

Committee: Social, Cultural, and Humanitarian Issues Committee (GA3)

Issue: Ethics of AI technologies and their human rights implications

Student Officer: Davidoula Georgopoulou

Position: Deputy President

PERSONAL INTRODUCTION

Dear delegates,

My name is Davidoula Georgopoulou and I am currently attending 11th grade at Arsakeio High School of Ioannina. It is my absolute honor to be serving as a Deputy President of the Social, Cultural, and Humanitarian Committee (GA3) at the 9th Arsakeia-Tositseia Schools Model United Nations conference.

Firstly, I would like to welcome you to the committee and congratulate you on your initiative to participate in this conference. I was introduced to the MUN community two years ago and since then I have fallen in love with its nature: it gives you a more holistic view of the world's problems and it teaches you the process of coming up with possible solutions and defending them, while fostering long-lasting friendships with like-minded peers. You will also acquire certain soft skills that will not only help in your career but also in your everyday life. For instance, you will learn the art of debating and start believing in teamwork, as collaboration always facilitates change.

This is going to be my first time chairing and I am particularly excited that it will be in the GA3 Committee. This year's topics are extremely prominent and through this study guide I will introduce you to the issue of "Ethics of AI technologies and their human rights implications".

This study guide will serve as a basis for your preparation and you can also take advantage of the bibliography, which you will find towards the end of the guide. However, I would like to stress that you should not limit your studying only to this document but also, conduct further research on your own.

Should you have any questions or need any help in general, do not hesitate to contact me at: davidoula.mg@gmail.com. I cannot wait to meet you all in November. I wish you a fruitful debate and good luck with your research!

Yours truly,

Davidoula Georgopoulou

INTRODUCTION

Artificial Intelligence (AI) is a transformative force that redefines the boundaries of human capability and potential. For instance, AI-powered recommendation systems influence the content we consume on streaming platforms, the products we buy online, and even the news we read. Virtual assistants like Siri, Alexa, and Google Assistant have become ubiquitous, providing us with instant information and performing tasks on our behalf. In customer service, chatbots efficiently handle our inquiries without us even realizing that we are interacting with a machine.

AI has become so integrated into our daily lives that people often tend to overlook its presence and stop thinking of it as AI. It is so pervasive that many remain unaware of both its impact and our reliance upon it.

Despite the robust opportunities provided, these rapid changes also raise profound ethical concerns and can seriously threaten human rights. Such risks of AIs have already begun to intensify beyond existing inequalities and further harm already marginalized and vulnerable groups.

With each step forward in technological innovation, we must answer a myriad of questions: How does the development of AI impact our right to privacy, autonomy, and freedom of expression? What are the ethical considerations surrounding AI in areas such as healthcare, criminal justice, and social welfare? How can we mitigate algorithmic bias and discrimination and ensure that AI systems operate with fairness and justice? Is AI in the best interests of humanity and human well-being?

In no other field is the ethical compass more relevant than in artificial intelligence. AI poses risks if not accompanied by proper ethical safeguards. Without these guardrails, AI will perpetuate biases and real-world discrimination, exacerbate divisions, and pose threats to fundamental human rights and freedoms. While human rights may not provide all the answers for every specific AI-related challenge, they establish a solid foundation of ethical principles that should guide decision-making and actions in the field of AI.

DEFINITION OF KEY TERMS

Artificial Intelligence (AI)

Artificial intelligence (AI) is the simulation of human intelligence using software-coded heuristics. It is an interdisciplinary field that merges computer science with robust datasets to facilitate problem-solving. It also encompasses the development of algorithms that enable expert systems to make predictions or classifications by analyzing input data. ¹

Human Rights

Human rights are inherent rights possessed by every individual simply by being human. They encompass a wide spectrum of moral principles, from the most fundamental -the right to life-, to those that enhance the quality of life, including the rights to education, employment, healthcare, and freedom among others. The Universal Declaration of Human Rights (UDHR) has declared 30 articles that serve as the foundation for existing and forthcoming human rights treaties and conventions. ²

¹Frankenfield, Jake. "Artificial Intelligence: What It Is and How It Is Used." *Investopedia*, Apr. 2023, www.investopedia.com/terms/a/artificial-intelligence-ai.asp.

²OHCHR. "What Are Human Rights?" *OHCHR*, www.ohchr.org/en/what-are-human-rights.



Figure 1: 30 Basic Human Rights List ³

Accountability in AI

Accountability in AI relates to the expectation that designers, developers, and deployers will comply with standards and legislation to ensure the proper functioning of AIs. However, the outcome of AIs is often opaque and unpredictable, hindering the detection of causes and reasons for unintended results. ⁴

Black Box AI

Black box AI models arrive at decisions or conclusions without offering any explanations for their outcomes. These models utilize intricate networks of artificial neurons, spreading data and decision-making processes across tens of thousands of

³ Venngage. "Human Rights Template." Venngage, venngage.com/templates/infographics/human-rights-template-7d71ff49-ab13-4aae-9325-5dd1fb64accf.

⁴ Accountability (OECD AI Principle) - OECD.AI. oecd.ai/en/dashboards/ai-principles/P9.

neurons, resulting in a complexity that may be just as difficult to understand as that of the human brain.⁵

AI Transparency

AI transparency refers to making the algorithms and decision-making processes understandable. AI Transparency builds trust in AI systems, giving users and stakeholders greater confidence that the system is being used appropriately.⁶

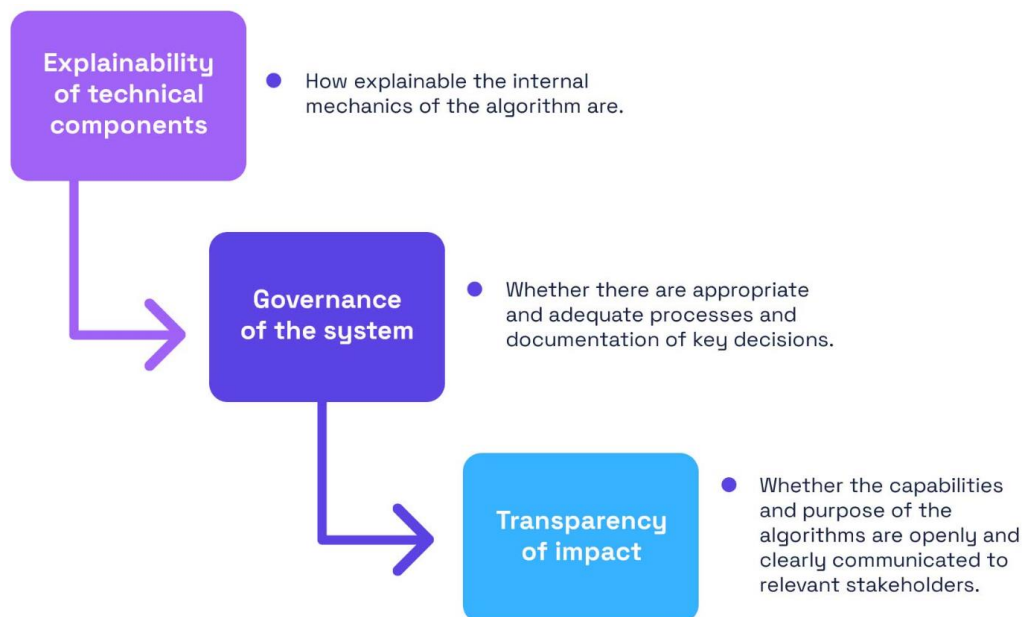


Figure 2: Principles of AI Transparency⁷

Privacy

Privacy is the right of individuals to control access to their personal information and to be free from intrusion into their private lives. In the context of AI, privacy concerns arise from the collection, use, and storage of personal data by AI systems. It is essential to ensure that AI systems do not manipulate individuals or discriminate

