AI: What Is It and Why Is It Good for Us and for Society?

Maxime C. Cohen[1]

September 3, 2024

1. Introduction

The term artificial intelligence (AI) was first introduced in the 1950s, but it wasn't until the launch of OpenAI's ChatGPT in November 2022 that it became a ubiquitous topic of discussion. In the last few months, not only has the number of AI companies offering wide ranges of solutions, agents, and APIs exploded, but everybody has been talking about AI and its potential impact on our society. Suddenly, most social media feeds are discussing AI-related topics; politicians, regulators, and nonprofit organizations are paying close attention to AI; C-suite executives are frequently mentioning the acronym "AI" in their speeches and earning calls; and it is now a buzzword among startup entrepreneurs and venture capitalists. At the same time, few truly know what AI is and the impact it has been having on our society and economy.

While large tech organizations have embraced AI for many years, recent advancements in the field have led to smaller firms and less tech-savvy companies beginning their AI journey. AI can serve as a clear competitive advantage by enabling companies to leverage data and algorithms to enhance efficiency and foster innovation (Iansiti and Lakhani, 2020). Many companies and organizations in various sectors have recently hired a chief AI officer (CAIO) or an equivalent position. Recent examples (all from 2024) include NASA, the United Nations, Children's National Hospital, and the U.S. Department of Homeland Security. In March 2024, the White House ordered all federal agencies to name CAIOs to oversee the federal U.S. government's approaches related to AI and manage the resulting risks. The list goes on, and this strong interest in AI embraces all geographies and all sectors of the economy. Similarly, many universities around the world are now offering degrees and professional certificates in disciplines related to AI.

This paper aims to provide factual insights into the evolution of AI and how it's benefiting humanity across multiple industries and dimensions as well as a multifaceted view of the future of AI and how we can maximize the advantages of this pivotal moment.

The paper begins by offering a brief history of AI and defining the concept. It also explains how generative AI has been a game changer for the AI industry. It then presents the numerous ways that AI is benefiting humanity already today. Next, it describes the recent democratization of AI and discusses how we can ensure that this process will continue to deliver high value. It then presents an overview of the recently coined term "AI economy," which happens to be a competitive market with hundreds of players of all sizes. Finally, the paper concludes by acknowledging the risks and challenges associated with AI and asserts that through careful regulation and ongoing dialogue, the positive impacts of AI can be maximized, ultimately outweighing the drawbacks.

---

[1] Maxime C. Cohen is the Scale AI Chair Professor of Retail and Operations Management at McGill University, the chief AI officer at ELNA Medical, an AI advisor at IVADO Labs, and the director of AI at Intégral. He has served as a consultant or advisor for several companies and startups including Google AI, IBM Research, Meta, Microsoft, Oracle, and Waze (the full list can be found at www.maximecohen.com). The opinions presented in this report are solely those of the author and do not reflect the views of any affiliated or collaborating institutions. The author acknowledges financial support from Microsoft Corporation.

1.1. Brief history and definition of AI

Although AI has been a subject of research and development for many years, it has seen remarkable advancements in the past decade. This section provides a brief overview of this evolution. Before formally defining and demystifying AI, it is important to understand a pivotal event often coined the "data revolution." Data is like fuel for AI: the more data one has access to, the better AI models will typically perform.[2] It has become common to consider both the quantity and quality of data owned by a company as a pillar of its evaluation. Data is considered an asset in a way similar to land, labor, and capital. In fact, several companies (both large and small) have been using their data (or sometimes collected curated external data) as collateral for financial loans. Not only can data be monetized by selling it to third parties and generate an additional revenue stream, but the data itself is also considered an asset. Data is collected and stored at an unprecedented speed. The following figures illustrate this claim. First, the fact that 90% of the world's data was generated in the last two years is nothing short of mind-blowing. This data can take different forms, including text, photos, and videos. Second, 120 zettabytes of data were generated per day in 2023 (a zettabyte corresponds to 21 zeros).[3] For example, the giant retailer Walmart is known to collect 2.5 petabytes (15 zeros) of data from 1 million customers per hour. Similarly, YouTube processes approximately 500 hours of video uploads every minute, translating to 720,000 hours of new content each day, amounting to several petabytes of data. So, one may naturally ask, why are businesses collecting and storing all this data? And why is this revolution happening right now? One can list at least three reasons: (i) the cost of collecting and storing data has drastically reduced in recent years, so it became affordable for many companies; (ii) compute power capabilities have considerably improved and have become affordable and available at a global scale, and (iii) executives and business managers started to fully understand the power and richness hidden in data, so that the return on investment is clear. It is safe to say that data has become more accessible today compared to the last decade and that many companies in most sectors are collecting massive amounts of data on their customers, their suppliers, and all parts of their processes. This data can then be leveraged in various ways to improve firms' performances.

It is now time to formally define AI. While there exists a myriad of definitions, a common one is as follows: AI is a synergy of technologies that enable a machine to sense, act, and learn in a way that resembles human intelligence and to perform tasks such as visual perception, speech recognition, decision-making, and translation between languages. It is important to distinguish between AI and automation: AI is much more than automating through hard-coded rules; instead, it will make (and constantly adjust) decisions by analyzing data. Examples of AI applications are endless and include AI chess players, self-driving cars, automated financial investing, chatbots, personalized recommendations, and many more.

Experts generally consider that AI has undergone three significant phases.[4] The initial phase (1950s–1970s) laid the groundwork for AI, with applications in algebraic problem-solving, geometric theorem proving, and language learning. The subsequent period (1970s–1990s) saw the

[2] In the context of AI, data can be used in various ways, including for training algorithms, validation, fine-tuning, and information retrieval. Discussing this topic in greater detail is beyond the scope of this paper.

[3] https://www.forbes.com/sites/forbestechcouncil/2024/03/04/from-5mb-hard-drives-to-180-zettabytes-the-data-migration-challenge/

[4] Providing a comprehensive review of AI developments over the past decades would be excessively lengthy and fall outside the scope of this paper. Instead, this paper aims to briefly highlight some of the key milestones, particularly focusing on the last decade.

development of symbolic algorithms and expert systems, marking AI's industrialization phase. The third period, from the 1990s to today, has been characterized by industry investments and R&D by tech companies alongside academic researchers, with a shift to data-driven approaches and large-scale applications. The 2010s saw significant advancements in deep learning, driven by the proliferation of big data and increased computational power. A seminal review on deep learning can be found in LeCun et al. (2015). This era marked a transition from shallow learning models with zero or one hidden layer to deep learning models with many hidden layers. The growing accessibility of code, data, and knowledge has further accelerated AI development. During this period, Google researchers made seminal contributions to the development of AI models with a significant impact on the field. First, Google developed TensorFlow (Abadi et al., 2016), an open-source machine learning framework, which has become one of the most widely used tools in AI research and applications. TensorFlow provides a comprehensive ecosystem for building and deploying machine learning models, making it accessible to both researchers and practitioners. Second, a very influential contribution from Google is the development of the Transformer architecture. Introduced in the paper "Attention Is All You Need" by Vaswani et al. (2017), the Transformer model has revolutionized natural language processing (NLP). In fact, the letter "T" in ChatGPT stands for Transformer. Third, Google introduced BERT by Devlin et al. (2018), a groundbreaking language model based on the Transformer architecture, notable for its impressive improvement on a variety of NLP tasks (e.g., question answering, sentiment analysis). Other contributions from Google include DeepMind, Neural Machine Translation, which has significantly improved the quality of translations, and AutoML, a suite of machine learning products that allows developers to build high-quality custom models with minimal effort. Similarly, Meta has made several seminal contributions to the development of AI models. One of Meta's most notable contributions is PyTorch by Paszke et al. (2019), an open-source deep learning framework that has become a leading AI tool due to its flexibility and ease of use. Other contributions from Meta include RoBERTa (Liu et al., 2019), a model improved from BERT by optimizing the training process and using more data, OPT (Open Pretrained Transformer), which is used for a variety of language understanding tasks, Detectron2, an open-source platform for object detection and segmentation, SimCLR, and the M2M-100 model, which is a multilingual translation model that can directly translate between any pair of 100 languages. Other technology companies have also made various seminal AI contributions, including Alibaba (City Brain, ET Medical Brain), Amazon (AWS Machine Learning Services, Alexa), Apple (Siri, Core ML), Baidu (PaddlePaddle, Apollo), IBM (Watson), Microsoft (Azure AI, Turing project, Brainwave, HoloLens, Seeing AI), NVIDIA (powerful GPUs, cuDNN, TensorRT), Open AI (GPT models, DALL-E, CLIP), and Tencent (AI Lab, WeChat AI). As can been seen in Figure 1, there is a striking inflection point in 2010 with the unveiling of large-scale deep learning models.

The exponential increase in available data, which is now routinely collected through various capture modes (e.g., computers, smartphones, sensors), has enhanced AI performance.[5] Both data and compute power are critical inputs for AI systems. The data will help the model learn the right correlations and patterns, whereas the compute power capabilities make it possible to crunch large amounts of data in reasonable timeframes. A famous successful AI example is Alpha Go, which has often been considered an important milestone in terms of AI capabilities and impact. More

[5] Publication of the ImageNet database in 2009 and the subsequent competitions spurred advancements in image classification, culminating in the creation of AlexNet in 2012, which achieved near-human image recognition capabilities. Another pivotal moment was the development of GPT-2 in 2019, marking significant progress in generative AI (generative AI will be formally defined in Section 1.3).

specifically, AI revolutionized the game of Go[6] with the development of AlphaGo, an AI program created by DeepMind, a Google subsidiary.[7] In 2016, AlphaGo made headlines by defeating Lee Sedol, one of the world's top Go players, in a five-game match, a feat previously thought to be years away. This breakthrough demonstrated AI's ability to tackle and excel in tasks requiring intuition and strategic thinking, highlighting the potential of AI to solve complex problems and revolutionize various fields beyond gaming. An interesting conclusion was that even though AI was trained on human-generated game instances, it was still able to discover new, unexplored strategies beyond the typical ways that professional humans were playing. It shows that AI can potentially perform well in tasks that involve reasoning and creativity skills. Today, in the generative AI era, it is common for AI tools to perform various creative tasks, such as creating images, writing documents, and composing music.

