Artificial Intelligence in Development and **Humanitarian Work**

Promises, Paradoxes, and Perils

A Project of

impactskills

UNIVERSITÀ DI TORINO

Dipartimento di Culture, Politica e Società

MANCHESTER 1824
The University of Manchester

With the Support of

Ministero degli Affari Esteri e della Cooperazione Internazionale

*Publication produced under the project: **Potential and Risks of Artificial Intelligence Predictive Models in the Development and Humanitarian Sector** funded by Italian Ministry of Foreign Affairs and International Cooperation.*

The views expressed in this volume are to be attributed to the authors and do not necessarily reflect the views of the Ministry of Foreign Affairs and International Cooperation, nor partners and sponsors.

SCIENTIFIC SUPERVISORS

Egidio Dansero, full professor of Economic and Political Geography (M-GGR/02)
Vice Rector Vicar of Sustainability and Development Cooperation, Scientific Director of the Master ICT for Development and Social Good, Dip. CPS, University of Turin.
Gianluca Iazzolino, professor of Digital for Development, University of Manchester

RESEARCH PRINCIPAL

Ron Salaj, Head of Research and Strategy, Impactskills

SCIENTIFIC COMMITTEE

Alberto Vanolo, professor of Economic-Political Geography (M-GGR/02), Dep. CPS, University of Turin
Chiara Certomà, professor of Political-economic Geography, Dep. Memotef, University Sapienza Rome
Elisa Bignante, professor of Economic and Political Geography (M-GGR/02), Dep. CPS, University of Turin
Guido Boella, full professor of Computer Science, Vice-rector Vicar of the University of Turin for the promotion of relations with companies and trade associations of enterprises
Marinella Belluati, professor of Cultural and Communication Processes (SPS/08), Dep. CPS, University of Turin
Monica Cerutti, Researcher, Dip. of Informatics, University of Turin.
Paola Minoia, professor of Economic-Political Geography (M-GGR/02), Dep. CPS, University of Turin
Silvia Pochettino, founder and CEO, ImpactSkills

GRAPHIC DESIGNER

Diletta Vignolo

Acknowledgments

We would like to express our deepest gratitude to Elisa Bignante and Paola Minoia for their invaluable review and constructive feedback on this report. Their insightful suggestions and thoughtful guidance have greatly enhanced the quality and clarity of our work.

*We extend our heartfelt thanks to the panelists who participated in the public workshop, **"Exploring AI's Contradictory Role in International Development and Humanitarianism"**, held on July 3rd, 2024, as part of the project **"Risks and Opportunities of the AI Predictive Models in the International Development and Humanitarian Field"**. The workshop benefited immensely from the contributions of **Giulio Coppi (AccessNow), Jen Persson (Defend Digital Me), Guido Boella (University of Turin)**, as well as the organizing team comprising **Gianluca Iazzolino (University of Manchester), Egidio Dansero (University of Turin), Silvia Pochettino (Impactskills), Luca Indemini (Impactskills), and Ron Salaj (Impactskills)**.*
*We are also deeply grateful to **Filippo Lonardo, the representative from the Ministry of Foreign Affairs of Italy,** for his valuable input. Their engagement fostered a vibrant and constructive discussion, enriching the exploration of AI's role in the international development and humanitarian field.*

Finally, we would like to thank the entire research team and collaborators for their collective efforts and support throughout this project. Although not directly involved in specific contributions, their presence and encouragement have been integral to the successful completion of this report.

Contents and Index

Introduction

The times we live in increasingly reflect what Ziauddin Sardar terms 'Post-Normal Times'[1]: an era marked by chaos, complexity, and contradictions, where traditional orthodoxies are dying, and new ones have yet to emerge. This period is shaped by a polycrisis—a series of interconnected global crises, including wars, climate emergencies, economic inequalities, distrust in democracy, and disruptive influence of technologies such as Artificial Intelligence (AI).

Like other technologies, AI is not neutral; but it embodies social relations, values, biases, politics, and power dynamics. AI technologies do not emerge in a vacuum; their development and deployment are deeply embedded in and influenced by historical, cultural, and political-ideological contexts and beliefs. Langdon Winner's seminal question, "Do artifacts have politics?"[2] finds renewed relevance in our (post-normal) times, prompting us to ask: "Does AI have politics?"

Langdon Winner's assertion is that "artefacts have politics". Winner argues that technical arrangements can perpetuate or challenge social orders, often producing results that serve some interests while disadvantaging others. Winner illustrates this with a compelling example from urban planning: Robert Moses, in redesigning New York parkways, ensured that bridges providing access to beaches and recreational parks were constructed so low that buses could not pass underneath them. At the time, this effectively excluded Black and lower-income populations, who relied on public transport, from these recreational spaces. Without any explicit apartheid laws or overt impropriety, Moses implemented a covert but impactful form of segregation, maintaining his beaches as "for whites only." Winner concludes that not only do artefacts have politics, but their biases are particularly insidious because they hide under the guise of objectivity, efficiency, or practicality. Similar dynamics have been empirically evidenced in numerous critical AI studies, explored further in Chapter 4 of this research.