⁵ Yasar, Kinza, and Ivy Wigmore. "Black Box AI." *WhatIs.com*, Mar. 2023, www.techtarget.com/whatis/definition/black-box-AI.

⁶ *What Is AI Transparency?* www.holisticaai.com/blog/ai-transparency#:~:text=A%20major%20motivation%20for%20AI,system%20is%20being%20used%20appropriately.

⁷ *What Is AI Transparency?* www.holisticaai.com/blog/ai-transparency.

against them based on their personal data. Privacy remains one of the most recurrent concerns about AI technologies.^{8 9}

Big Data

Big Data refers to gathering and analyzing vast amounts of data to understand and predict human behavior. Big data or predictive analytics is a new field that relies on sophisticated algorithms to process millions of data points and then, make accurate predictions about future human actions and trends.¹⁰

AI Winter

“In the history of AI, an AI winter is a period of reduced funding and interest in artificial intelligence research.”¹¹

Deep Learning

Deep learning is a machine learning technique that teaches computers by example, meaning they are trained to perform classification tasks directly from provided images, text, or sound. For example, deep learning is used for driverless cars, enabling them to recognize a stop sign or to distinguish a pedestrian from a lamppost. However, deep learning models are only as good as the data they are trained on. If

⁸ Elliott, David, and Eldon Soifer. “AI Technologies, Privacy, and Security.” *Frontiers in Artificial Intelligence*, vol. 5, Frontiers Media, Apr. 2022, <https://doi.org/10.3389/frai.2022.826737>.

⁹ Van Rijmenam Csp, Mark. “Privacy in the Age of AI: Risks, Challenges and Solutions.” *Dr Mark Van Rijmenam, CSP | Strategic Futurist Speaker*, Apr. 2023, www.thedigitalspeaker.com/privacy-age-ai-risks-challenges-solutions/#:~:text=In%20the%20context%20of%20AI,based%20on%20their%20personal%20data.

¹⁰ Emeritus. (2023b). Big Data vs Data Analytics: key differences. *Emeritus - Online Certificate Courses | Diploma Programs*. <https://emeritus.org/in/learn/big-data-vs-data-analytics/#:~:text=Type%20of%20Data%20in%20Big,analytics%20is%20mostly%20structured%20data>.

¹¹ *January 2005 - AI Newsletter*. (2005, January 1). https://www.ainewsletter.com/newsletters/aix_0501.htm#w

the training data contains biases, such as racial or gender biases, the AI system will perpetuate these biases.¹²

Neural Networks

Neural networks are the backbone of deep learning algorithms. They are called “neural” because they mimic how neurons in the brain signal one another and communicate. They are designed to recognize patterns and relationships in data, enabling them to perform tasks such as decision-making.

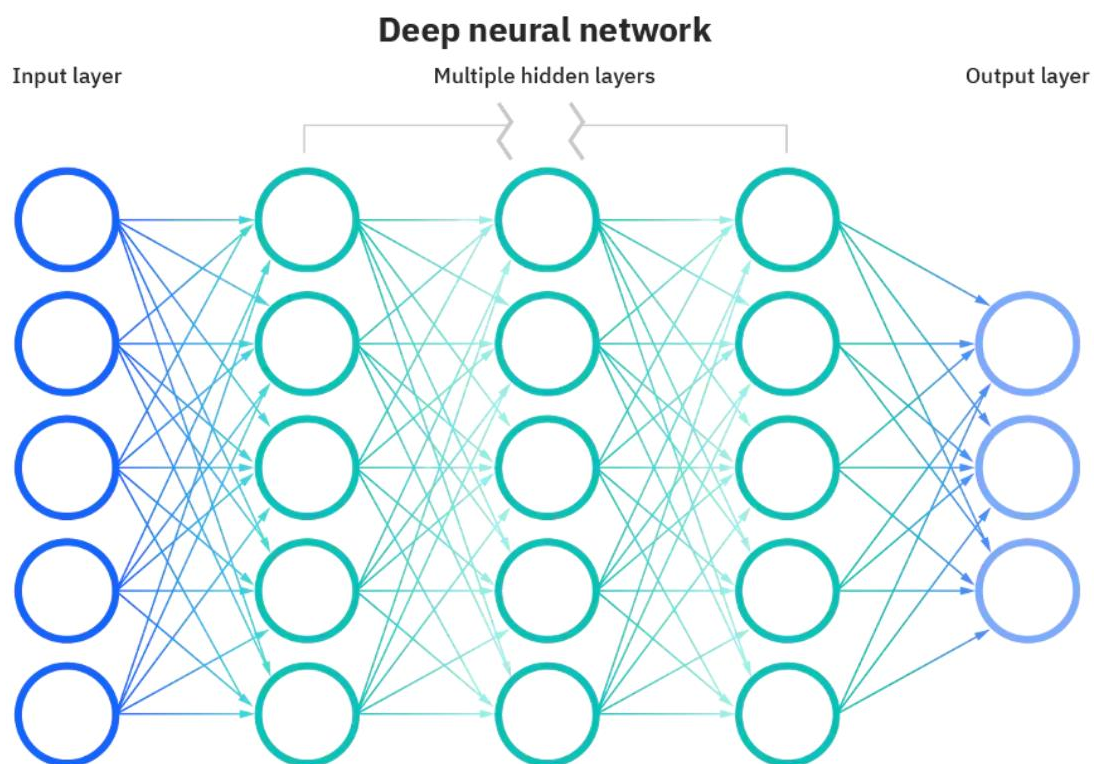


Figure 3: Visualization of a Deep Neural Network ¹³

Data Breach

¹² What is deep learning? | How it works, techniques & applications. (n.d.). MATLAB & Simulink. <https://www.mathworks.com/discovery/deep-learning.html#:~:text=Deep%20learning%20is%20a%20machine,a%20pedestrian%20from%20a%20lamppost>.