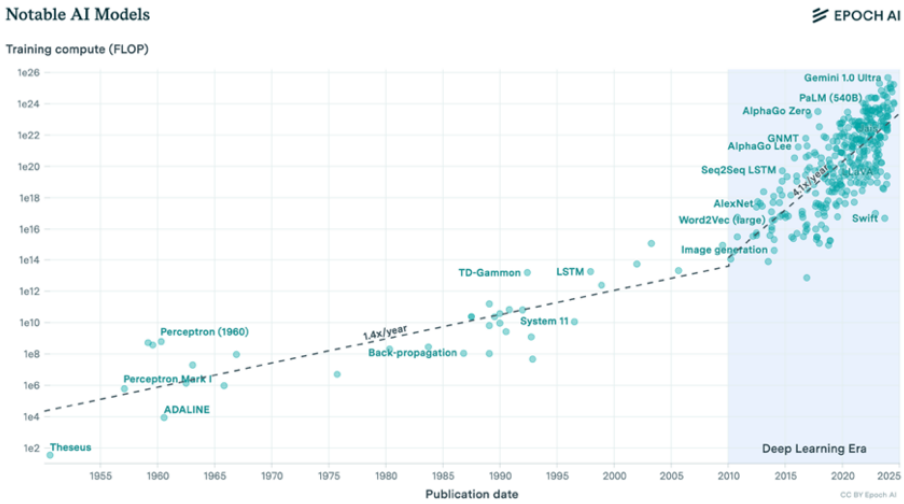

Figure 1. Evolution of notable AI models (source: https://epochai.org/data/notable-ai-models, date accessed: August 2024).

Since 2015, AI has demonstrated near-human performance across various domains and has outperformed world champions in increasingly complex games. Google's development of BERT, the first large language model (LLM), paved the way, but it was ChatGPT that truly brought AI into the public spotlight. And since then, a plethora of models, applications, and chatbot experiences have been released, and new ones are still released every week.

The recent rapid evolution and scaling of LLMs—often referred to as foundation models—has been a dominant focus of recent AI efforts. Models have quickly grown in size and scale, for example, the early GPT-2 in 2019 with 1.5 billion parameters to GPT-3 with 175 billion parameters, and beyond to models like Megatron-Turing NLG 530B. Today,[8] GPT-4 is ten times the size of its predecessor, GPT-3, with roughly 1.8 trillion parameters.[9] That said, with great size comes great computational cost and latency as well as significant carbon emissions. Stanford's AI

[6] Go is a complex, abstract strategy board game. It is considered one of the world's oldest board games and thought to have originated in China some 4,000 years ago. The game happens to have more possible moves than atoms in the universe and was considered a significant challenge for AI due to its deep strategic elements. AlphaGo utilized deep learning to analyze vast amounts of data from professional games and self-play to ultimately master the game.

[7] https://deepmind.google/technologies/alphago/

[8] Due to the rapid evolution of the LLM landscape, many models and versions quickly become outdated and are replaced by newer iterations. Therefore, all examples, references, and model names mentioned in this paper reflect the information available at the time of writing.

[9] https://www.kdnuggets.com/2023/07/gpt4-details-leaked.html

Index report concludes that the training costs and computational requirements of frontier (i.e., cutting-edge) models "have reached unprecedented levels: OpenAI's GPT-4 used an estimated $78 million worth of compute to train, while Google's Gemini Ultra cost $191 million for compute."[10] At the same time, larger models can often (yet not always) perform well across general-purpose tasks, but (i) they may be very expensive to run, as they require powerful compute clusters, (ii) they can be slow in responding to queries, (iii) they require significant time to train, and (iv) yet they can often be outperformed by smaller models trained for specialized tasks within specific verticals. As a result, the industry has shifted from focusing on rising performance at all costs to the development of smaller, cost-effective models, hence trading away some intelligence for speed and affordability. These models are often referred to as small language models (SLMs, e.g., Google's Gemma, Microsoft's Phi-3, Mistral AI's NeMo, Open AI's GPT-Neo, and GPT-4o mini) and aim to be inexpensive to run for API users (sometimes even on personal laptops or mobile devices), be less resource intensive, and hence mitigate carbon emissions and allow broader access. Many experts believe that SLMs can achieve performance comparable to larger models when trained for specific tasks while requiring a fraction of the cost for deployment and operation, exhibiting lower latency, and being more environmentally responsible.

1.2. Several types of AI

At a high level, one can segment AI (or analytics) into four types: descriptive AI, predictive and discriminative AI, prescriptive AI, and generative AI.[11] Descriptive AI aims to answer questions about what happened in the past. It uses historical data to learn new (sometimes surprising) patterns and identify interesting trends that would be challenging for humans to discern. A common output of descriptive AI is a data dashboard, which is a visualization tool of the most pertinent data fields (another output can simply be a report based on exploratory data analyses). In a way, dashboards allow a democratization of data access to all functions of the company. For example, a medical clinic could be interested in monitoring the number of patients seen by each doctor in each specialty, or the proportion of new versus recurring patients. Similarly, a fashion retailer could be interested in understanding the daily sales of each product in each store. Another part of descriptive AI is information retrieval, which refers to the process of searching for and extracting relevant information from large datasets. Predictive AI is more ambitious and aims to answer questions pertaining to what is likely to happen in the future. The goal is to use past data to predict future outcomes. Examples include predicting future demand or sales, the daily number of Covid-19 cases, and customer life-time value. In a similar vein, discriminative AI refers to tasks that aim to classify data into different classes or categories. The goal is to predict labels or categories for given inputs by learning patterns from the data, such as identifying whether an email is spam, recognizing an image, and computing the likelihood of a loan to default. Prescriptive AI is even more ambitious and aims to recommend which concrete actions should be taken. It analyzes data and suggests the best course of action to achieve a desired outcome by relying on optimization models and data-driven algorithms. Examples of applications are endless and include price optimization, inventory management, personalized recommendations, online matching, procurement and fulfillment, staffing and scheduling, fleet optimization, portfolio management, chatbots, and many more. Finally, generative AI creates new content, such as text, images, or audio, by learning patterns from existing data. Generative AI will be further discussed in Section 1.3.

[10] https://aiindex.stanford.edu/report/

[11] Many experts consider that descriptive and predictive applications do not qualify as real AI and are better categorized as data analytics.

As mentioned, AI is fueled by data as a critical input. Today, it has become common to use several types or modes of data. Multimodal AI refers to AI systems that can process and integrate information from multiple types of input modalities, such as text, images, audio, and video. Unlike traditional AI systems that typically specialize in one type of data, multimodal AI combines and correlates diverse data sources to enhance understanding and decision-making. By leveraging multiple modalities, these AI systems can achieve a more holistic and accurate representation of the environment or context they are analyzing, in areas like autonomous driving, healthcare diagnostics, and human–computer interaction.

It is also common to talk about explainable AI and interpretable AI, which refer to the development of AI systems that provide clear, understandable explanations for their decisions and actions. Explainable AI focuses on creating models that can articulate the reasoning behind their outputs in human-understandable terms, thereby increasing trust and transparency. This is crucial in applications where understanding the decision-making process is vital (Doshi-Velez and Kim, 2017), such as in healthcare diagnostics, financial services, and legal decisions. Interpretable AI, on the other hand, emphasizes the inherent transparency of the AI models themselves. It involves designing algorithms that are inherently understandable and whose internal workings can be easily examined and understood by humans (Ribeiro et al., 2016). Both explainable and interpretable AI aim to bridge the gap between complex machine learning models and human users, ensuring that AI systems are not only powerful and accurate but also trustworthy and accountable.

An interesting, successful type of AI system is human-in-the-loop (HITL) AI systems that integrate human expertise into the AI decision-making process to enhance its accuracy, reliability, and ethical considerations and not fully automate the process without human validation. In HITL systems, humans actively participate in training, refining, and evaluating AI models, providing critical oversight and intervention where and when necessary. This approach is particularly valuable in complex or high-stakes domains, such as medical diagnostics or autonomous driving, where nuanced human judgment can complement AI's data-driven insights. By involving humans, HITL AI systems can correct errors, incorporate contextual knowledge, and address ethical concerns. This collaboration between humans and machines can help ensure that AI systems remain aligned with human values and standards, promoting more robust and responsible AI deployment (Amershi et al., 2014). Keeping a human in the loop can be seen as a guardrail, ultimately increasing trust and trustworthiness and reducing several risks.

Finally, a commonly discussed type of AI is artificial general intelligence (AGI), which refers to a type of AI that has the ability to understand, learn, and apply knowledge across a wide range of tasks at a level comparable to human intelligence. Unlike narrow AI, which is designed to perform specific tasks such as language translation or image recognition, AGI aims to fully replicate the cognitive abilities of humans. This includes reasoning, problem-solving, abstract thinking, and generalizing knowledge from one domain to another. The development of AGI involves creating systems that can autonomously improve their performance and adapt to new situations without human intervention. Achieving AGI represents a significant milestone in AI research, and experts have diverging opinions regarding the predicted timeline to reach it.

1.3. Generative AI is a game changer

Without further ado, let's talk about the technology of the day, generative AI. Generative AI is a type of AI that is capable of generating content (e.g., text, images, audio, video, code, blog articles) in response to prompts. A prompt is a natural language text or query that describes the task a user wants to accomplish using AI. This technology was revealed to the public with the release of ChatGPT in November 2022. A mind-blowing fact is that it took only five days for ChatGPT to reach one million users (in comparison to 2.5 years for Airbnb, ten months for Facebook, and five months for Spotify).[12] As of today, ChatGPT has over 180 million users and sees around 600 million monthly visits. When one sees such astronomic numbers, it can only mean that the technology is truly disruptive and tremendously beneficial. And indeed, the very initial benefits were to spur significant amounts of capital and investment in AI by large corporations like Amazon, Google, and Microsoft, smaller startups like Anthropic, Cohere, Inflection, and others, and companies at all layers of the tech stack (e.g., NVIDIA in accelerator chips, CoreWeave and Oracle in cloud compute). Nowadays, one can count a multitude of generative AI tools that aim to tackle different problems and generate various types of information, such as in the generation of text, images, audio and music, video, multimodal output, 3D models, and code.

