial progress in NLP tasks, such as machine translation, summarisation, and sentiment analysis.

Developed by OpenAI, the GPT series leverages the potential of transformer networks to produce unparalleled language models. GPT-1's 2018 debut

demonstrated impressive performance across NLP tasks without task-specific training. The subsequent releases of GPT-2 in 2019 and GPT-3 in 2020 delivered even greater language comprehension and generation. Boasting 175 billion parameters, GPT-3 is capable of producing coherent, contextually relevant text, answering questions, and writing code--- cementing its position as one of the most versatile AI models in existence.

When OpenAI released its web based interface to GPT-3, called 'ChatGPT', it reached a user base of one hundred million in one week, less time than any other application in human history, including such luminaries as Facebook, Instagram and Snapchat.
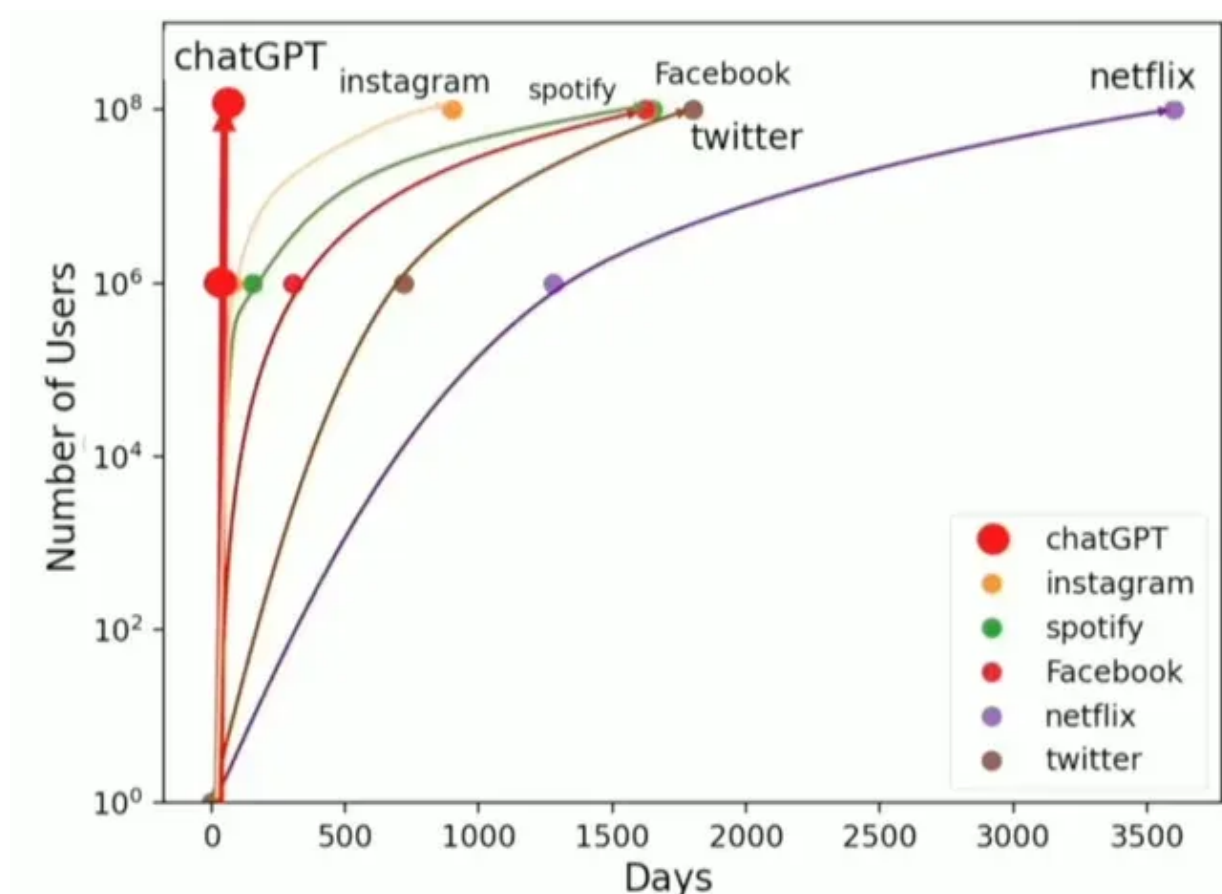
Number of Days to 1M and 100M Users by Technology [Source]