¹³ Team, I. D. a. A., & Team, I. D. a. A. (2023). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the difference? *IBM Blog*. <https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/>

A data breach refers to an incident where sensitive, confidential, or protected information is stolen by unauthorized individuals through cyberattacks or other security vulnerabilities. Data breaches can affect various types of data, like personal information, financial data, medical records, login credentials, intellectual property, and more.¹⁴

Human-Centric AI

Human-centered AI learns from human input and collaboration, focusing on algorithms that exist in a larger, human-based system. Human-centered AI is defined by systems that are continuously improving because of human input while providing an effective experience between humans and robots. The approach develops and deploys AI technologies that prioritize human values, well-being, and rights. Human-centric AI aims to ensure that AI systems are designed and used in ways that align with human interests and contribute to the common good.¹⁵

BACKGROUND INFORMATION

A Brief History of AI

When we talk about the history of AI we must refer to the 1940s, the Second World War, and the advent of the first computers. It was at that time that the idea of the human-machine analogy was born, that human intelligence could be simulated through the use of machines. The British scientist Alan Turing is considered one of the fathers of modern computer science, as he fundamentally contributed to the success of the Allies in the Second World War by decoding the Enigma device and cracking the German military transmissions.

The term "artificial intelligence" was coined in 1956 at the Dartmouth Conference, where researchers gathered to discuss the possibility of creating

¹⁴ *Data breach.* (n.d.).

<https://www.trendmicro.com/vinfo/us/security/definition/data-breach>

¹⁵ "Human-centered Artificial Intelligence." *www.cognizant.com*,
www.cognizant.com/us/en/glossary/human-centered-ai#:~:text=What%20is%20human%2Dcentered%20AI,experience%20between%20human%20and%20robot.

machines with human-like intelligence.¹⁶ After initial enthusiasm in the early years of AI research, progress faced setbacks and challenges during the 1970s-1980s. This period is commonly referred to as the "AI winter" due to dwindling funding and interest caused by unrealistic expectations and limitations in computing power and algorithms. The industry experienced a second AI winter in the mid-1990s, as progress hit roadblocks again.¹⁷

However, the issues were soon resolved with a series of innovations. For instance, neural networks, inspired by the structure of the human brain, promised to push the limits of AI and gained popularity. Advances in algorithms and computing power enabled neural networks to tackle tasks like image recognition and natural language processing. AI research also shifted its focus towards solving complex problems in areas like medicine and finance.¹⁸

In the 2000s, the rise of the Internet and the availability of vast amounts of data led to a focus on machine learning techniques like supervised learning, and companies like Google, Facebook, and others started leveraging AI. Deep learning, a subset of machine learning using deep neural networks, made remarkable advancements in computer vision, speech recognition, and more. Breakthroughs in hardware, such as Graphics Processing Units (GPUs), accelerated deep learning training. AI became increasingly integrated into various aspects of society, including

¹⁶ Potenza, Davide. "AI History: The Dartmouth Conference." *Klondike*, June 2023, www.klondike.ai/en/ai-history-the-dartmouth-conference/#:~:text=1956%20thus%20marks%20the%20official,proposed%20to%20call%20Artificial%20Intelligence.

¹⁷ Glover, Ellen. "What Is AI Winter?" *Built In*, Dec. 2022, builtin.com/artificial-intelligence/ai-winter.

¹⁸ Jaspreet. "A Concise History of Neural Networks - Towards Data Science." *Medium*, 19 Sept. 2022, towardsdatascience.com/a-concise-history-of-neural-networks-2070655d3fec#:~:text=The%20idea%20of%20neural%20networks,McCulloch%20and%20mathematician%20Walter%20Pitts.

self-driving cars, virtual assistants, and healthcare applications. Ethical considerations surrounding AI, such as bias, transparency, and accountability, gained prominence.¹⁹

Privacy and Data Protection:

AI is often data-driven, meaning algorithms use vast amounts of data to train and improve their performance. The accumulation of large datasets increases the risk of data breaches, since AI systems may inadvertently or intentionally misuse personal data, leading to privacy violations and potential harm to individuals.

For instance, there is no question that digital platforms such as TikTok, Facebook, YouTube, Google, and Amazon have accumulated rich datasets for every account holder. Users provide on social media platforms extensive information voluntarily, including their interests, preferences, locations, and many more. This vast amount of user-generated data is valuable to companies as it allows them to target specific audiences with personalized content and advertisements. This practice, also known as surveillance advertising or behavioral advertising, is deeply harmful to privacy and the flow of information. Users may not always be fully aware of the extent of data collection and how their information is being used. This data can be vulnerable to hacking, misuse, or lead to potential identity theft.^{20 21}

Experts have also raised concerns about the ethical implications of healthcare data storage and data security practices for years, and AI is taking up a larger share of that conversation. A good illustration of this point is Facebook since it uses AI to store and act on users' mental health data with no legal safeguards in place. In late 2017, Facebook rolled out a "suicide detection algorithm" in an effort to promote suicide

¹⁹ "What Is the History of Artificial Intelligence (AI)?" *Tableau*, www.tableau.com/data-insights/ai/history#:~:text=1952%3A%20A%20computer%20scientist%20named,it%20came%20into%20popular%20usage.

²⁰ Hogue, Preston. "Terms of Use: User Privacy and the Algorithms Behind Social Media." *SecurityWeek*, Dec. 2020, www.securityweek.com/terms-use-user-privacy-and-algorithms-behind-social-media.

²¹ "Social Media Privacy." *EPIC - Electronic Privacy Information Center*, epic.org/issues/consumer-privacy/social-media-privacy.

awareness and prevention. The system uses AI to gather data from posts and then predict people's mental state and propensity to commit suicide. Of course, this is ostensibly a positive use case for AI in healthcare. But benevolent intent aside, the fact remains that Facebook is gathering and storing users' mental health data without their consent.²²

Maintaining absolute anonymity on the Internet remains a concept akin to a superpower. This is notably evident as organizations employ big data analytics, even the mere notion of having anonymized data files becomes impossible. Even in the rare occurrence of a data file being made to be completely "anonymous," several security teams can combine these valued files with others, making the task of identifying an individual quite easy. In an attempt to safeguard their confidential data from hackers and cybercriminals, a majority of organizations opt for the practice of "Data masking." As its name implies, data masking, sometimes referred to as data obfuscation, involves the meticulous process of camouflaging authentic data using less significant elements or datasets. In other words, real data is hidden by other less important characters or data sets. Again, this does not necessarily ring true for big-data-driven settings.²³

Last but not least, ChatGPT, a language model, is another example of AI with possible implications for human rights. ChatGPT can generate human-like text based on a given input. One of the primary concerns surrounding ChatGPT is its potential to be weaponized for disinformation. In recent years, the world has witnessed an alarming rise in the spread of fake news and propaganda, which has had severe consequences on political stability and social cohesion. With ChatGPT's ability to generate realistic and coherent text, the technology can easily create convincing yet false narratives. Moreover, the data ChatGPT was trained on is now the subject of a new lawsuit against OpenAI. In a class action lawsuit filed on June 28 against OpenAI and its partner Microsoft, it is claimed that OpenAI used stolen data to train and