According to a 2023 McKinsey report,[13] generative AI has the potential to create between $2.6 and $4.4 trillion in annual value to the global economy by boosting productivity. This impact is anticipated through the automation of creative tasks, enhanced efficiency in various industries, and the creation of new products and services. Generative AI tools have already revolutionized numerous fields through their ability to autonomously create new content. These tools leverage massive datasets and deep learning techniques to generate human-like content. GPT models, for instance, can write coherent essays, craft creative stories, and simulate conversations with remarkable fluency. In art, tools like OpenAI's DALL-E and Midjourney can create detailed and imaginative images from textual descriptions, while tools like Artbreeder and DeepArt allow users to generate and customize artwork, blending styles and features. In music, platforms like Amper Music and Jukedeck compose original pieces in various styles, enabling personalized soundtracks and aiding musicians in the creative process. In software development, there are many AI-powered coding assistants available today, designed to help programmers write code more efficiently and accurately by suggesting code snippets. Examples include AlphaCode, Amazon CodeWhisperer, AskCodi, Cursor, GitHub Copilot, Kite, Phid, Snyk, and Tabnine. Healthcare is also benefiting, with generative AI aiding in drug discovery, as demonstrated by Insilico Medicine's AI that identified potential treatments for fibrosis. Additionally, tools like Runway ML facilitate video editing and special effects creation, democratizing access to advanced video production capabilities. These innovations not only enhance productivity but also open new avenues for creative expression and problem-solving across diverse industries. Section 2.3 will discuss in further detail how generative AI tools boost productivity.

While generative AI presents several risks and challenges that require careful regulation in its development, adoption, and large-scale deployment,[14] it also offers many benefits for individuals,

[12] https://www.digitalinformationworld.com/2023/01/chat-gpt-achieved-one-million-users-in.html

[13] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

[14] Many such risks and challenges are listed in Section 6.

society, and the economy. The remainder of this paper will explore how, with appropriate regulatory measures, the positive impacts of AI can far outweigh its potential drawbacks.

2. How does AI benefit humanity today?

Today, AI significantly benefits humanity across various domains by enhancing productivity, creativity, and problem-solving capabilities. From healthcare to education, AI-driven technologies improve efficiency, personalize experiences, and provide innovative solutions to complex challenges. In healthcare, AI aids in diagnosing diseases and discovering new treatments, while in education, it can offer personalized learning experiences tailored to one's skills and capabilities. As mentioned, AI's impact also extends to creative fields, where it can generate art, music, and literature, pushing the boundaries of human creativity. Overall, AI is transforming industries by disrupting economic paradigms (Agrawal et al., 2022b) and is improving the quality of life on a global scale. At a high level, AI can benefit us in three ways: the power of prediction, personalization, and boosting productivity.

2.1. The power of prediction

As discussed, predictive AI analyzes vast amounts of historical data to forecast future outcomes with remarkable accuracy, given sufficient and high-quality data. These models address several societal issues and are thus beneficial to humanity. Agrawal et al. (2022a) explain that as AI reduces the cost of predictions, it becomes more economically viable to use it in many applications. For instance, in healthcare, AI predicts disease outbreaks, infection rates, and patient outcomes by analyzing medical records and genetic data. Methodologies such as machine learning, neural networks, and regression analysis are commonly employed in such contexts. In climate science, AI forecasts weather patterns and natural disasters, such as earthquakes, to improve preparedness for such devastating events. Deep learning algorithms are used to interpret complex climate data and improve prediction models. Other applications include predicting optimal airplane maintenance schedules and advancing preventive medicine through personalized health risk assessments. Researchers leverage AI to generate valuable predictions for businesses, vulnerable populations, and government agencies, showcasing its wide-ranging impact. This article will elaborate on some of these specific application domains, highlighting how AI advancements transform data into actionable insights across various sectors to ultimately benefit humanity.

2.2. The power of personalization

A notable use case where AI can excel is personalization. AI is revolutionizing personalization by tailoring experiences and services to individual preferences and needs. Through sophisticated algorithms and data analysis, AI can first predict user behaviors and responses and then provide a personalized recommendation or action for each type of individual depending on contextual and personal information. In some cases, the personalization can be very fine-grained, and the recommendations are made at the individual level, whereas in other cases, it can be done for a group of individuals with similar characteristics. The right level of aggregation depends on the quantity and quality of the data and on the homogeneity between subjects.

AI is increasingly being implemented in user-facing solutions in various industries, revolutionizing how individuals interact with technology and services. From virtual assistants like Amazon's Alexa and Apple's Siri that use NLP to respond to user queries, to personalized recommendations on platforms like Netflix and Spotify that employ machine learning algorithms to recommend content based on user preferences, AI is enhancing user experiences in unprecedented ways. More specifically, streaming services like Netflix use AI to suggest movies and shows based on viewing history and other features (e.g., time of day, pace at which users watched a full season of a show, pace at which users stopped watching a movie). E-commerce platforms leverage AI-driven recommendation engines to suggest products tailored to each user's browsing and purchase history, enhancing customer satisfaction and sales. Interestingly, AI-driven personalization extends to several applications that can substantially benefit us: first and foremost is healthcare, where AI-driven personalization can help create individualized treatment plans by analyzing detailed patient data; other examples include mental health, dating platforms, personalized fitness programs, and customized investment portfolios. In many cases, a one-size-fits-all approach does not perform well and can even be detrimental to specific groups of people. Leveraging granular data and advanced AI techniques to produce the right customization has the potential to enhance user satisfaction, increase engagement, and lead to significantly better outcomes, hence making AI a powerful tool for delivering personalized experiences. Several examples of applications where AI can leverage personalization for social good are discussed in Section 2.4.

2.3. The power of boosting productivity

With the recent rise of generative AI, many individuals and small firms found innovative ways to leverage this technology to enhance efficiency and productivity. AI can be leveraged to significantly boost productivity across various industries by automating tedious tasks and optimizing workflows in daily life. In manufacturing, AI-powered robots and predictive maintenance systems reduce downtime and increase efficiency. In the corporate world, AI-driven tools streamline administrative tasks, such as scheduling, data entry, and customer service, through chatbots, freeing up employees' time to focus on more strategic work. In agriculture, AI can be used to optimize crop management and yields by analyzing weather patterns and soil conditions. Additionally, AI enhances decision-making processes by quickly analyzing large datasets to identify trends and opportunities. By leveraging AI, businesses can achieve greater efficiency, reduce operational costs, and enhance overall productivity. The goal is to use AI as an assistant for painful repetitive tasks, allowing humans to be more efficient and more productive by reallocating their time to high-value-added tasks.

Several recent studies have documented the productivity gains from using generative AI for various tasks. Researchers at GitHub, Microsoft, and MIT have found that developers who use GitHub Copilot completed a JavaScript task 55.8% faster than those who do not (Peng et al., 2023). Nowadays, most developers cannot imagine writing software code without the assistance of generative AI. Both the productivity and quality gains are enormous, and the barrier to entry has significantly reduced (e.g., to learn a new programming language or to seamlessly build a website or an app). In Noy and Zhang (2024), the authors conducted an experiment asking college-educated professionals to complete incentivized midlevel professional writing tasks. The randomly assigned treatment group could use generative AI while completing their writing tasks. They found that the participants using generative AI were more productive, efficient, and enjoyed the tasks more. The

average time to complete the task decreased by 40% and the output quality rose by 18% when the professionals used generative AI. Brynjolfsson et al. (2023) found that a generative AI-based conversational assistant increased the productivity of customer support agents by 14% on average. It should be noted that while the productivity impact on novice/low-skilled workers was 34%, the impact on experienced/highly skilled workers was minimal, hence reducing inequalities among workers by helping the inexperienced ones more. Dell'Acqua et al. (2023) studied the performance implications of AI on realistic, complex, and knowledge-intensive tasks in collaboration with the Boston Consulting Group. Consultants were assigned to one of three conditions: no AI access, GPT-4 AI access, or GPT-4 AI access with a prompt engineering overview. Consultants using AI completed 12.2% more tasks on average, completed them 25.1% more quickly, and produced work with a 40% higher quality. Finally, Eloundou et al. (2023) show that, with access to an LLM, around 15% of all worker tasks in the U.S. could be completed significantly faster and at the same level of quality, whereas embedding software on top of LLMs (i.e., LLM-powered software) increases this share to 47–56% of all tasks. Interestingly, this effect applies to all wage levels, hence suggesting that LLMs can have a considerable economic impact on the labor market.

While the above studies examined the impact on the productivity of workers performing specific tasks, other studies have investigated the impact on the productivity of firms who use AI. Czarnitzki et al. (2023) considered a sample of German firms and found that there are positive and significant associations between the use of AI and firm productivity, and that both the use of AI and the intensity with which firms exploit the potential of AI significantly increase sales and value-added. Babina et al. (2024) found that AI-investing firms experience higher growth in sales, employment, and market valuations. Yu and Qi (2024) considered the 100 largest publicly traded U.S. companies and found that generative AI has created a positive and statistically significant effect on labor productivity for the firms who adopted it.

In this context, it is important to remind ourselves that in most cases, these AI tools aim at boosting productivity and not replacing humans. In many tasks, human validation and tuning is a critical component to ensure that the outputs are appropriate and accurate. A common way to look at these AI tools is as "copilots" to help decision makers by delegating some tedious, low-added-value tasks to AI. In fact, in February 2023, Microsoft launched Copilot, which is a generative AI tool integrated to Microsoft 365 that "leverages the power of AI to boost productivity, unlock creativity, and helps you understand information better with a simple chat." Copilot offers numerous benefits by integrating AI into everyday tasks, significantly enhancing productivity and efficiency. For instance, in coding, Copilot assists developers by suggesting code snippets, autocompleting functions, and identifying potential errors, thus accelerating the development process. In writing, it can aid in drafting emails, creating reports, and generating content by providing relevant suggestions and improving language quality. For data analysis, Copilot can automate data entry, generate insights, and create visualizations, making complex data more accessible and actionable. By streamlining these tasks, Copilot helps users focus on higher level strategic work by reducing the time spent on repetitive and routine activities. Several companies have started adopting tools like Copilot at the corporate level. For example, Beiersdorf, a skin care industry leader, has accelerated its research and development processes by leveraging AI to analyze vast amounts of data efficiently while also reducing its carbon footprint.[15] Cambon et al. (2023) summarized various studies done by Microsoft to test how their Copilot tool can boost

[15] https://ms-f1-sites-03-ea.azurewebsites.net/en-gb/story/1739992869010829656-beiersdorf-azure-data-lake-consumer-goods-en-germany

productivity. The tasks in their studies were related to (1) summarizing a missed meeting, (2) enterprise information retrieval, (3) email writing, (4) enterprise security operations, and (5) coding. Their results support the hypothesis that Copilot substantially increases productivity. The typical productivity boost is in the form of increased speed without a decrease in quality. More precisely, the authors report that Copilot users completed tasks more quickly (between 26% and 73%) when compared to those not using Copilot. Similar productivity tools were launched by other technology companies, such as the introduction of Amazon Q (as part of AWS AI services), Gemini by Google, Projects by Anthropic, and Slack AI by Salesforce.