The rise of transformer networks and GPT has led to a myriad of practical applications across industries Microsoft has incorporated GPT into applications like Excel and PowerPoint, enhancing their natural language capabilities, while GitHub Copilot utilises GPT for code generation, boosting developer productivity and code quality

It isn't only technology companies that have gotten into the race. Many open source large language models, datasets and tools are now available from web sites such as Hugging Face and Papers Wiih Code. Many of the models provided by these sites share similar levels of performance to the latest commercial offerings from the likes of Microsoft, Meta and Google.
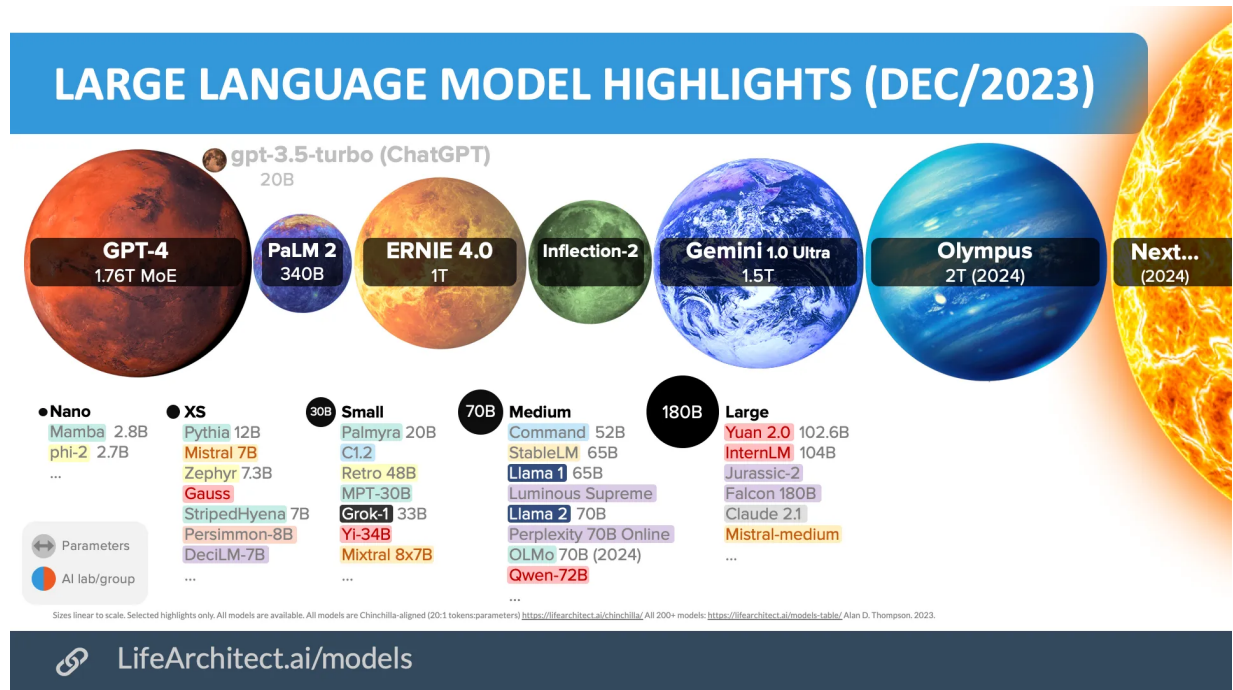
Language Model Sizes [Source]

Use cases among market sectors

Finance

AI agents have significantly reshaped the financial sector, transforming trading, risk assessment, and fraud detection. Companies like Kensho harness AI to analyze vast amounts of market data and generate valuable insights, radically altering investment strategies and providing data-driven decision-making tools for portfolio management. For instance, in 2017, JPMorgan Chase introduced COIN (Contract Intelligence), an AI tool that can analyze legal documents in seconds, saving over 360,000 hours of human labor annually. Additionally, a report by Autonomous Research estimates that AI technologies could help reduce operational costs in the financial

services industry by 22%, or roughly $1 trillion by 2030. Bloomberg Inc. has announced its own large language model devoted to finance, Bloomberg GPT.