²² Richards, S., & Richards, S. (2022, July 11). AI in Healthcare: Data Privacy and Ethics Concerns - Lexalytics. *Lexalytics*. <https://www.lexalytics.com/blog/ai-healthcare-data-privacy-ethics-issues/>

²³ cyberinsiders. "What Are the Biggest Privacy Issues Associated with Big Data?" *Cybersecurity Insiders*, 20 Apr. 2020, www.cybersecurity-insiders.com/what-are-the-biggest-privacy-issues-associated-with-big-data/.

develop its products. The lawsuit claims that OpenAI stole data from "millions of unsuspecting consumers worldwide". Furthermore, the lawsuit claims that OpenAI is "harvesting massive amounts of personal data from the internet" such as private conversations, medical data, and more, without asking for users' permission. ^{24 25}

Bias and Discrimination

When AI makes headlines, most of the time it's because of problems with bias and fairness. Some of the most infamous issues have to do with facial recognition, policing, and health care, but across many industries and applications, there are missteps where machine learning is contributing to creating a society where some groups or individuals are disadvantaged.

One well-known example is the COMPAS system used in the United States criminal justice system, which predicts the likelihood of a defendant reoffending. A study by ProPublica found that the system was biased against African-American defendants, as they were more likely to be labeled as high-risk even if they had no prior convictions. ²⁶

Within the healthcare sector, an AI system designed to forecast patient mortality rates came under scrutiny due to inherent biases against African-American patients. Investigations revealed that the system exhibited a tendency to attribute elevated risk scores to African American patients, even when variables like age and health condition were identical. This bias can lead to African-American patients being denied access to healthcare or receiving subpar treatment.

Another example of bias in AI systems is the facial recognition technology used by law enforcement agencies. It was found that facial recognition technology was

²⁴ Legal IT Insider. (2023b). OpenAI and Microsoft face class action lawsuit for allegedly violating copyright and privacy laws. *Legal IT Insider*. <https://legaltechnology.com/2023/06/29/openai-and-microsoft-face-class-action-lawsuit-for-allegedly-violating-copyright-and-privacy-laws/>

²⁵ Ortiz, S. (2023, June 29). OpenAI sued for "stealing" data from the public to train ChatGPT. *ZDNET*. <https://www.zdnet.com/article/openai-sued-for-stealing-data-from-the-public-to-train-chatgpt/>

²⁶ Angwin, Julia, et al. "Machine Bias *." *Auerbach Publications eBooks*, 2022, pp. 254–64. <https://doi.org/10.1201/9781003278290-37>.

significantly less accurate for people with darker skin tones, leading to higher rates of false positives. This bias can have serious consequences, such as wrongful arrests or convictions.²⁷

Type of Bias	Description	Examples
Sampling Bias	Occurs when the training data is not representative of the population it serves, leading to poor performance and biased predictions for certain groups.	A facial recognition algorithm trained mostly on white individuals that performs poorly on people of other races.
Algorithmic Bias	Results from the design and implementation of the algorithm, which may prioritize certain attributes and lead to unfair outcomes.	An algorithm that prioritizes age or gender, leading to unfair outcomes in hiring decisions.
Representation Bias	Happens when a dataset does not accurately represent the population it is meant to model, leading to inaccurate predictions.	A medical dataset that under-represents women, leading to less accurate diagnosis for female patients.
Confirmation Bias	Materializes when an AI system is used to confirm pre-existing biases or beliefs held by its creators or users.	An AI system that predicts job candidates' success based on biases held by the hiring manager.
Measurement Bias	Emerges when data collection or measurement systematically over- or under-represents certain groups.	A survey collecting more responses from urban residents, leading to an under-representation of rural opinions.
Interaction Bias	Occurs when an AI system interacts with humans in a biased manner, resulting in unfair treatment.	A chatbot that responds differently to men and women, resulting in biased communication.

Table 1: Different Types of Bias ²⁸

Accountability:

AI systems are not simply conventional tools or traditional social structures. Their technological characteristics often introduce opacity and unpredictability into their outcomes, making it difficult to pinpoint the causes behind unintended results. Multiple elements contribute to AI systems yielding unfavorable outcomes. The inherent nature of AI systems presents challenges in attributing accountability for these consequences. To illustrate, determining responsibility for medical errors or incorrect diagnoses made by AI-driven diagnostic systems can be intricate. Similarly, AI-powered drones with self-governing flight capabilities raise queries about liability in situations involving accidents or violations of privacy regulations.

²⁷ Hall, Reva Schwartz Apostol Vassilev, Kristen K. Greene, Lori Perine, Andrew Burt, Patrick. "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence | NIST." *NIST*, Mar. 2023, www.nist.gov/publications/towards-standard-identifying-and-managing-bias-artificial-intelligence.

²⁸ <https://arxiv.org/ftp/arxiv/papers/2304/2304.07683.pdf>

Features	Examples		
Context (what for?)	Electoral	Juridical	Administrative
Range (about what?)	Choices of political direction, laws, and recruitment	Conducts, omissions, and decisions	Policy implementation
Agent (who?)	Representatives, leaders, parties, governments, and institutional bodies	Natural persons, legal persons, states, and assets	Public officials and institutions
Forum (to whom?)	Citizens, voters, taxpayers, political parties, and institutions	Individual and collective entities (including states and courts)	Citizens, auditors, inspectors, and ombudsman
Standard (according to what?)	Reliability, coherence, and ideology	Legal rules, principles, and precedents	Efficiency, effectiveness, and legal norms
Process (how?)	Public debate (media), internal or external vigilance (e.g., judicial review), and elections	Judicial and extra-judicial review	Auditing, internal supervision, and judicial review
Implications (what follows?)	Electoral outcomes, political reputation, careers, and funding	Reparations, remands, detentions, fines, and prohibitions	Certifications, validations, revocations, penalties, suspensions, and seizures

Table 2: Examples of Accountability Types ²⁹

MAJOR COUNTRIES AND ORGANISATIONS INVOLVED

United States of America

The U.S. government, research institutions, and industry leaders have been actively developing AI ethics guidelines and principles. For instance, The U.S. has implemented several privacy laws and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) and the California Consumer Privacy Act (CCPA), to safeguard individuals' data rights and privacy in the context of AI. Moreover, in various sectors, including healthcare, education, and public services, the United States has been exploring the potential of AI for social good. Efforts are being made to utilize AI technologies to address societal challenges while ensuring that human rights, such as access to quality healthcare and education, are upheld. ^{30 31}

China

One of the main concerns is the extensive use of AI-powered surveillance and facial recognition technology for social control. The Chinese government has implemented a vast network of surveillance cameras equipped with AI algorithms to monitor citizens' movements and behaviors. This mass surveillance has implications

²⁹ <https://arxiv.org/ftp/arxiv/papers/2304/2304.07683.pdf>

³⁰ CCPA. theccpa.org.