Interestingly—and perhaps still a bit controversially—to write content like this paper, generative AI tools like ChatGPT and Claude can serve as great assistants. Not only can they correct typos and grammatical errors, but they can also shorten or elaborate on specific parts, strengthen the transition between paragraphs, suggest great sources and references, propose section titles, and deliver a first draft for specific topics. And yes, such tools were extensively (and carefully) used when writing this paper, and the productivity gain is assessed to exceed 50%.

2.4. Examples of positive impacts

While AI introduces new challenges and debates in our society, this section discusses several domains and applications where AI has already made a remarkable positive impact. The content in this section is by no means exhaustive and includes only a subset of the applications where AI has had (or will soon have) a great social impact. A comprehensive list would be prohibitively long and grow longer every day.

Healthcare. Many experts firmly believe that healthcare is the domain where AI can have the most significant impact, provided its adoption and governance are properly regulated. It goes without saying that carefully regulating AI adoption, its governance, and compliance are paramount in an application like healthcare. Many hospitals, medical researchers, and AI experts have been actively working on developing safe data-driven technologies to solve various problems in the healthcare landscape. It all starts with putting the right data infrastructure in place and collecting as much relevant granular information as possible while conforming to privacy and security protocols (often, formal consent from both patients and healthcare professionals needs to be received). AI applications in healthcare are endless and have the potential to enhance the quality of care and increase access for a larger number of patients. These applications include developing personalized treatments and drug plans, preventive care, rare disease detection, enhanced cancer screening processes, improved radiology diagnoses accuracy, improved organ donation and allocation processes, predicting the number of infections at a granular level, predicting the number of hospitalizations as well as the hospital length of stay and discharge time for each patient, predicting the likelihood of relapse, making the logistics for hospitals and vaccine deliveries more efficient, and the list goes on.

A very promising (perhaps, even revolutionary) AI application in healthcare is drug discovery, where AI has been used to accelerate the identification of new drug candidates, optimize clinical trials, and personalize medication. For example, DeepMind has developed AlphaFold, an AI system that aims to accurately predict protein structures, a crucial step in drug discovery (Jumper et al., 2021). Companies like Atomwise use AI to predict how small molecules interact with

proteins, streamlining the process of finding potential drug compounds. Several companies employ AI to design novel drug candidates and identify biomarkers for various diseases, significantly speeding up drug development timelines. Another example is companies like BenevolentAI that leverage AI for biomedical data to discover new drug targets and repurpose existing drugs for new indications. Several research studies have highlighted that AI-driven approaches can substantially expedite drug discovery processes and improve the success rates of clinical trials (see, e.g., Vamathevan et al., 2019). The integration of AI into drug discovery is transforming the industry by enhancing efficiency, reducing costs, and accelerating the development of new therapeutics.

A final interesting recent application is leveraging AI to boost clinicians' productivity and reduce their fatigue and burnout. Several studies have shown that physicians spend nearly half of their time on administrative tasks, primarily on documentation, and this is a major contributing factor to clinician fatigue and burnout. An important part of healthcare documentation requires clinicians to write a summary consultation note after each patient visit. Recent advancements in generative AI have led to the emergence of AI medical scribes, a new solution that promises to significantly reduce the time spent on documentation. By leveraging AI to help with documentation tasks, clinicians can reallocate their time to what they do best, namely, seeing patients and diagnosing their conditions. AI medical scribes have the potential to save precious minutes for each patient encounter, reduce clinician fatigue, and provide a better experience to patients. Adopting AI medical scribes is another great example in which AI can serve as a copilot to complement humans and gain efficiency. For more details on AI medical scribes, including their benefits and challenges, refer to Cohen and Toubiana (2024) and Tierney et al. (2024).

Wellness and fitness. Another closely related domain is wellness and fitness. Countless AI applications and tools have been developed with the promise of enhancing one's well-being, physical health, and mental health. For example, smart watches and fitness trackers are now constantly tracking daily activity, captured by the number of steps, calories burned, heart rate, sleep duration and quality, blood oxygen levels, stress levels, respiratory rate, skin and body temperature, and sometimes even blood pressure. This massive amount of data can then be used to train AI algorithms to make personalized suggestions about physical activity (e.g., sending a notification encouraging exercise), specific diets, sleep schedules, and food supplements. In fact, many research studies and commercial solutions have focused on AI-driven personalized behavior changes to improve lifestyle and longevity. For example, Thrive AI Health, a new company funded by the OpenAI Startup Fund and Thrive Global, is devoted to building an AI health coach. The company's mission is to use AI to democratize access to expert-level health coaching in order to improve health outcomes and address health inequities. The goal is to leverage AI to deliver hyper-personalized behavioral change across the five key behaviors that Thrive focuses on: sleep, food, movement, stress management, and connection.[16]

Similarly, AI has significantly advanced mental healthcare by enhancing accessibility, personalization, and efficiency. Companies like Woebot Health and Wysa have developed AI-driven chatbots that provide users with cognitive-behavioral therapy techniques and therapeutic conversations (to help users manage stress and anxiety with interactive conversations), respectively, hence offering immediate, affordable support and reducing the barrier of accessing mental health services. Another noteworthy example is Mindstrong Health, which leverages AI to

[16] https://time.com/6994739/ai-behavior-change-health-care/

analyze smartphone usage patterns and detect early signs of mental health issues, enabling timely intervention. These technologies aim to democratize mental healthcare, making it more accessible and scalable. Research supports the efficacy of these AI tools; for instance, Fitzpatrick et al. (2017) found that Woebot's AI-based interventions can be as effective as traditional therapy methods for some users. The integration of AI into mental healthcare is transforming the field by providing scalable, personalized, and timely support, potentially revolutionizing how mental health services are delivered. Recent advancements in generative AI and personalized agents can naturally be very useful in this context and can offer highly personalized experiences at low cost.

Climate science and ecology. While AI technologies, particularly large-scale machine learning models, LLMs, and energy-intensive data centers, contribute significantly to carbon emissions, AI is also making a profound positive impact on climate science and ecology by enhancing predictive modeling, optimizing resource management and energy consumption, and improving conservation efforts. For instance, IBM's Green Horizons initiative uses AI to predict air pollution levels and provide actionable insights to improve air quality in cities, leveraging vast amounts of environmental data globally. ClimateAi utilizes machine learning algorithms to offer precise climate risk assessments for agriculture, helping farmers adapt to climate change by optimizing crop yields and reducing losses. In the field of ecology, Microsoft's AI for Earth initiative supports projects that use AI to monitor ecosystems, track wildlife populations, and predict deforestation, thereby aiding in biodiversity conservation and the sustainable management of natural resources.[17] Additionally, Google's DeepMind applies AI to model and predict the effects of climate change on ice sheet dynamics, contributing to more accurate sea-level rise forecasts.[18] These examples illustrate how AI technologies are providing critical tools for scientists and policymakers to address and mitigate the impacts of climate change, thereby fostering a more sustainable and resilient planet.

Similarly, AI is significantly enhancing wildlife conservation efforts by providing innovative tools for monitoring, protecting, and managing wildlife populations. Wild Me uses computer vision and deep learning to identify individual animals based on their unique patterns and features from photographs, enabling researchers to effectively track endangered species like whale sharks and African elephants. Google's Wildlife Insights platform utilizes AI to analyze camera trap images, quickly sorting through millions of photos to identify species and monitor biodiversity, helping conservationists make data-driven decisions. Microsoft's AI for Earth initiative supports projects that use AI to combat poaching by predicting poaching activities and enabling more efficient patrolling. These applications demonstrate how AI is positively impacting wildlife conservation by providing powerful tools to gather and analyze vast amounts of data, ultimately improving the ability to protect and preserve endangered species and their habitats.

Legal systems. AI has also had a substantial impact in legal practice. More specifically, AI is transforming legal systems by enhancing efficiency, accuracy, and accessibility in various legal processes. Ravel Law, acquired by LexisNexis, uses AI to perform advanced legal research and analytics, allowing lawyers to quickly find relevant case law and gain insights from judicial opinions. LawGeex leverages AI to automate contract review and approval, significantly reducing the time required to process legal documents and ensuring higher accuracy in identifying potential

[17] https://news.microsoft.com/apac/features/ai-for-earth-helping-save-the-planet-with-data-science/

[18] https://deepmind.google/discover/blog/using-ai-to-fight-climate-change/

issues. Companies like ROSS have developed AI-driven legal research tools to streamline the process of finding legal precedents, thereby improving lawyers' productivity and decision-making. DoNotPay, self-branded as "Your AI Consumer Champion," is an AI-powered chatbot that helps individuals contest parking tickets, seek compensation for flight delays, and handle small claims court cases, making legal assistance more accessible and affordable for the public. These are just a small number of examples among the growing number of legal-tech AI companies that illustrate how AI is revolutionizing the legal sector by automating routine tasks, enhancing research capabilities, and making legal services more accessible (and affordable) to a broader audience.

Other examples. Another example lies in leveraging data and AI solutions to help supply chain management, logistics, and retail operations, from detecting panic buying for products amid the pandemic (Adulyasak et al., 2023) to developing resilient supply chains (Cohen and Tang, 2024). Another application domain is the various studies that aim to leverage AI to improve traffic conditions and the algorithms that govern self-driving cars in order to reduce the number of car crashes and thus decrease the number of fatal accidents on the road. This is another example of an important societal issue for which data and AI can help tremendously.