Cryptocurrencies

AI-driven tools are being employed to analyze and predict cryptocurrency trends, enabling informed investment decisions in a highly volatile market. Platforms such as CoinPredictor and CryptoSage offer AI-powered analytics to assist investors in navigating the complex world of digital assets. For example, a study by the University of Cambridge found that 84% of surveyed cryptocurrency exchanges and wallets use AI-based systems for their trading and investment activities. These AI-driven tools have been credited with a 20% reduction in prediction errors compared to traditional methods. One suh application is *Satoshi*. Satoshi is an application created by startup FalconX in San Francisco that acts like an investment advisor for those punters wishing to invest in this area. The developers are planning to incorporate large language models to allow the application to provide timely and accurate investmment information to users.

Healthcare

AI agents are making strides in diagnostics, drug discovery, and personalized medicine. PathAI, for example, employs AI to improve pathology diagnoses, enhancing the accuracy and efficiency of disease detection. By implementing AI, PathAI claims to have reduced diagnostic errors by up to 50%. Similarly, DeepMind's AlphaFold has made groundbreaking advancements in protein folding prediction, which could revolutionize drug development and disease understanding. A Nature article reported that AlphaFold's accuracy surpassed that of other methods by a significant margin, achieving a median Global Distance Test (GDT) score of 92.4 out of 100.

Customer Service

AI-powered chatbots, such as those deployed by IBM Watson, have revolutionized customer support by providing instant, personalized assistance, improving customer satisfaction and operational efficiency. AI-driven virtual assistants can now handle a wide range of queries, freeing up human agents to tackle more complex tasks. According to a study by Juniper

Research, AI-powered chatbots are expected to save businesses approximately $8 billion per year by 2022, up from $20 million in 2017, and reduce customer service response times by up to 80%.

Supply Chain Management

AI agents optimize inventory management, demand forecasting, and logistics. Companies like ClearMetal leverage AI to enhance supply chain visibility and streamline operations, reducing costs and improving efficiency across the entire supply chain. In a case study, ClearMetal reported that one of its clients, a global retailer, achieved a 40% reduction in inventory holding costs, a 10% reduction in transportation costs, and improved their on-time delivery rate by 20%. This illustrates the transformative potential of AI in supply chain management, as it enables companies to make data-driven decisions, increase responsiveness to market fluctuations, and achieve a competitive advantage. Furthermore, AI-powered tools can help organizations identify and mitigate risks in the supply chain, ensuring business continuity and resilience in the face of unforeseen disruptions, such as natural disasters or global pandemics. Ultimately, the integration of GPT and other AI technologies into supply chain management can revolutionize the way businesses operate, driving growth and fostering sustainable practices.