³¹ *Health Insurance Portability and Accountability Act of 1996 (HIPAA) | CDC.*
www.cdc.gov/phlp/publications/topic/hipaa.html.

for privacy and freedom of expression, raising questions about individual rights and government intrusion into personal lives. Moreover, there are concerns about the use of AI-driven technologies for censorship and online surveillance, leading to limitations on freedom of information and expression. China's Great Firewall is an example of how AI is used to control and censor online content, restricting access to information and suppressing dissenting voices.³²

European Union

As part of its digital strategy, the European Union (EU) aims to establish regulations for AI to ensure responsible and beneficial development. In April 2021, the European Commission proposed the first-ever EU regulatory framework for AI, which involves analyzing and categorizing AI systems based on the risks they present to users. These varying risk levels will determine the extent of regulation applied. Once approved, these regulations will become the world's first set of rules on AI. The primary objective of the EU Parliament is to guarantee that AI systems utilized within the EU adhere to strict safety, transparency, traceability, non-discrimination, and environmental standards. Additionally, the involvement of human oversight, rather than complete automation, will be emphasized to avert any harmful consequences. On 14 June 2023, the EU Parliament finalized its negotiating position on the AI Act. The subsequent phase will involve discussions with EU member countries in the Council to shape the final version of the law. The target is to reach a consensus by the end of this year.³³

³² https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/snapshot_vi-_countering_the_rise_of_digital_authoritarianism_0.pdf

³³ *EU AI Act: First Regulation on Artificial Intelligence | News | European Parliament.* 6 Aug. 2023, www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?&at_campaign=20226-Digital&at_medium=Google_Ads&at_platform=Search&at_creation=RSA&at_goal=TR_G&at_advertiser=Webcomm&at_audience=ai%20europe&at_topic=Artificial_intelligence_Act&at_location=GR&gclid=CjwKCAjwtuOIBhBREiwA7a_gf1oxAWtyNRldcKfhv-Kum7PBERAxhURR8Clh5XvPBXClcyQ1GRVYgGB0CeYMQAvD_BwE.

Partnership of AI

Partnership on AI involves 17 countries and is a non-profit community of academic, civil society, industry, and media organizations addressing the future of AI. The organization's mission is to ensure that AI is developed and deployed in a manner that respects and upholds human rights, fosters transparency, promotes accountability and addresses potential biases and discrimination. This organization encourages its members to openly share information about their AI technologies, datasets, and algorithms to promote greater understanding and public scrutiny. The organization actively seeks input from individuals and communities impacted by AI systems to understand their needs, concerns, and preferences. Moreover, the Partnership on AI is dedicated to addressing potential biases and discrimination by promoting research.³⁴

Future of Life Institute

The Future of Life Institute (FLI) is a non-profit organization dedicated to addressing global challenges, particularly those posed by emerging technologies, including artificial intelligence (AI). FLI supports research and initiatives that focus on AI safety, aiming to ensure that AI systems are robust and secure, minimizing the risk of unintended consequences that could infringe upon human rights. It is registered in Pennsylvania, with offices in Campbell, US, and Brussels, Belgium and members from the US, Europe, Latin America and Asia and gives grants to researchers from all over the world.³⁵

³⁴ Home - Partnership on AI." Partnership on AI, 13 July 2023, partnershiponai.org.

³⁵ Future of Life Institute. "Home - Future of Life Institute." *Future of Life Institute*, 13 Mar. 2023, futureoflife.org.

TIMELINE OF EVENTS

Date	Description of Event
2011	The United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression releases a report highlighting the importance of protecting human rights in the digital age, including the implications of AI on freedom of expression. ³⁶
2014	The European Court of Justice recognizes the "right to be forgotten" in a landmark ruling: questions about the balance between privacy and freedom of expression in the context of AI-driven data processing. ³⁷
February 2019	The United Nations published ethical guidelines for AI, emphasizing principles such as transparency, accountability, and non-discrimination, to ensure the development and deployment of AI technologies that respect human rights. ³⁸
April 2019	Algorithmic Accountability Act: Introduced in the United States, this bill aims to address biases and discrimination in AI systems by requiring companies to assess and mitigate the impacts of their algorithms. ³⁹

³⁶ Frank, La Rue. "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue :: Addendum." *United Nations Digital Library System*, 27 May 2011, digitallibrary.un.org/record/706200?ln=en.

³⁷ "ENISA Statement Regarding the Landmark Online Privacy ECJ-ruling on Google Search Engine and the Right to Be Forgotten." *ENISA*, www.enisa.europa.eu/news/enisa-news/enisa-statement-regarding-the-landmark-online-privacy-ecj-ruling-on-google-search-engine-and-the-right-to-be-forgotten.

³⁸ *Shaping Europe's Digital Future*, 8 Apr. 2019, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

³⁹ [https://www.congress.gov/bill/116th-congress/house-bill/2231#:~:text=Introduced%20in%20House%20\(04%2F10%2F2019\)&text=This%20](https://www.congress.gov/bill/116th-congress/house-bill/2231#:~:text=Introduced%20in%20House%20(04%2F10%2F2019)&text=This%20)

2016	AI Now Symposium/ Annual Events: Organized by the AI Now Institute at New York University, this symposium critically examines the social implications of AI, bringing together scholars, activists, and policymakers to discuss issues such as bias, discrimination, privacy, and accountability in AI systems. ⁴⁰
2023 (Frequent Events)	Global Summit on AI for Good: Organized by the International Telecommunication Union (ITU), this summit focuses on leveraging AI for global challenges while upholding human rights and ethical considerations. ⁴¹

PREVIOUS ATTEMPTS TO SOLVE THE ISSUE

UN Attempts

The UN Guiding Principles (UNGPs)

The UN Guiding Principles (UNGPs) on Business and Human Rights, adopted by the UN Human Rights Council on 16 June 2011, are a set of 31 principles directed at governments and businesses that clarify their duties in the context of business operations. While the principles do not explicitly mention AI technology, they can be applied to AI and its impact on human rights, as AI is becoming increasingly relevant in business operations, and its deployment can have significant implications for human rights. ⁴²

bill%20requires%20specified%20commercial,artificial%20intelligence%20or%20mac hine%20learning.

⁴⁰ AI Now Institute. "AI Now Symposia Archives." *AI Now Institute*, ainowinstitute.org/series/ai-now-symposia.

⁴¹ "AI for Good Global Summit - IoT for All." *IoT for All*, 1 June 2023, www.iotforall.com/events/ai-for-good-global-summit#:~:text=This%20year%2C%20the%20highly%20anticipated,provide%20invaluable%20insights%20and%20discussions.

⁴² "Guiding Principles on Business and Human Rights | United Nations Development Programme." *UNDP*, www.undp.org/laopdr/publications/guiding-principles-business-and-human-rights.