The fashion industry has seen various applications of AI with a positive impact. Notorious examples are leveraging AI to optimize designs and sustainable packaging as well as reduce waste. For example, the Portuguese fashion company Fashable uses generative AI to create sustainable designs faster, reducing fabric waste and promoting sustainability.[19] Many other similar examples of fashion companies exist. In electricity and energy, the impact has also been considerable. For example, the German electric utility company E.ON SE employs AI-driven drones to inspect power lines, ensuring worker safety and uninterrupted energy supply.[20]

There is a myriad of additional examples of applications where AI has been very beneficial for humanity and our society, such as humanitarian logistics (e.g., predicting natural disasters, managing crisis response and recovery, and coordinating relief efforts), alleviating poverty and hunger (e.g., waste management), finding love (e.g., online dating algorithms), and finding a job (especially in the remote work environment). Other notorious examples include recent progress in AI-enhanced home technologies and AI for education (e.g., personalized learning, assisting instructors by automating tasks, and free accessible AI tutors). Many official reports have been written on the impact that AI can have on education and pedagogy, including by the UK Parliament,[21] the World Bank,[22] and the World Economic Forum.[23] A comprehensive discussion on some of the above topics can be found in the book by Bapna and Ghose (2024). As discussed, the above examples are only a small, non-comprehensive list. Many other application domains have seen a big impact, too, and the list will continue to expand in the years to come. At the same time, it is important to acknowledge that AI has several risks and challenges (see Section 6).

[19] https://customers.microsoft.com/en-us/story/1558909662014453187-fashable-retail-azure

[20] https://www.eon.com/en/about-us/media/press-release/2022/eon-extends-innovation-cooperation-with-microsoft.html

[21] https://researchbriefings.files.parliament.uk/documents/POST-PN-0712/POST-PN-0712.pdf

[22] http://documents1.worldbank.org/curated/en/099734306182493324/pdf

[23] https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Learning_2024.pdf

3. Democratization of AI

3.1. AI for all

In the past, AI was mainly used by large companies with technical teams and big budgets. Today, AI tools and technologies are no longer restricted to a small set of firms but are rather widely accessible across various sectors and to individuals with diverse levels of expertise. Advances in AI frameworks, open-source platforms, and cloud-based services have made powerful AI tools available to small businesses, startups, and even individual developers. Similarly, compute capabilities made available by large tech companies are offering the possibility to process large amounts of data efficiently. In particular, the broad access and global reach of the public cloud has allowed almost instantaneous diffusion of the technology worldwide. As a result, more people than ever can now leverage AI for tasks such as data analysis, NLP, and smart automation, without needing extensive technical knowledge. This democratization fosters innovation, enables personalized solutions, and drives economic growth by allowing a broader range of users to harness the transformative power of AI. It has become very affordable (and sometimes even free) to access the latest AI advancements, allowing small companies and developers to enjoy access to high-performing models and tools. But democratization is not limited to companies. Recent advancements in generative AI have brought a democratization of AI tools to the masses. Individuals now have access to expert advice at no cost.

For instance, generative AI can now be routinely employed to answer various questions and provide advice on a wide range of topics. While the answers are not guaranteed to always be fully accurate and precise, they can still be very helpful in many cases. In addition, the quality and accuracy of such tools have drastically improved in the past few months and will continue to improve. Imagine a scenario where someone has legal questions. In many cases, getting an answer to a specific legal question would require having access to a legal professional, which can be costly and take time, thus raising issues of affordability, prohibiting access to the most vulnerable people. Today, generative AI tools can provide answers to several legal questions in many jurisdictions (e.g., Bhambhoria et al., 2024). The same applies to questions about taxes, visa applications, and other administrative inquiries.

Generative AI has also shown promising capabilities in addressing medical inquiries, leveraging its ability to analyze vast medical databases and generate contextually relevant responses. For instance, platforms like Ada Health utilize AI to provide personalized medical advice based on user inputs and symptoms, enhancing accessibility to healthcare information. Moreover, research at the Mayo Clinic indicates that AI-powered systems can assist in diagnosing complex conditions by interpreting medical images with high accuracy, potentially improving patient outcomes.[24] For example, AI-powered chatbots can offer preliminary health advice and triage patients, directing them to the appropriate care based on their symptoms. Additionally, AI can assist doctors by summarizing patient records, suggesting evidence-based treatment options, and staying updated with the latest medical research. While one needs to be cautious about the various risks and challenges, this type of AI integration enhances accessibility to medical knowledge, improves diagnostic accuracy, and supports better patient outcomes. These advancements underscore the transformative potential of generative AI in augmenting medical decision-making and patient care

[24] https://www.mayoclinic.org/giving-to-mayo-clinic/our-priorities/artificial-intelligence

through innovative technological solutions. Still in the healthcare context, AI technologies can directly help patients especially given the shortage of healthcare professionals worldwide. A recent example is the concept of "AI nurses" launched by companies like Hippocratic AI, which recently announced a collaboration with NVIDIA.[25] This concept represents an innovative approach to augmenting healthcare delivery through AI-driven virtual assistants designed to support healthcare professionals by handling routine tasks such as patient monitoring, appointment scheduling, and basic diagnostic assistance. Companies like Hippocratic AI aim to alleviate the administrative burden on nurses, allowing them to focus more on direct patient care and complex medical decision-making. By leveraging NLP and machine learning capabilities, these AI nurses are available 24/7 and can provide timely information and personalized patient guidance as well as facilitate efficient communication between patients and healthcare providers. This integration of AI into nursing practice exemplifies the potential of technology to enhance healthcare efficiency and patient outcomes in increasingly digitalized healthcare environments.

These types of AI applications are offering high-quality (and often affordable) services to the masses. Once reserved for a select subset of individuals, thanks to AI developments, many such services are becoming available to everyone at a low cost. While the above discussion has focused on recent examples in healthcare, legal help, and tax assistance, many additional application domains appear in the marketplace every week.

3.2. Building on top

Another interesting aspect is the fact that startups and individuals, who have now access to LLMs and other sophisticated AI models, can take them to the next level by developing them further. More specifically, consider the situation where one has access to a general-purpose LLM like Claude, Cohere's Command R+, Gemini, GPT4, Llama 3, or Mistral 7B. These models are trained using a massive amount of data, and they can be used to generate content on all domains and topics. If a company is interested in developing an LLM specifically trained on a particular domain, it is possible to use one of the LLMs available and fine-tune it with data (either publicly available or proprietary) from a specific domain or task. For example, OpenAI allows enterprise customers to use their own data to fine-tune their latest LLM (GPT-4o) so they can build custom models to complete specific tasks. The goal of this fine-tuning process is to improve the model's ability to understand specific use-case instructions. More precisely, developers or companies can upload their text-based data to OpenAI's servers, and the fine-tuning process usually takes a few hours, hence allowing for quick customizations. Developers can further build AI agents based on LLMs. AI agents are autonomous programs designed to perform specific tasks or functions by perceiving their environment, processing information, and taking actions to achieve clearly designated goals. AI agents are used in numerous applications, such as customer support chatbots, marketing content generators, personal finance advisors, educational tutors, recipe generators, and travel assistants. The development of AI agents continues to evolve, aiming to create more intelligent, efficient, and adaptable systems that can operate seamlessly in dynamic and complex environments.

This is a big advancement in terms of opening the market and allowing a collaborative ecosystem to foster progress and innovation. Since one could potentially apply this idea to any field (e.g., creating a model to specifically generate advertising campaigns or creating one to generate answers

[25] https://qz.com/nvidia-wants-replace-nurses-with-ai-1851349522

about U.S. taxes), it offers a wide range of opportunities. Section 5 will discuss the importance of partnerships and collaborations to foster AI development and advance our collective understanding and positive impact.

3.3. Increasing fairness

If designed carefully, AI tools and algorithms have the potential to significantly mitigate biases, thereby promoting fairness and reducing discrimination and inequality in various domains. Research has shown that AI can help counteract human biases by processing data objectively and making decisions based on statistical patterns rather than subjective factors. For instance, a study by Obermeyer et al. (2019) demonstrated that AI algorithms could reduce racial disparities in healthcare by more accurately assessing patients' needs compared to traditional methods. Furthermore, AI-driven recruitment platforms have been shown to mitigate gender bias in hiring processes by focusing on candidates' skills and qualifications rather than demographic characteristics (Dastin, 2022). Additionally, algorithms in criminal justice systems have been developed to minimize racial disparities in sentencing by providing judges with data-driven insights into recidivism risk, thereby reducing the influence of subjective biases (Angwin et al., 2016). A last example is a recent study showing that generative AI can reduce price discrimination in the housing market (Tanlamai et al., 2024). These examples highlight how AI tools can enhance objectivity and equity in decision-making processes across sectors, illustrating their potential to contribute positively to societal equality and justice.

At the same time, AI systems can inherit biases from the data they are trained on (which is generated by humans, who themselves have biases), or from the algorithms used. These biases can arise from various sources, including historical societal prejudices embedded in datasets or the design algorithmic choices made by developers. For example, facial recognition algorithms have shown biases in accurately identifying individuals from different racial or ethnic backgrounds (Buolamwini and Gebru, 2018). Without careful design, AI algorithms can perpetuate biases (and misinformation) and even amplify them. Addressing these biases is crucial to ensure that AI systems make fair and equitable decisions across diverse populations. Recognizing and mitigating biases in AI is an ongoing challenge that requires scrutiny, ethical considerations, and continual refinement of algorithms to promote unbiased outcomes. Ethical AI and fairness in the context of AI-driven decisions is an active research topic (see, e.g., Binns, 2018; Jobin et al., 2019).

4. Ensuring that this process continues to deliver

4.1. Safe data collection

As discussed, data can be seen as the fuel for AI. In most AI applications, both the quantity and the quality of data available are important. Without data, there is very little that can be done in terms of solving problems using AI models and algorithms. While the specific data requirements highly depend on the focal use case and application domain, a typical rule of thumb is the more data, the better. At the same time, data quality (e.g., its variation, representativeness, and heterogeneity) is also crucial. As the old saying goes: "garbage in, garbage out," meaning that if the quality of the data to train AI models is low, the output will also be of poor quality. When collecting and storing data (with the goal of using it to train AI models), several considerations are

typically needed. First, one needs to decide the level of data aggregation. The answer has several dimensions, including temporal aggregation (e.g., data at the month/week/day/hour/transaction levels), geographical aggregation (e.g., data at the country/city/zip code/user levels), and object aggregation (e.g., data at the category/brand/product levels). A more granular dataset is typically more costly and can be noisier but includes much richer information, whereas a higher aggregation level is more compact at the expense of losing some data variation. The right trade-off depends on the specific use case and is often based on experience.