The Labor Conundrum: GPT's Impact on the labor market

The rise of GPT and advanced AI models, including the potential development of Artificial General Intelligence (AGI), has sparked concerns about their influence on the labor market, with particular attention given to the software development industry. A recent Goldman Sachs report has contributed to the discussion, examining the industries and job roles potentially disrupted by GPT and its ilk, highlighting the significant labor market implications.
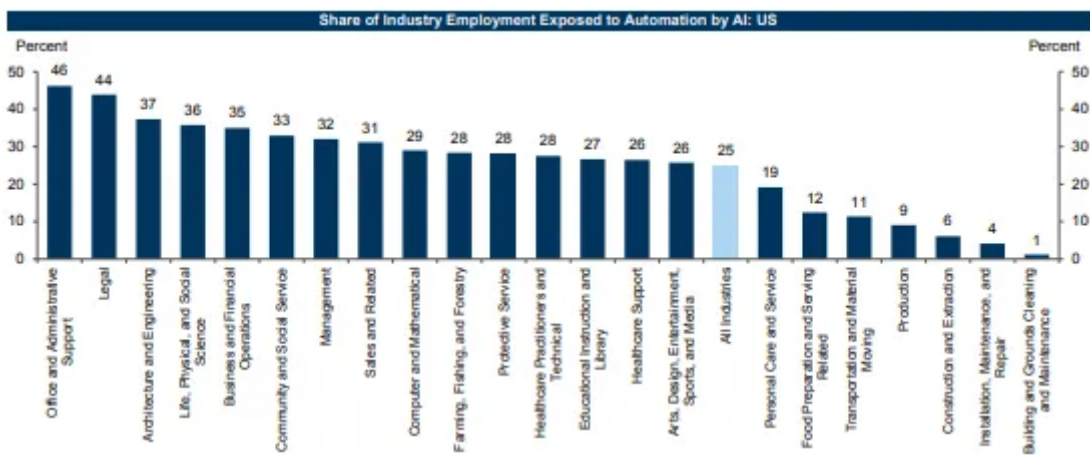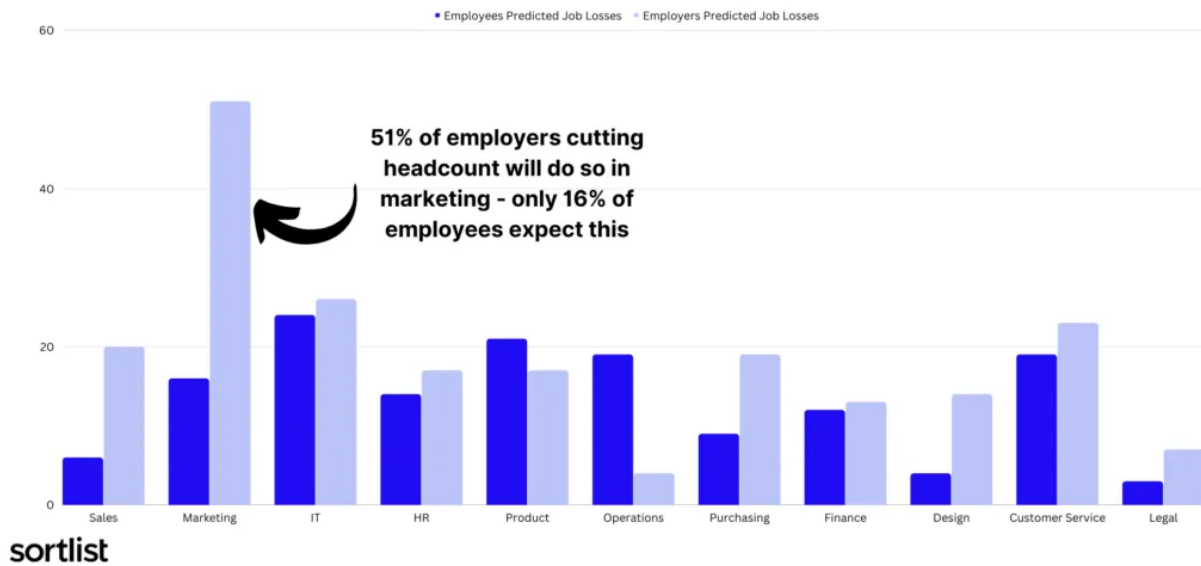
Share of Industry Employment Expose to Automation by AI: US

Goldman Sachs' analysis reveals the finance and customer service sectors as particularly susceptible to job displacement due to GPT. However, the software development industry is also facing considerable changes. The report estimates that the growing capabilities of AI-powered tools like GPT-driven code generation and natural language processing may transform the way developers work, streamlining the coding process and potentially reducing the need for entry-level programming roles.

Examples of users coding applications and websites using GPT models include GitHub's Copilot, an AI-powered code completion tool that assists developers by automatically generating code snippets based on context and natural language descriptions. Similarly, OpenAI's Codex, the model behind Copilot, has demonstrated the ability to generate functional code based on user prompts. As these technologies mature, they may further automate certain aspects of software development, enabling developers to focus on more complex, creative, and strategic tasks.

The precise effect of GPT and AGI on software developers remains nebulous, yet their increasing prevalence across various sectors serves as a stark reminder of the importance of proactively addressing potential labor market repercussions of AI and automation. Policymakers, businesses, and educational institutions must unite to equip the workforce for this transition, investing in upskilling and reskilling initiatives and fostering a culture of lifelong learning.