The mandate of the Special Rapporteur

The mandate of the Special Rapporteur was created by the Human Rights Council to protect and promote freedom of opinion and expression, offline and online, in light of international human rights laws and standards. It also examines the impact of AI on freedom of expression and access to information. The mandate was renewed again for an additional three years in March 2023.⁴³

The Secretary's General Roadmap for Digital Cooperation

On 11 June 2020, a set of recommended actions was presented to the international community to help ensure all people are connected, respected, and protected in the digital age. The Secretary-General's Roadmap for Digital Cooperation is the result of a multi-year, multi-stakeholder, global effort to address a range of issues related to the Internet, AI, and other digital technologies. The panel's report emphasized the need for human rights to be at the center of digital cooperation efforts.⁴⁴

Non-UN Attempts

European Union (EU)

The General Data Protection Regulation (GDPR), enforced by the EU, is the strongest privacy and security law in the world and establishes comprehensive data protection and privacy rules. It addresses the potential risks and rights violations associated with AI systems that process personal data. The GDPR also defines individuals' fundamental rights in the digital age, the obligations of those processing data, methods for ensuring compliance, and sanctions for those in breach of the rules.

⁴³ OHCHR. "Special Rapporteur on Freedom of Opinion and Expression." *OHCHR*, www.ohchr.org/en/special-procedures/sr-freedom-of-opinion-and-expression#:~:text=Purpose%20of%20the%20mandate&text=The%20mandate%20of%20the%20Special,human%20rights%20law%20and%20standards.

⁴⁴ United Nations. "Secretary-General's High-level Panel on Digital Cooperation | United Nations." *United Nations*, www.un.org/en/sg-digital-cooperation-panel.

This regulation updated and modernized the principles of the 1995 data protection directive. It was adopted in 2016 and entered into application on 25 May 2018.⁴⁵

NATO (North Atlantic Treaty Organization)

The NATO Cooperative Cyber Defense Centre of Excellence (CCDCOE), a NATO-affiliated institution, conducts research and hosts discussions on various aspects of cybersecurity, including ethical considerations. Their mission is to support the member nations and NATO with unique interdisciplinary expertise in the field of cyber defense research, training, and exercises covering the focus areas of technology, strategy, operations, and law. While not AI-specific, their work contributes to the broader discussions on cybersecurity and human rights in the digital age. The report “Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States’ Strategies and Deployment” offers a high-level view of the role of AI-enabled and autonomous technologies in the militaries of NATO Allies as of January 2021. It is the first academic work of this kind that focuses specifically on military AI in NATO countries.⁴⁶

European Commission

On 8 April 2019, Ethics Guidelines for Trustworthy Artificial Intelligence were presented by the European Commission, emphasizing human-centricity, transparency, fairness, and accountability. The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. Those were: Human agency and oversight, Technical Robustness and safety, Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental well-being, and Accountability.⁴⁷

⁴⁵ EUR-Lex - 32016R0679 - EN - EUR-Lex. eur-lex.europa.eu/eli/reg/2016/679/oj.

⁴⁶ CCDCOE. ccdcOE.org/library/publications/artificial-intelligence-and-autonomy-in-the-military-an-overview-of-nato-member-states-strategies-and-deployment.

⁴⁷ “Ethics Guidelines for Trustworthy AI.” *Shaping Europe’s Digital Future*, 8 Apr. 2019, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

POSSIBLE SOLUTIONS

Fairness and Bias

First and foremost, to ensure fairness and reduce bias in AI systems, it is essential to start with a foundation of diverse and representative data. By using datasets that accurately reflect the demographics and characteristics of the population it serves, developers can reduce biases that might arise from skewed or incomplete data. This approach helps in creating AI models that deliver equitable outcomes for all users, regardless of their background.

Bias detection and mitigation are crucial steps in the development of AI systems. Implementing algorithms and techniques to identify and address biases is essential to avoid perpetuating unfair or discriminatory results. Regular assessments of the AI model's performance can also help identify any unintended biases and allow for necessary corrective measures to be taken promptly.⁴⁸

Transparency and Explainability

Proper safeguards and guidelines need to be in place to ensure that AI systems are transparent and explainable to the extent required for ethical and responsible use, such as Interpretable AI. Interpretable AI can describe its internal processes in a user-friendly manner, but it may compromise completeness to align with users' cognitive abilities and knowledge. Thus, striking the right balance between transparency and other considerations is crucial, and it requires a thoughtful and context-specific approach.⁴⁹

Promoting public understanding of AI and data is essential in creating a transparent and accountable AI ecosystem. This can be achieved through open and accessible education, civic engagement, digital skills training, and AI ethics programs.

⁴⁸ "Tackling Bias in Artificial Intelligence (and in Humans)." *McKinsey & Company*, 6 June 2019, www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans.

⁴⁹ Novelli, Claudio, et al. "Accountability in Artificial Intelligence: What It Is and How It Works." *AI & Society*, Springer Nature, Feb. 2023, <https://doi.org/10.1007/s00146-023-01635-y>.

Media and information literacy initiatives also play a crucial role in ensuring that the public is well-informed about AI's capabilities, limitations, and potential implications. By fostering a knowledgeable and engaged public, the conversation around AI can become more inclusive and democratic, and decisions regarding AI can be made with greater awareness and consensus.⁵⁰

Accountability and responsibility

AI systems should prioritize traceable features, incorporating mechanisms for oversight, impact assessment, audit, and compliance to ensure alignment with human rights norms. This goal defines the design and development standards to be met throughout the entire lifecycle of an AI, but it is rather generic if it is not implemented by good practices.

Transparency is vital in this process. AI systems must provide clear and comprehensible information to individuals about how their personal data will be used, along with the rights they have regarding their data, such as access, correction, erasure, and restricting its processing. Establishing effective processes for responding to individuals' data access and correction requests is crucial to maintaining data integrity. Moreover, there should be an efficient system to handle complaints or concerns related to data protection.⁵¹

Enforcement further strengthens accountability by linking the monitoring and evaluation of AI performance to formal or informal consequences, ensuring that AI systems remain responsible and compliant with established norms.

Technological and organizational peculiarities of AIs also call for a sociotechnical approach to accountability, a point insufficiently addressed in the relevant debate. The sociotechnical approach emphasizes the interconnectedness and interdependence of social and technical factors in designing and implementing systems, including AI systems. It recognizes that technology is not developed or used

⁵⁰ "Ethics of Artificial Intelligence." *UNESCO*, June 2023, www.unesco.org/en/artificial-intelligence/recommendation-ethics.