First, one needs to collect and store data in a safe and secure way that satisfies all regulatory and privacy requirements. As more and more data is routinely collected in our everyday lives, it is crucial to ensure that it is collected properly and carefully. A critical concern is around data privacy. Ensuring that data security and privacy measures protect sensitive information from breaches and unauthorized access would allow safeguarding individuals' privacy and, thus, maintaining trust in AI systems. As AI relies on vast amounts of data for training and improving its algorithms, any compromise in data integrity can have significant repercussions, including inaccurate predictions, biased outcomes, and ethical concerns. Moreover, data breaches can result in the exposure of confidential personal, financial, or health information, leading to identity theft, financial loss, and reputational damage for both individuals and organizations. In addition to privacy concerns, the integrity and quality of data are crucial for developing reliable and effective AI models. Secure data storage prevents tampering, ensuring that the data remains accurate and unaltered over time. This is essential for maintaining the validity of AI training processes and the subsequent performance of AI applications. Implementing robust encryption methods, access controls, and audit trails can mitigate risks associated with data breaches and unauthorized modifications, providing a secure environment for data handling. Furthermore, adherence to legal and regulatory frameworks, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), is essential for organizations involved in AI development. These regulations mandate data protection measures, granting individuals greater control over their personal data and imposing severe penalties for noncompliance. By following these guidelines, organizations not only ensure legal compliance but also demonstrate their commitment to ethical AI practices. In the era of big data, where information is constantly being generated and collected from various sources, establishing secure data collection and storage protocols is a fundamental aspect of responsible AI development. This includes adopting advanced cybersecurity measures, regularly updating security protocols, and educating employees about data security best practices. Moreover, fostering transparency in data-handling processes and involving stakeholders in data governance can enhance trust and accountability.

Second, the topic of fake data has considerable implications in the AI world. Fake data can take several forms, including synthetic data and fabricated data. Such data has substantial implications, influencing both the development and deployment of AI systems. On the one hand, synthetic data can be intentionally used to bolster the AI training processes and to test AI algorithms before deploying them in the real world. This is especially useful when real-world data is scarce, sensitive, or difficult/expensive to obtain. By generating synthetic data that mimics real data (e.g., preserves the same statistical and structural properties), researchers can create extensive datasets that enhance the robustness, generalizability, and scalability of AI models. This approach is particularly valuable in fields like healthcare, where privacy concerns severely limit data availability. On the other hand, the use of fake data also presents significant risks and ethical concerns. If synthetic

data is not accurately representative or contains biases, it can lead to flawed AI models that will perform poorly in real-world scenarios, potentially causing harm or perpetuating inequalities. Furthermore, malicious use of fake data, such as feeding AI systems with deliberately misleading information, can undermine the integrity and trustworthiness of AI applications. For instance, adversarial attacks leveraging fake data can deceive AI systems, leading to incorrect decisions in critical areas like autonomous driving or security surveillance. The proliferation of deepfakes,[26] a form of synthetic media, exemplifies the potential for misuse, as they can spread misinformation and erode public trust. Therefore, while fake data offers opportunities for advancing AI, it necessitates rigorous validation, ethical considerations, and robust safeguards to ensure its beneficial and responsible use in the AI ecosystem. It is thus important to put robust systems in place to carefully check the veracity of data and to not base important societal decisions driven by AI algorithms before validating data quality and representativeness.

Third, as discussed, the topic of human biases in data is also important and needs to be accounted for. Human biases in data are a critical concern when designing AI algorithms, as they can significantly influence the outcomes and the fairness of AI-driven decisions. Biases can enter data at various stages, from data collection to labeling, reflecting the prejudices, stereotypes, and subjective judgments of those involved. For instance, historical data on hiring practices might reflect gender or racial biases, which, if used to train AI recruitment tools, can perpetuate and even exacerbate these biases, ultimately leading to discriminatory hiring decisions. Similarly, biased data in law enforcement records can result in AI systems that unfairly target certain demographic groups, reinforcing systemic inequalities. The implications of biased data are profound, as AI models trained on such data are likely to inherit and propagate these biases, undermining the objectivity and fairness of their predictions and recommendations. This may not only affect individual lives but also erode trust in AI technologies, hindering their widespread adoption and potential benefits. An additional common example in this context is biased credit-scoring algorithms that can unjustly deny loans to qualified applicants from marginalized communities.

Addressing human biases in data is crucial for developing fair, ethical, and trustworthy AI systems. Ensuring that AI systems are trained on unbiased and representative data is fundamental to achieving equitable outcomes and fostering public trust in AI technologies. This involves implementing rigorous data auditing practices, employing diverse datasets, and incorporating fairness constraints into AI algorithms. Techniques such as bias detection and mitigation, adversarial debiasing, and fairness-aware machine learning are essential in identifying and correcting biases in training data and model outputs. Additionally, fostering a diverse and inclusive team of AI developers and stakeholders can help in recognizing and addressing potential biases from multiple perspectives. As AI continues to permeate various aspects of our society, addressing data biases becomes increasingly important to prevent the reinforcement of existing inequalities and to promote ethical and responsible use of AI. As discussed in Section 3.3, if implemented properly, AI algorithms can in fact alleviate biases and ensure that AI-driven decisions lead to fair and equitable outcomes.

[26] A deepfake is a synthetic media (image, video, or audio recording) created or manipulated using AI. Nowadays, AI can generate highly realistic but entirely fabricated media that depict people saying or doing things they never actually said or did. These deepfakes can then be used for malicious purposes like spreading misinformation, defamation, or fraud.

4.2. Importance of open source

Recent progress in AI development and deployment has been spectacular and has helped advance our understanding of several important societal problems. As mentioned, one way to ensure that future progress continues to deliver value is by fostering a collaborative ecosystem where individuals and companies of all sizes can easily contribute and innovate. Numerous initiatives in the AI space have promoted open-source tools, with developers publicly sharing their models and algorithms for others to build upon. In the context of AI, open source refers to the practice of making the source code, models, and sometimes even the training data publicly available.

Open source is critically important in the context of AI as it fosters innovation, accessibility, collaboration, reproducibility, and transparency within the field. By making AI tools, frameworks, and algorithms freely available (or by offering an affordable option), open source democratizes access to cutting-edge technology, allowing researchers, developers, and organizations of all sizes to contribute to and benefit from the latest AI advancements. Specific examples include TensorFlow and PyTorch, two of the most popular open-source AI frameworks developed by Google and Meta, respectively. These platforms provide powerful tools for building and training machine learning models, enabling a wide range of applications from NLP to computer vision. Another related example is OpenAI's GPT models, which, while not open source, have a commercial API service that allows developers to build applications leveraging their advanced natural language understanding capabilities. A final example is Llama 3, developed by Meta; it is an advanced, open-source, generative AI model. Its open-source nature is one of its standout benefits, as it allows researchers, developers, and organizations to freely access, modify, and build upon the model, fostering innovation and collaboration in the AI community. Llama 3 is aimed at delivering high-quality text generation, understanding, and conversational abilities, making it a powerful tool for applications ranging from content creation to complex problem-solving. Its accessibility and flexibility position it as a valuable resource for those looking to integrate cutting-edge AI into their projects without the constraints of proprietary systems. The open-source community also benefits from datasets such as ImageNet and COCO, which are crucial for training and benchmarking AI models in image recognition tasks. GitHub repositories and collaborative platforms like Hugging Face's model hub further exemplify the importance of open source by hosting pretrained models and allowing users to share their own models, facilitating rapid experimentation with and deployment of AI technologies.

This collective effort accelerates the pace of innovation, as diverse perspectives and expertise converge to solve complex problems and improve state-of-the-art models. Open source also promotes transparency, enabling the wider community to scrutinize and validate AI algorithms, which is essential for building trust and ensuring ethical AI practices. Additionally, open-source resources provide educational opportunities, equipping students and professionals with the skills needed to develop and implement AI solutions. This collaborative ecosystem not only drives technological progress but also ensures that the benefits of AI are widely distributed, enhancing the overall impact of AI on society.

While the open-source philosophy is possible in some cases (e.g., academia, not-for-profit organizations, or small parts of tech companies), it is not always possible due to the competitive nature of certain industries and the need to ensure a positive return on investment. In such cases,

offering affordable licenses and forming partnerships and alliances could be a great way to foster innovation and ensure that advancements will thrive. Partnerships and alliances are crucial in the context of AI, as they enable the pooling of resources, expertise, and data to drive innovation and address complex challenges that individual entities might struggle to tackle alone. For instance, the partnership between IBM and MIT for the MIT-IBM Watson AI Lab leverages the strengths of both institutions to advance AI research in areas such as healthcare, cybersecurity, and ethical AI. Another example is the collaboration between Microsoft and OpenAI, which combines OpenAI's cutting-edge research with Microsoft's Azure cloud platform to scale AI capabilities and make advanced AI tools accessible to developers and businesses. Another recent partnership involves Apple and OpenAI, aiming to integrate ChatGPT into iOS, iPadOS, and macOS experiences, thereby enhancing various Apple services like Siri and writing tools. The AI for Earth initiative by Microsoft, which partners with various environmental organizations, aims to address global environmental challenges by using AI solutions for issues like climate change, biodiversity conservation, and water management. Google has also engaged in significant partnerships, such as its partnership with Mayo Clinic, which aims to transform patient care by applying AI to radiotherapy planning and improving the accuracy of patient diagnoses. Furthermore, the alliance between NVIDIA and various car manufacturers, including Tesla, highlights how partnerships can accelerate the development of autonomous driving technologies by integrating NVIDIA's AI-powered hardware with the automotive industry's expertise. These collaborations exemplify how strategic partnerships can amplify the impact of AI, driving progress, fostering innovation, and ensuring that the benefits of AI are widely shared across different sectors and communities. More examples of AI partnerships are discussed in Section 5.1.