Concern about job loss vs planned cuts (% responses by department)

- Employees Predicted Job Losses
- Employers Predicted Job Losses

51% of employers cutting headcount will do so in marketing - only 16% of employees expect this

sortlist

Concern about job loss vs planned cuts [Source]

While GPT and AGI-driven tools may automate specific tasks, they are also expected to create new job opportunities in creative, problem-solving, and technically advanced fields. As the nature of software development evolves, demand for skills such as AI integration, ethical AI development, and advanced algorithm design will likely grow. Mitigating job losses hinges on adopting strategies to upskill and reskill affected workers, ensuring their preparedness for new roles in the evolving labor market.

The Double-Edged Sword: Biases and Misinformation in the Age of GPT

The ascent of GPT and its sophisticated language generation capabilities have brought forth a host of concerns, including potential biases and the abuse of misinformation. This analysis delves into the complex interplay of these issues, providing concrete examples to illustrate the inherent challenges and risks.

A primary concern with GPT models is the manifestation of biases, often rooted in the training data. Since these AI models learn from large datasets containing text from the internet, they inadvertently absorb and perpetuate

existing biases. For instance, OpenAI's GPT-3, released in June 2020, attracted criticism for generating sexist and racist content, mirroring the biases found in its training data.

To counter this issue, OpenAI has acknowledged the importance of research focused on reducing biases and has initiated efforts to develop improved guidelines for its human reviewers. Nonetheless, striking the right balance between reducing biases and maintaining model performance remains an ongoing challenge.

The potential misuse of GPT for spreading misinformation is another significant concern. A prime example is the website "ThisPersonDoesNotExist.com," which uses GPT's sister model, StyleGAN, to generate highly realistic images of non-existent individuals. Launched in February 2019, the site quickly garnered attention, raising concerns about the potential abuse of such technology for nefarious purposes, such as creating deepfake images and videos.

In a study conducted in September 2020 by researchers at Georgetown University, GPT-3 was found to be remarkably adept at producing disinformation. The AI model was given several prompts, including conspiracy theories and false news, and it generated highly convincing responses, posing a risk for the spread of misinformation.

A disquieting example of GPT's abuse for political misinformation surfaced in February 2021 when the Lithuanian news site "15min.lt" reported that a deepfake article, supposedly written by a well-known politician, had been published. The fraudulent piece, generated using a GPT-like model, called for the withdrawal of Lithuania from NATO and the European Union.

The growing presence of GPT models and their potential misuse necessitates the development of robust countermeasures. Initiatives such as the "Deepfake Detection Challenge," launched by Facebook in September 2019, aim to encourage the development of technologies capable of detecting and mitigating deepfake content.

Navigating the Policy Maze: GPTs and AI Regulation

As the prominence of GPTs and other AI technologies burgeons, governments across the globe find themselves confronting the intricate task

of devising policies and regulations that address the ethical, social, and economic ramifications. The convoluted nature of these matters, amplified by the swift evolution of technology, has produced a diverse array of policy reactions, oscillating between stringent regulation and laissez-faire approaches.

Data Privacy and Security

The European Union's trailblazing General Data Protection Regulation (GDPR), which came into effect in 2018, embodies a robust attempt to safeguard individual privacy in the digital age. GPT models, which rely on vast amounts of data for training, must adhere to GDPR's stringent provisions on data collection, usage, and storage. Some critics argue that GDPR's rigidity may stifle innovation, while others assert that a robust privacy framework is vital in the age of AI.

Algorithmic Accountability and Transparency

The black-box nature of AI models like GPTs poses challenges to algorithmic accountability and transparency. In response, the EU has proposed the Artificial Intelligence Act (AIA), which, if enacted, would require companies to ensure that AI systems are transparent, traceable, and accountable. The AIA could mandate extensive documentation and disclosure, compelling developers to unveil the inner workings of their AI models. Critics warn of potential trade-offs between transparency and intellectual property protection, while proponents argue that accountability is crucial to prevent AI misuse.

Bias and Discrimination

AI models such as GPTs can inadvertently perpetuate biases present in their training data, leading to unfair and discriminatory outcomes. In the United States, the Algorithmic Accountability Act (AAA), proposed in 2019, sought to tackle this issue by mandating that companies perform impact assessments on high-risk automated decision systems. Though the AAA has not been enacted, it highlights the growing awareness of the need to address AI-driven bias and discrimination.