⁵¹ Zenonos, Alexandros, PhD. "Artificial Intelligence and Data Protection - Towards Data Science." *Medium*, 27 Jan. 2023, towardsdatascience.com/artificial-intelligence-and-data-protection-62b333180a27.

in isolation but is deeply embedded in social contexts and human interactions. The sociotechnical approach seeks to understand and address the complex interactions between technology, people, organizations, and the broader societal context. In the context of AI, the sociotechnical approach advocates for considering not only the technical aspects of AI development but also the social, ethical, and economic implications of AI systems. It involves involving diverse stakeholders, including end-users and affected communities, in the AI design process to ensure that the technology aligns with their needs, values, and preferences.⁵²

Human-centered AI

Human-centered AI specifically focuses on placing human values, interests, and well-being at the forefront of AI development and deployment. It emphasizes designing AI systems that prioritize human needs and respect human rights. This approach aims to create AI technologies that augment human capabilities, rather than replacing or undermining human decision-making. Human-centered AI also involves incorporating principles such as transparency, fairness, interpretability, and accountability into the AI development process. The goal is to ensure that AI systems are not only technically proficient but also ethical, trustworthy, and aligned with human values.⁵³

Collaboration and Multidisciplinary Approaches

AI and human rights intersect across various fields, including technology, law, ethics, sociology, and policy. Collaboration among experts from different disciplines ensures a comprehensive understanding of the complexities involved, allowing for more effective solutions and policies. One critical aspect of collaboration is to integrate ethics into AI development. By involving ethicists and technologists in the early stages of AI creation, it becomes possible to design ethical frameworks and

⁵² "Accountability in Artificial Intelligence: What It Is and How It Works." *AI & Society*, Springer Nature, Feb. 2023, <https://doi.org/10.1007/s00146-023-01635-y>.

⁵³ "Human Centered AI - Definition, Benefits, Cons, Applications." *clickworker.com*, 23 Nov. 2022, www.clickworker.com/ai-glossary/human-centered-ai.

guidelines that prioritize human rights protection. This proactive approach can help prevent potential violations and biases.⁵⁴

FURTHER READING

Here you can find some articles that can help with your research.

1. This Recommendation addresses ethical issues related to the domain of AI to the extent that they are within the United Nations Educational, Scientific and Cultural Organization (UNESCO's) mandate. <https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence>
2. Recommendation on the Ethics of AI by UNESCO <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
3. Montreal Declaration for a Responsible Development of AI [https://monoskop.org/images/d/d2/Montreal Declaration for a Responsible Development of Artificial Intelligence 2018.pdf](https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_Artificial_Intelligence_2018.pdf)
4. Ethics Guidelines for Trustworthy AI by the European Commission <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>

BIBLIOGRAPHY

1. Frankenfield, Jake. "Artificial Intelligence: What It Is and How It Is Used." *Investopedia*, Apr. 2023, www.investopedia.com/terms/a/artificial-intelligence-ai.asp.
2. OHCHR. "What Are Human Rights?" *OHCHR*, www.ohchr.org/en/what-are-human-rights.
3. Venngage. "Human Rights Template." *Venngage*, venngage.com/templates/infographics/human-rights-template-7d71ff49-ab13-4aae-9325-5dd1fb64accf.
4. *Accountability (OECD AI Principle)* - *OECD.AI*. oecd.ai/en/dashboards/ai-principles/P9.
5. Yasar, Kinza, and Ivy Wigmore. "Black Box AI." *WhatIs.com*, Mar. 2023, www.techtarget.com/whatis/definition/black-box-AI.

⁵⁴ Ebadi, Bushra. "Collaboration Is Necessary for Ethical Artificial Intelligence." *Centre for International Governance Innovation*, 20 Aug. 2018, www.cigionline.org/articles/collaboration-necessary-ethical-artificial-intelligence.

6. *What Is AI Transparency?* www.holistica.ai/blog/ai-transparency#:~:text=A%20major%20motivation%20for%20AI,system%20is%20being%20used%20appropriately.
7. *What Is AI Transparency?* www.holistica.ai/blog/ai-transparency.
8. Elliott, David, and Eldon Soifer. "AI Technologies, Privacy, and Security." *Frontiers in Artificial Intelligence*, vol. 5, Frontiers Media, Apr. 2022, <https://doi.org/10.3389/frai.2022.826737>.
9. Van Rijmenam Csp, Mark. "Privacy in the Age of AI: Risks, Challenges and Solutions." *Dr Mark Van Rijmenam, CSP | Strategic Futurist Speaker*, Apr. 2023, www.thedigitalspeaker.com/privacy-age-ai-risks-challenges-solutions/#:~:text=In%20the%20context%20of%20AI,based%20on%20their%20personal%20data.
10. Emeritus. (2023b). Big Data vs Data Analytics: key differences. *Emeritus - Online Certificate Courses | Diploma Programs*. <https://emeritus.org/in/learn/big-data-vs-data-analytics/#:~:text=Type%20of%20Data%20in%20Big,analytics%20is%20mostly%20structured%20data.>
11. *January 2005 - AI Newsletter*. (2005, January 1). https://www.ainewsletter.com/newsletters/aix_0501.htm#w
12. *What is deep learning? | How it works, techniques & applications*. (n.d.). MATLAB & Simulink. <https://www.mathworks.com/discovery/deep-learning.html#:~:text=Deep%20learning%20is%20a%20machine,a%20pedestrian%20from%20a%20lamppost.>
13. Team, I. D. a. A., & Team, I. D. a. A. (2023). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference? *IBM Blog*. <https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/>
14. *Data breach*. (n.d.). <https://www.trendmicro.com/vinfo/us/security/definition/data-breach>
15. "Human-centered Artificial Intelligence." www.cognizant.com, www.cognizant.com/us/en/glossary/human-centered-ai#:~:text=What%20is%20human%20Dcentered%20AI,experience%20between%20human%20and%20robot.
16. Potenza, Davide. "AI History: The Dartmouth Conference." *Klondike*, June 2023, www.klondike.ai/en/ai-history-the-dartmouth-

conference/#:~:text=1956%20thus%20marks%20the%20official,proposed%20to%20call%20Artificial%20Intelligence.