4.3. Facilitating custom models

Facilitating AI custom models is crucial for maximizing the technology's potential across diverse applications. Custom models are tailored to specific tasks, industry-related problems, and user needs, ensuring higher accuracy and pertinence in their outputs. For instance, in healthcare, custom AI models can analyze patient data to provide personalized treatment plans (models could even be customized for specific medical conditions and types of patients). This level of customization allows businesses to address niche challenges, improve efficiency, and enhance user experiences across several verticals. Additionally, it enables organizations to maintain competitive advantages by leveraging AI that is finely tuned to their operational contexts. By facilitating the development and deployment of custom AI models, we can drive innovation, foster greater adoption of AI technologies, and achieve more impactful outcomes across many domains.

Similarly, developing small, dedicated AI models trained with domain-specific data is also important. Such dedicated models can be trained to provide highly accurate, relevant, and efficient solutions tailored to specific fields or tasks. Unlike general (and often much larger) AI models, these specialized models are trained on datasets that are directly relevant to specific industries or tasks, enabling them to understand and address unique challenges and nuances. For example, in medical diagnostics, an AI model trained exclusively on dermatological images can more accurately detect skin conditions than a general model trained on a massive dataset with all sorts of different images. In finance, models trained on specific market data can better predict stock movements or detect fraudulent transactions. These dedicated models also require less computational power and resources, making them more accessible and cost-effective for

organizations with limited infrastructure, while also being more environmentally sustainable. By focusing on domain-specific data, small AI models deliver precision, enhance performance, and drive innovation within specialized areas, leading to improved outcomes and increased efficiency. Interestingly, a common approach to developing a small, dedicated model is by starting with a large, generic model (either open source or easily accessible) and carefully fine-tune it with a context-specific dataset (either proprietary or publicly available) to achieve better results for specific tasks. This fine-tuning process can sometimes be sophisticated and involve several validation phases. The development of custom models can either be done internally by firms or be led by third-party AI providers. One such provider is Cohere, who has partnerships with Google Cloud and Oracle and focuses on working with enterprise customers rather than individuals. They build LLMs that allow AI to learn from new data and can be customized for specific corporate applications, such as interactive chatbots, search engines, and content generation for websites and marketing campaigns.

5. The AI economy

The term "AI economy" has been increasingly used to describe the economic impact and ecosystem surrounding AI. For example, a report by PwC titled "Sizing the Prize" discusses how AI could contribute up to $15.7 trillion to the global economy by 2030, highlighting the transformative potential of AI across industries and its role in shaping future economic landscapes.[27] Similarly, a McKinsey Global Institute report explores how AI technologies are expected to contribute to economic growth, productivity gains, and job creation across various sectors globally.[28] These insights underscore the widespread recognition of AI's potential to shape economies and drive significant changes in business operations and societal dynamics.

More specifically, the AI economy refers to the ecosystem of economic activities and value creation driven by AI technologies and applications. It encompasses a broad spectrum of industries, from technology giants to startups, all leveraging AI to innovate, streamline processes, and create new business opportunities. At its core, the AI economy revolves around the development and deployment of AI technologies, including machine learning, NLP, computer vision, and robotics. This ecosystem fosters growth in various sectors by enhancing productivity, optimizing resource allocation, and enabling new forms of consumer interaction through personalized services and intelligent automation. It also fuels job creation in AI-related fields, ranging from data scientists and AI researchers to software developers and ethical AI specialists. As AI continues to evolve and permeate industries worldwide, the AI economy represents a transformative force driving global competitiveness and economic growth in the digital age.

5.1. The role of large tech companies

Large tech companies are playing an important role in driving the AI revolution by spearheading groundbreaking research, developing sophisticated AI tools, and democratizing access to AI technologies. As mentioned, companies like Google, through its DeepMind subsidiary, have achieved significant milestones in AI research. Microsoft has also been pivotal in making AI more

[27] https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

[28] https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

accessible through platforms like Azure AI, enabling individual developers and businesses of all sizes to easily integrate AI capabilities into their operations. An interesting initiative by Microsoft is the AI for Good Lab, which focuses on leveraging AI to address some of the world's most pressing challenges, spanning sustainability, humanitarian efforts, and health, with the goal of creating a more equitable and sustainable future.[29] This lab include projects like AI for Earth, which develops AI-driven solutions for environmental conservation, and AI for Accessibility, which aims to enhance the lives of people with disabilities. Additionally, Amazon's AWS offers a wide range of AI services, from NLP to computer vision, empowering businesses to integrate machine learning into their operations seamlessly, from predictive analytics to automated customer service. Meta utilizes AI algorithms for content moderation, using advanced pattern recognition to detect and remove harmful content swiftly, enhancing user safety and experience on its platform. Tesla has pushed boundaries in autonomous driving with its advanced neural networks and distributed fleet-based data capture network, aiming to revolutionize transportation safety and efficiency. IBM's Watson AI has analyzed vast amounts of medical data to support clinicians in making informed decisions, from personalized treatment plans to diagnostic insights. These examples underscore how large tech companies are not just advancing AI capabilities but also applying them across sectors to foster collaboration, innovation, efficiency, and societal benefits globally in diverse applications from healthcare diagnostics to autonomous vehicles. Overall, large companies can build the infrastructure and provide access to everyone via licenses or partnerships.

Large corporations are also paving the way for new companies, business models, and startups to scale and turn themselves into established providers. As discussed, there are numerous partnerships between large tech companies and smaller AI startups (especially those that developed foundation models). Examples, as of today, include the following:[30]

- Amazon has partnerships with AI21 Labs, Cohere, Falcon LLM, Meta, Mistral AI, Runaway AI, and Stability AI; and has both investments and partnerships with Anthropic and Hugging Face.

- Google has partnerships with Cohere, Contextual AI, Falcon LLM, InstaDeep, Meta, Midjourney, Mistral AI, and NVIDIA; and has both investments and partnerships with AI21 Labs, Anthropic, Hugging Face, and Runaway AI.

- Microsoft has partnerships with Falcon LLM, Hugging Face, Meta, NVIDIA, and Perplexity AI; and has both investments and partnerships with Adept AI, Inflection AI, Mistral AI, and OpenAI.

- NVIDIA has partnerships with Amazon, Google, Meta, Microsoft, Mistral AI, and Stability AI; and has both investments and partnerships with Adept AI, Cohere, Hugging Face, Inflection AI, and Recursion.

These types of partnerships allow smaller companies to accrue funding and advance the frontier of AI technology in ways that would not have been possible otherwise. Such partnerships enhance

[29] https://www.microsoft.com/en-us/research/group/ai-for-good-research-lab/

[30] https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf

the traditional venture capital funding approach for startups by offering not only the necessary capital to cover operating expenses but also the technology needed to succeed, such as cloud computing credits and access to infrastructure.

Another important role of large tech companies is to ensure the responsible development and adoption of AI within our society. One such example is the access principles released by Microsoft earlier this year.[31] These principles are commitments to promote innovation and competition in the AI economy that are underpinned by the concepts of access, choice and fairness, and societal responsibility. More specifically, the principles articulate five tenets that define Microsoft's goals as it focuses on AI access, including its role as an infrastructure and platforms provider, as follows: (1) a responsibility to enable innovation and foster competition; (2) responsibilities begin by meeting obligations under the law; (3) the need to advance a broad array of AI partnerships; (4) the commitment to partnership extends to customers, communities, and countries; and (5) the need to be proactive and constructive, as a matter of process, in working with governments and the IT industry in the design and release of new versions of AI infrastructure and platforms. In the same vein, Microsoft released a responsible AI transparency report in May 2024,[32] as well as various other policy implications in terms of privacy, copyright commitment, and advancing safe and trustworthy AI practices.

In a similar spirit, other tech companies have also released various AI principles and ethical guidelines on responsible AI applications, such as Apple's privacy and security policy,[33] AWS's responsible AI policy,[34] Google's AI principles,[35] IBM's AI ethics,[36] Meta's five pillars of responsible AI,[37] and OpenAI's charter.[38] Laying out the foundations for formally defining the ethical principles of responsible and trustworthy AI is a critical aspect of ensuring secure and non-harmful adoption of this technology. Given that this technology is evolving very quickly, tech companies play a key role in laying out these foundations and constantly adjusting them in collaboration with government entities and regulatory bodies.

5.2. Competitive market

The AI market has become fiercely competitive, spanning multiple layers and application domains, driven by several large tech companies, many innovative startups, and established enterprises alike. It is common to split the AI market into different layers, including the following.

At the hardware and data storage layer, companies like AMD, Google (whose chips are primarily designed by Broadcom), Intel, NVIDIA, and Qualcomm offer GPUs optimized for AI and deep learning tasks, which are crucial for training and inference processes in all types of applications.

---

[31] https://blogs.microsoft.com/on-the-issues/2024/02/26/microsoft-ai-access-principles-responsible-mobile-world-congress/

[32] https://www.microsoft.com/en-us/ai/responsible-ai

[33] https://www.apple.com/ca/legal/privacy/en-ww/

[34] https://aws.amazon.com/machine-learning/responsible-ai/policy/

[35] https://ai.google/responsibility/principles/; https://ai.google/responsibility/responsible-ai-practices/

[36] https://www.ibm.com/impact/ai-ethics

[37] https://ai.meta.com/responsible-ai/

[38] https://openai.com/charter/

At the cloud platform and tools layer, services like AWS, Google Cloud, IBM Cloud, and Microsoft Azure offer comprehensive suites of AI services, including machine learning APIs and pre-built models, empowering developers and businesses to efficiently integrate AI into their applications. Sometimes, this layer also includes the development frameworks that facilitate the creation of AI models, such as TensorFlow and PyTorch, as well as libraries and toolkits, such as scikit-learn for machine learning and Hugging Face Transformers for NLP.

At the model layer, we find the core of the AI stack composed of the algorithms that analyze data, learn from it, and make predictions or decisions based on that data. At a high level, the AI models can take several forms, such as generative AI or foundation models (e.g., Claude 3.5, Gemini, Grok, GPT-4, Llama 3, Mistral, and Stable Diffusion), specific AI models (e.g., BERT and FaceNet), and hyperlocal AI models (i.e., specialized models customized for specific application domains or use cases). These models are often distributed via APIs that allow developers to leverage cutting-edge AI capabilities without having to build their own models. API access is usually available directly from developers (e.g., Claude API, Llama API, and OpenAI API) or via providers that offer easy API access to third-party generative AI models, such as Amazon Sagemaker, Google Vertex, Hugging Face, Microsoft Azure, and OpenAI.