Misinformation and Content Moderation

The potential of GPTs to generate high-quality text raises concerns about the proliferation of misinformation and deepfake content. Governments have been considering policy interventions to combat AI-generated falsehoods, such as the US Deep Fakes Accountability Act Go(2019), which aims to enforce transparency and accountability for deepfake content creators. Balancing content moderation and free speech remains a contentious issue, with some arguing that heavy-handed regulation may infringe upon freedom of expression.

World Governments' Legislative and Regulatory Responses to GPT

As GPT and other advanced AI models gain prominence, governments worldwide are devising legislative and regulatory responses to address ethical, social, and economic concerns raised by these technologies. The following is an overview of the responses in various jurisdictions:

- **United States**: The US has primarily focused on sector-specific regulations and guidelines. In 2019, the Algorithmic Accountability Act was proposed, which would require companies to perform impact assessments on high-risk automated decision systems, including AI models like GPT. The US has also introduced the Deep Fakes Accountability Act, aimed at enforcing transparency and accountability for creators of deepfake content, which could be generated by GPT-like models.

- **Canada**: Canada has been proactive in establishing an AI ecosystem that fosters innovation while addressing ethical concerns. In 2017, the Canadian government launched the Pan-Canadian Artificial Intelligence Strategy to promote AI research and development. Additionally, the Canadian government has been working on the Digital Charter, a framework designed to protect citizens' data privacy and ensure responsible AI development.

- **European Union**: The EU has been at the forefront of AI regulation, with the General Data Protection Regulation (GDPR) serving as a foundation
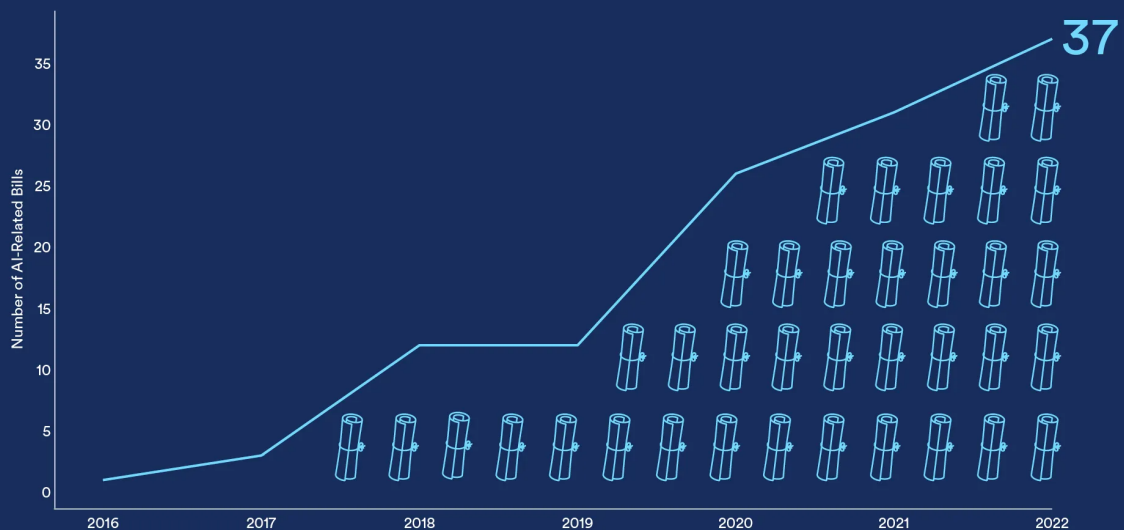
for data privacy in the digital age. GPT models, which rely on vast amounts of data for training, must adhere to GDPR's provisions on data collection, usage, and storage. The EU has also proposed the Artificial Intelligence Act (AIA), which aims to establish a legal framework for AI, ensuring that AI systems are transparent, traceable, and accountable.

- **EU Member States**: Individual EU member states have been developing their AI strategies and regulations. For example, France has introduced a national AI strategy that focuses on research, development, and ethical AI deployment. Germany, too, has unveiled a national AI strategy with a focus on innovation and ethical AI development. Italy, citing the GDPR, has banned the use of OpenAI's ChatGPT application within its boundaries.