17. Glover, Ellen. "What Is AI Winter?" *Built In*, Dec. 2022, builtin.com/artificial-intelligence/ai-winter.
18. Jaspreet. "A Concise History of Neural Networks - Towards Data Science." *Medium*, 19 Sept. 2022, towardsdatascience.com/a-concise-history-of-neural-networks-2070655d3fec#:~:text=The%20idea%20of%20neural%20networks,McCulloch%20and%20mathematician%20Walter%20Pitts.
19. "What Is the History of Artificial Intelligence (AI)?" *Tableau*, www.tableau.com/data-insights/ai/history#:~:text=1952%3A%20A%20computer%20scientist%20named,it%20came%20into%20popular%20usage.
20. Hogue, Preston. "Terms of Use: User Privacy and the Algorithms Behind Social Media." *SecurityWeek*, Dec. 2020, www.securityweek.com/terms-use-user-privacy-and-algorithms-behind-social-media.
21. "Social Media Privacy." *EPIC - Electronic Privacy Information Center*, epic.org/issues/consumer-privacy/social-media-privacy.
22. Richards, S., & Richards, S. (2022, July 11). AI in Healthcare: Data Privacy and Ethics Concerns - Lexalytics. *Lexalytics*. <https://www.lexalytics.com/blog/ai-healthcare-data-privacy-ethics-issues/>
23. Cyberinsiders. "What Are the Biggest Privacy Issues Associated with Big Data?" *Cybersecurity Insiders*, 20 Apr. 2020, www.cybersecurity-insiders.com/what-are-the-biggest-privacy-issues-associated-with-big-data/.
24. Legal IT Insider. (2023b). OpenAI and Microsoft face class action lawsuit for allegedly violating copyright and privacy laws. *Legal IT Insider*. <https://legaltechnology.com/2023/06/29/openai-and-microsoft-face-class-action-lawsuit-for-allegedly-violating-copyright-and-privacy-laws/>
25. Ortiz, S. (2023, June 29). OpenAI sued for "stealing" data from the public to train ChatGPT. *ZDNET*. <https://www.zdnet.com/article/openai-sued-for-stealing-data-from-the-public-to-train-chatgpt/>
26. Angwin, Julia, et al. "Machine Bias *." *Auerbach Publications eBooks*, 2022, pp. 254–64. <https://doi.org/10.1201/9781003278290-37>.
27. Hall, Reva Schwartz Apostol Vassilev, Kristen K. Greene, Lori Perine, Andrew Burt, Patrick. "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence | NIST." *NIST*, Mar. 2023,

www.nist.gov/publications/towards-standard-identifying-and-managing-bias-artificial-intelligence.

28. <https://arxiv.org/ftp/arxiv/papers/2304/2304.07683.pdf>
29. <https://arxiv.org/ftp/arxiv/papers/2304/2304.07683.pdf>
30. CCPA. theccpa.org.
31. *Health Insurance Portability and Accountability Act of 1996 (HIPAA)* | CDC. www.cdc.gov/phlp/publications/topic/hipaa.html.
32. ¹ https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/snapshot_v1_countering_the_rise_of_digital_authoritarianism_0.pdf
33. ¹ *EU AI Act: First Regulation on Artificial Intelligence | News | European Parliament*. 6 Aug. 2023, www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?&at_campaign=20226-Digital&at_medium=Google_Ads&at_platform=Search&at_creation=RSA&at_goal=TR_G&at_advertiser=Webcomm&at_audience=ai%20europe&at_topic=Artificial_intelligence_Act&at_location=GR&gclid=CjwKCAjwtuOlBhBREiwA7aCjwKCAjwtuOlBhBREiwA7agf1oxAWtyNRldcKfhv-Kum7PBERAxhURR8Clh5XvPBXClcyQ1GRVYgGBoCeYMQAvD_BwE.
34. Home - Partnership on AI." Partnership on AI, 13 July 2023, partnershiponai.org.
35. Future of Life Institute. "Home - Future of Life Institute." *Future of Life Institute*, 13 Mar. 2023, futureoflife.org.
36. Frank, La Rue. "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue :: Addendum." *United Nations Digital Library System*, 27 May 2011, digitallibrary.un.org/record/706200?ln=en.
37. "ENISA Statement Regarding the Landmark Online Privacy ECJ-ruling on Google Search Engine and the Right to Be Forgotten." *ENISA*, www.enisa.europa.eu/news/enisa-news/enisa-statement-regarding-the-landmark-online-privacy-ecj-ruling-on-google-search-engine-and-the-right-to-be-forgotten.
38. *Shaping Europe's Digital Future*, 8 Apr. 2019, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
39. [https://www.congress.gov/bill/116th-congress/house-bill/2231#:~:text=Introduced%20in%20House%20\(04%2F10%2F2019\)&text=This%20](https://www.congress.gov/bill/116th-congress/house-bill/2231#:~:text=Introduced%20in%20House%20(04%2F10%2F2019)&text=This%20)

bill%20requires%20specified%20commercial,artificial%20intelligence%20or%20mac
hine%20learning

40. AI Now Institute. "AI Now Symposia Archives." *AI Now Institute*,
ainowinstitute.org/series/ai-now-symposia.
41. ¹ "AI for Good Global Summit - IoT for All." *IoT for All*, 1 June 2023,
www.iotforall.com/events/ai-for-good-global-
summit#:~:text=This%20year%2C%20the%20highly%20anticipated,provide%20inval
uable%20insights%20and%20discussions.
42. "Guiding Principles on Business and Human Rights | United Nations
Development Programme." *UNDP*,
www.undp.org/laopdr/publications/guiding-principles-business-and-human-
rights.
43. OHCHR. "Special Rapporteur on Freedom of Opinion and Expression."
OHCHR, www.ohchr.org/en/special-procedures/sr-freedom-of-opinion-and-
expression#:~:text=Purpose%20of%20the%20mandate&text=The%20mandat
e%20of%20the%20Special,human%20rights%20law%20and%20standards.
44. United Nations. "Secretary-General's High-level Panel on Digital Cooperation
| United Nations." *United Nations*, www.un.org/en/sg-digital-cooperation-
panel.
45. *EUR-Lex - 32016R0679 - EN - EUR-Lex*. eur-
lex.europa.eu/eli/reg/2016/679/oj.
46. *CCDCOE*. ccdcoe.org/library/publications/artificial-intelligence-and-
autonomy-in-the-military-an-overview-of-nato-member-states-strategies-
and-deployment.
47. "Ethics Guidelines for Trustworthy AI." *Shaping Europe's Digital Future*, 8 Apr.
2019, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-
ai.
48. "Tackling Bias in Artificial Intelligence (and in Humans)." *McKinsey &
Company*, 6 June 2019, www.mckinsey.com/featured-insights/artificial-
intelligence/tackling-bias-in-artificial-intelligence-and-in-humans.
49. Novelli, Claudio, et al. "Accountability in Artificial Intelligence: What It Is and
How It Works." *AI & Society*, Springer Nature, Feb. 2023,
https://doi.org/10.1007/s00146-023-01635-y.
50. "Ethics of Artificial Intelligence." *UNESCO*, June 2023,
www.unesco.org/en/artificial-intelligence/recommendation-ethics.

51. Zenonos, Alexandros, PhD. "Artificial Intelligence and Data Protection - Towards Data Science." *Medium*, 27 Jan. 2023, towardsdatascience.com/artificial-intelligence-and-data-protection-62b333180a27.
52. "Accountability in Artificial Intelligence: What It Is and How It Works." *AI & Society*, Springer Nature, Feb. 2023, <https://doi.org/10.1007/s00146-023-01635-y>.
53. "Human Centered AI - Definition, Benefits, Cons, Applications." *clickworker.com*, 23 Nov. 2022, www.clickworker.com/ai-glossary/human-centered-ai.
54. Ebadi, Bushra. "Collaboration Is Necessary for Ethical Artificial Intelligence." *Centre for International Governance Innovation*, 20 Aug. 2018, www.cigionline.org/articles/collaboration-necessary-ethical-artificial-intelligence.