Finally, the application layer consists of the countless opportunities for consumer-facing applications. For example, virtual assistants like Alexa and Siri compete for capabilities in NLP and smart home integration,[39] while in e-commerce, companies like Alibaba and Amazon leverage AI for personalized recommendations and predictive analytics to enhance customer engagement and sales. In healthcare, startups like PathAI focus on pathology diagnostics, utilizing deep learning to analyze medical images and aid in disease diagnosis. In marketing, companies like Jasper AI can generate content such as ads, blog posts, website copy, and social media posts.

As illustrated in Figure 2,[40] the AI landscape is extremely crowded, with thousands of companies of various sizes operating across the different layers of the AI market. The number of domain applications and solution providers has grown drastically in the last few years, and especially in the past several months. According to various sources, there are approximately 78,000 AI companies worldwide and over 115 million companies currently using AI.[41] The World Economic Forum predicted that approximately 97 million people will be necessary to fill the work demands of the surging AI industry,[42] and this forecast may need to be adjusted upward due to the recent surge in AI applications. Finally, corporate and private investments in AI, particularly in generative AI, have skyrocketed, as illustrated by data from Stanford's AI Index report.[43]

[39] As of now, Alexa and Siri are not powered by LLMs and instead are using a combination of AI technologies, including NLP, machine learning, and rule-based systems, to understand and respond to user queries. However, these systems continuously evolve by integrating more advanced AI components and may potentially incorporate LLMs in the future.

[40] The intent is not for the reader to read all the names displayed in the figure but instead to illustrate how crowded the current AI market is across the different layers.

[41] https://tracxn.com/d/sectors/artificial-intelligence/__cbMnXfS2GfFo4Vi2dxZyUy7l4O8WyzVYLseb9keW5cI/companies

[42] www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

[43] https://aiindex.stanford.edu/report/

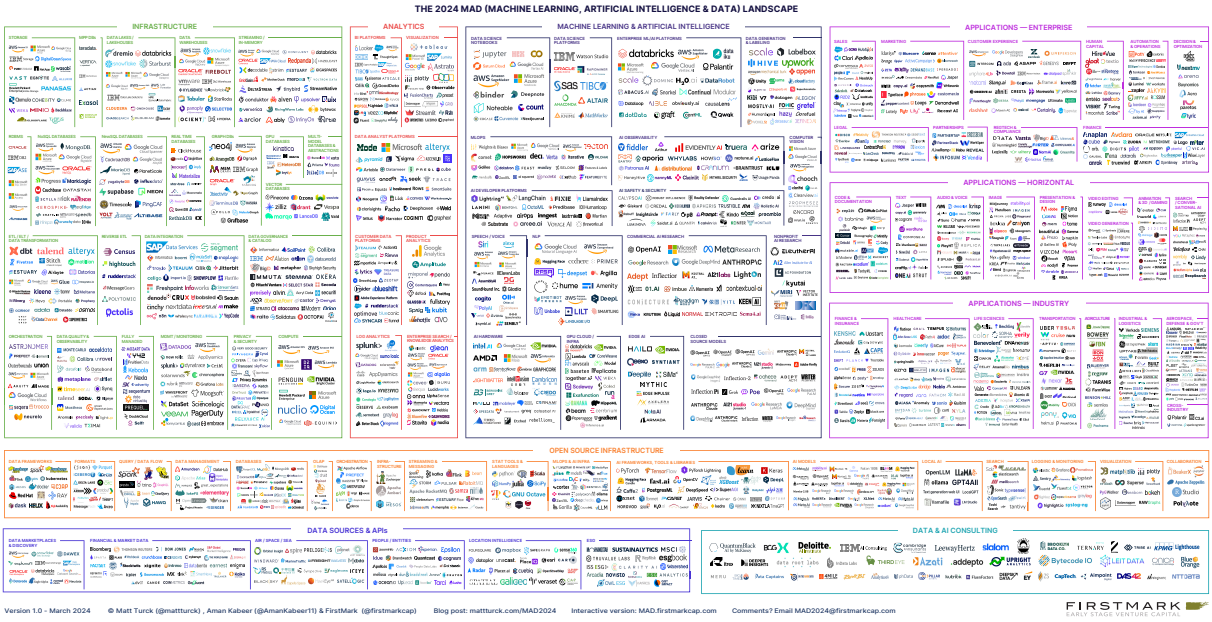

Figure 2. Illustration of the crowded AI landscape (source: https://mattturck.com/landscape/mad2024.pdf, date accessed: August 2024).

6. Conclusion

While AI offers transformative benefits across various sectors, it is crucial to also acknowledge the potential pitfalls, risks, and challenges that accompany its rapid development and deployment. For instance, as discussed, AI systems can perpetuate and amplify biases present in their training data, potentially leading to unfair and discriminatory outcomes in areas like hiring and law enforcement. Additionally, the lack of transparency in AI decision-making processes, often referred to as the "black box" problem, can make it difficult to understand and trust AI-driven outcomes. Other important challenges include preserving the privacy of users (and the potential misuse of personal information), recognizing and being cautious of deepfakes and misinformation (e.g., in the form of hallucination), regulating plagiarism and copyright ownership for AI-created objects,[44] and ensuring that carbon emissions and environmental concerns are considered when developing and deploying AI tools. An additional critical issue is to mitigate security risks, given that AI systems can be vulnerable to adversarial attacks where malicious actors manipulate inputs to deceive the AI, leading to incorrect or harmful outputs. The deployment of AI in areas such as autonomous vehicles and healthcare can also raise ethical questions, including decisions about life and death and the extent of human oversight. Ultimately, it is imperative to protect ourselves against the risk that AI systems do not align with human values and priorities by embedding ethical principles into AI development. These issues underscore the importance of establishing robust regulatory frameworks to ensure that AI technologies are developed and deployed ethically and responsibly. Addressing these challenges requires a multidisciplinary approach, involving technologists, ethicists, policymakers, and society at large to ensure that AI is developed and deployed in a way that maximizes benefits while minimizing risks.

[44] Determining liability in cases where AI systems cause harm or make incorrect decisions can be complex, hence raising interesting new questions about accountability and legal responsibility.

Despite these challenges, this paper has discussed the fact that the benefits of AI are already immense for humanity and our society. If managed correctly, AI can significantly enhance efficiency, productivity, accuracy, and accessibility in fields such as healthcare, environmental conservation, and legal systems. As highlighted in this paper, AI is already making strides in improving mental healthcare, accelerating drug discovery, and revolutionizing wildlife conservation, among many other areas; however, this discussion only scratches the surface of the full potential of AI. Interestingly, a 2024 McKinsey report states that AI is already being used to advance all 17 UN Sustainable Development Goals, ranging from eradicating poverty and ensuring healthy lives and well-being for all at all ages to creating sustainable cities and communities and providing quality education for all.[45]

It is imperative that stakeholders from all sectors work together to navigate the ethical and practical challenges of AI. With careful regulation and continuous dialogue, the positive impacts of AI can be maximized, ultimately outweighing the drawbacks. The journey toward a more AI-integrated future is complex, but with thoughtful consideration and proactive measures, the benefits will be profound and far-reaching.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX symposium on operating systems design and implementation, 265-283.

Adulyasak, Y., Benomar, O., Chaouachi, A., Cohen, M. C., Khern-Am-Nuai, W. (2023). Using AI to detect panic buying and improve products distribution amid pandemic. AI & Society, 1-30.

Agrawal, A., Gans, J., Goldfarb, A. (2022a). Prediction machines, updated and expanded: The simple economics of artificial intelligence. Harvard Business Press.

Agrawal, A., Gans, J., Goldfarb, A. (2022b). Power and prediction: The disruptive economics of artificial intelligence. Harvard Business Press.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems, 1-13.

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine bias. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Babina, T., Fedyk, A., He, A., Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics, 151, 103745.

Bapna, R., Ghose, A. (2024). Thrive: Maximizing well-being in the age of AI. MIT Press.

Bhambhoria, R., Dahan, S., Li, J., & Zhu, X. (2024). Evaluating AI for law: Bridging the gap with open-source solutions. arXiv:2404.12349.

[45] https://www.mckinsey.com/capabilities/quantumblack/our-insights/ai-for-social-good

Binns, R. (2018, January). Fairness in machine learning: Lessons from political philosophy. In Conference on Fairness, Accountability and Transparency. PMLR, 149-159.

Brynjolfsson, E., Li, D., Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.

Cambon, A., Hecht, B., Edelman, B., Ngwe, D., Jaffe, S., ... Teevan, J. (2023). Early LLM-based tools for enterprise information workers likely provide meaningful boosts to productivity. MSR-TR-2023-43.

Cohen, M. C., Tang, C. S. (2024). The role of AI in developing resilient supply chains. Georgetown Journal of International Affairs.

Cohen, M. C., Toubiana, E. (2024). AI medical scribes: A revolution in healthcare documentation. Retrieved from https://maxccohen.github.io/AI-Medical-Scribes.pdf

Czarnitzki, D., Fernández, G. P., Rammer, C. (2023). Artificial intelligence and firm-level productivity. Journal of Economic Behavior & Organization, 211, 188-205.

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In Ethics of data and analytics (pp. 296-299). Auerbach Publications.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Working paper.

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Doshi-Velez, F., Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. arXiv preprint arXiv:2303.10130.

Fitzpatrick, K. K., Darcy, A., Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2), e7785.

Iansiti, M., Lakhani, K. R. (2020). Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. Harvard Business Press.

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. arXiv:1907.11692.

Noy, S., Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. doi:10.1126/science.aax2342.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32.

Peng, S., Kalliamvakou, E., Cihon, P., Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub copilot. arXiv:2302.06590.

Ribeiro, M. T., Singh, S., Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

Tanlamai, J., Khern-am-nuai, W., Cohen, M. C. (2024). Generative AI and price discrimination in the housing market. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4764418

Tierney, A. A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Kipnis, P., ... Lee, K. (2024). Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catalyst Innovations in Care Delivery, 5(3).

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, arXiv preprint arXiv:1706.03762.

Yu, J., Qi, C. (2024). The impact of generative AI on employment and labor productivity. Review of Business, 44(1).