- **China**: China's AI development has been driven by its national AI strategy, announced in 2017, which aims to achieve global AI supremacy by 2030. While China has not enacted comprehensive AI regulations, it has established AI ethics guidelines, such as the Beijing AI Principles, which emphasize the responsible and ethical use of AI technologies.

- **Japan**: Japan has adopted a proactive approach to AI regulation, focusing on fostering innovation while ensuring ethical AI development. In 2019, Japan released the AI R&D Guidelines, which outline principles for AI research and development, emphasizing transparency, accountability, and user privacy.

Governments worldwide are grappling with the complex task of regulating AI technologies like GPT while striking a balance between fostering innovation and ensuring public safety and individual rights. As AI continues to advance, it is crucial for policymakers to monitor developments and adapt regulations accordingly to mitigate risks and harness the potential of these technologies.

The policy landscape surrounding GPTs and AI technologies remains fraught with challenges and conflicting interests. Striking the right balance between fostering innovation, ensuring public safety, and preserving individual rights will be vital to shaping the future of AI development and deployment.

Number of AI-Related Bills Passed into Law Globally [Source] [Data]

The Great AI Arms Race: A Global Contest for Supremacy

The rapid advancement of AI technologies, including GPTs and other sophisticated models, has ignited a global race for AI dominance. Nations are vying for supremacy in this strategic domain, investing heavily in research, development, and talent acquisition. The impact of this technological contest transcends economic and scientific realms, extending to geopolitical and military implications.

US and China: Front-runners in the AI Marathon

The United States and China have emerged as the primary contenders in the AI arms race. A 2020 report by the Global AI Index revealed that the US holds the top position in AI research and development, accounting for 33% of global AI capabilities. In contrast, China, the closest competitor, accounts for

17%. The US boasts a robust AI ecosystem, with tech giants like Google, Microsoft, and IBM leading the charge, while China's national AI strategy (announced in 2017) aims to achieve global AI supremacy by 2030.

European Union: Striving for Sovereignty

The European Union, recognizing the strategic importance of AI, has sought to establish itself as a formidable player in this domain. In April 2021, the European Commission proposed a comprehensive AI regulatory framework, emphasizing human-centric and ethical AI development. Furthermore, the EU's €1 billion investment in the Human Brain Project (2013-2023) and its recent commitment of €140 billion to the Digital Europe Programme (2021-2027) demonstrate its resolve to bolster AI capabilities and foster innovation.

Russia: Military AI and Autonomy

Russia, though lagging in the broader AI race, has been actively pursuing AI technologies for military applications. President Vladimir Putin's assertion in 2017 that "whoever becomes the leader in this sphere [AI] will become the ruler of the world" underscores the nation's strategic interest in AI. Russia's military modernization efforts include the development of autonomous combat vehicles, such as the Uran-9, and investments in AI-driven decision-making systems.

The AI Defense Race: A Double-edged Sword

The pursuit of AI supremacy has not been confined to the civilian sphere. Nations are increasingly exploring AI's potential for defense and military applications, raising the stakes in this technological race. AI agents, which are autonomous software programs capable of learning, adapting, and making decisions, are a key component of these military advancements.

In February 2020, the US Department of Defense (DoD) released its AI strategy, outlining plans to integrate AI capabilities, including AI agents, into military operations across all domains, from intelligence and surveillance to logistics and battlefield decision-making. The Pentagon's Joint Artificial

Intelligence Center (JAIC) serves as the focal point for these efforts, with an estimated budget of over $800 million in 2021.

Conversely, China's military-civil fusion policy seeks to harness AI advancements, including AI agents, for national defense. The Chinese government has established the Central Military Commission's Science and Technology Commission to spearhead these efforts, focusing on the integration of AI into military applications, such as autonomous drones, robots, and command and control systems. In 2017, China announced its plan to establish an AI innovation center for defense, aiming to strengthen its military capabilities and ensure national security.

How the Chinese & US government compare on AI spending

in billions

- China (lower bound estimate), 2018
- China (upper bound estimate), 2018
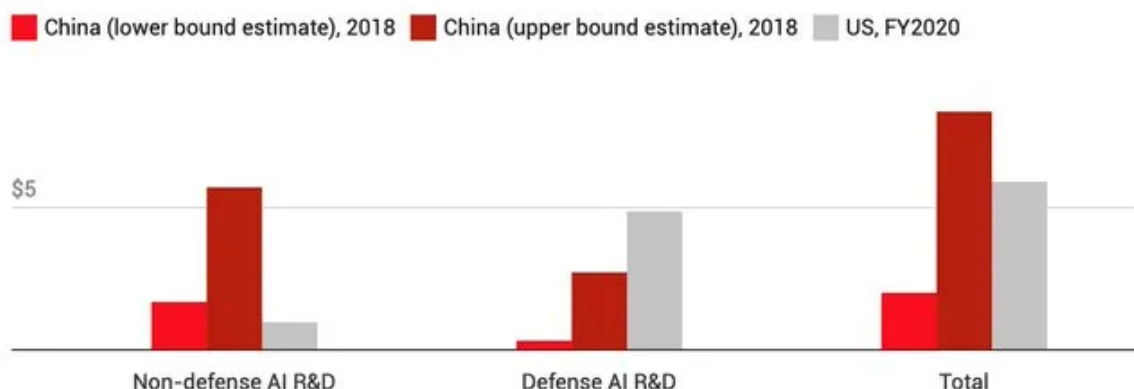- US, FY2020

Non-defense AI R&D · Defense AI R&D · Total

Chart: MIT Tech Review
· Source: China data from Center for Security and Emerging Technology; US data from US government & Bloomberg Government
· Created with Datawrapper

How the Chinese & US government compare on AI spending [Source]

Other countries are also active in the AI defense race. Russia, for instance, has invested in the development of autonomous combat vehicles, such as the Uran-9, and AI-driven decision-making systems for military operations. The United Kingdom, as part of its Integrated Review of Security, Defense, Development, and Foreign Policy, has committed to investing £6.6 billion in advanced defense research and development, including AI agents and autonomous systems, over the next four years.

The incorporation of AI into defense systems and military strategies raises ethical and security concerns, especially regarding the development of lethal autonomous weapons systems (LAWS). These systems, which rely on agents

to identify and engage targets without human intervention, have sparked a heated debate over their implications for international law, accountability, and the risk of unintended escalation.

In 2015, thousands of AI researchers and experts signed an open letter calling for a ban on LAWS, citing the potential for an AI-driven arms race and the risk of destabilizing global security. Since then, the United Nations has hosted several meetings under the Convention on Certain Conventional Weapons to discuss the challenges posed by LAWS, although no binding agreement has been reached to date.

As AI technologies continue to advance and permeate the defense domain, it is essential that policymakers, military leaders, and technology developers engage in an ongoing dialogue, striving to address the complex ethical, legal, and security challenges posed by the integration of AI into military operations.

Conclusion

GPT models, with their astounding capabilities, have the potential to transform various sectors and industries. However, this transformative power comes with a complex set of challenges, including biases, misinformation abuse, job displacement, and ethical concerns. As the development and deployment of these AI models progress, it is essential for researchers, policymakers, technology companies, and other stakeholders to collaborate and address these challenges effectively.

Firstly, responsible AI development should be prioritized, with research focusing on reducing biases and minimizing the potential for misuse. This would involve refining AI models, developing more robust datasets, and implementing transparent and accountable systems.

Secondly, policymakers must strive to balance innovation and regulation, enacting legislation that protects citizens and their rights without stifling the AI ecosystem. As the technology evolves, so too must the regulatory landscape, with governments adopting a proactive approach to policy design and implementation.

Moreover, the private sector must also contribute to fostering ethical AI development by adopting industry best practices, implementing guidelines,

and investing in AI ethics research. Collaborative efforts are crucial for overcoming the challenges posed by advanced AI models like GPT.

Lastly, educational institutions, businesses, and governments must join forces to address the potential labour market implications, investing in upskilling and reskilling initiatives to ensure a smooth workforce transition. Embracing a culture of lifelong learning will be key to adapting to the changing landscape brought about by GPT and other AI technologies.

In conclusion, while GPT models and related AI technologies present remarkable opportunities, their inherent challenges necessitate a concerted, collaborative effort from all stakeholders. By developing responsible guidelines, implementing safeguards, and fostering a culture of ethical AI development and use, we can harness the full potential of GPT models and propel humanity towards a more innovative, efficient, and equitable future.