In development and humanitarian efforts, where AI is frequently hailed as a solution to complex problems, such critiques take on heightened relevance. The existing polycrisis challenges the assumption that AI can provide universal, context-independent solutions. Instead, its deployment must be evaluated through the lens of historical inequities, power imbalances, and cultural diversity. Recent United Nations (UN) Sustainable Development Goals (SDG) reports reveal a paradox: despite rapid technological advancements, particularly in the field of AI, progress toward the SDGs has regressed in the past few years. This contradiction underscores the necessity for key actors in development and humanitarian sectors to recognize the politics of AI—how its ownership, design, deployment, and governance reflect underlying power dynamics, biases, and values. By acknowledging these embedded politics, stakeholders can critically assess AI systems to ensure they are implemented in ways that promote equity, sustainability, and inclusion, thereby aligning with the universal principles of the SDG Agenda: *leaving no one behind, fostering inclusivity, addressing inequalities, promoting sustainable development, and ensuring accountability and transparency*. Recognizing the politics of AI allows for proactive measures to

[1] Sardar, Z. (2009). *Welcome to Postnormal Times*. *Futures*, 42(5), pp. 435–444. [doi] Available at: https://doi.org/10.1016/j.futures.2009.11.028 [Accessed on 18 November 2024]
[2] Winner, L. (1977). *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA: MIT Press.

mitigate its potential to reinforce systemic inequalities, ensuring it serves as a tool for advancing these core SDG principles rather than undermining them.

While AI presents significant risks and paradoxes, its promises should not be completely dismissed, particularly when used as a complementary tool rather than a standalone decision-making system. Promising examples highlight how AI can support meaningful outcomes when contextualised and guided by human and community-centred approaches. For instance, initiatives like *Climate TRACE* leverage machine learning (ML) to monitor greenhouse gas emissions with unprecedented accuracy, empowering policymakers and communities with actionable data. Similarly, AI has been used to address spatial apartheid in South Africa, helping to identify systemic inequalities in resource distribution. These examples, explored in greater depth in the Chapter 3 of this research, demonstrate that when AI is designed with ethical foresight, transparency, and deployed alongside other tools, it can strengthen efforts toward sustainability and equity.

Yet, as Langdon Winner warns in his *Autonomous Technology*, technological systems often operate beyond human oversight, becoming autonomous entities with their own inertia. Winner observes that "members of technological society actually know less and less about the fundamental structures and processes sustaining them." (Winner, 1977: 295). He cautions that this growing gap between societal realities and human understanding risks neutralising the agency once central to citizenship. Overwhelmed by information and disconnected from decision-making, people lose the ability to participate actively in shaping technological futures. For policymakers, the challenge of the future lies in restoring citizens' *agency* with-and-in technological systems—shifting the inertia from autonomous technologies into the autonomy of citizens. This transformation would empower citizens to actively shape the systems that influence their lives, thereby addressing the uncertainties, challenges and contradictions of existing Post-Normal Times.

Why this research?

This research aims to provide a nuanced examination of the promises, paradoxes and perils associated with AI in development and humanitarian sectors, with a focus on its predictive and generative[3] capabilities. It investigates the various use cases of AI in these contexts, explores whether AI's promise to serve as a force for equitable and sustainable development is supported by independent evidence, and offers an overview of the main critiques of AI. The core questions guiding this report include:

- How are predictive and generative AI being integrated into development and humanitarian practices?
- What are the social, ethical, and operational challenges arising from their use?
- To what extent do these technologies contribute to—or detract from—the achievement of SDGs?

[3] Generative AI fundamentally operates as a predictive system. As Holly Lewis explains, whether it generates text, images, or motion, the process is based on predicting the next token, pixel, or frame, highlighting that in machine learning, generation is a form of prediction. For more see: Lewis, H. (2023). *Toward AI Realism*. [online] Available at: https://spectrejournal.com/toward-ai-realism/ [Accessed 18 November 2024]

- How can governance frameworks and policy interventions mitigate risks while engaging in AI realism?

Methodology

This research report draws upon a comprehensive review of academic literature, international and intergovernmental organisation reports, and case studies from development and humanitarian contexts. It incorporates evidence from peer-reviewed journals, industry publications, and policy documents to provide a robust understanding of the current state and implications of AI-driven predictive and generative technologies. Furthermore, a preliminary taxonomy categorises AI applications in humanitarian contexts into five overarching categories—predictive, generative, assistive, optimisation, and facial recognition technologies— offering an analytical and structured framework for understanding the functionality and use cases of AI in the sector. While a comparable classification could not be identified in the existing literature, we hope this (preliminary) taxonomy serves as a foundation for further research and supports the informed deployment of AI in humanitarian efforts.

The report presents four case studies to explore AI applications aligned with SDGs 4 (Education), 10 (Reduced Inequalities), and 13 (Climate Action). The case studies were identified and analysed using a combination of extensive academic literature reviews, general web searches, and snowball searching techniques. The snowball method involved drawing insights from citations and references within relevant reports and articles to ensure a comprehensive understanding of each application. The following criteria guided the selection of case studies:

1. *Alignment with SDGs:* Each case study was chosen for its direct relevance to three predefined SDGs, with a specific focus on education (SDG 4), inequality (SDG 10), and climate action (SDG 13).
2. *Demonstrated Impact:* Case studies were selected for their documented impact—whether positive or negative—to provide a balanced understanding of AI's potential and limitations.
3. *Dynamic Applications of AI:* Cases were chosen to showcase diverse uses of AI technologies in addressing critical social and environmental issues. These include predictive and generative AI models employed in innovative and contextually relevant ways.
4. *Diverse Contexts and Implementing Entities:* To ensure representativeness, the selected case studies span a range of geographic, thematic, and institutional contexts. This includes diversity in the types of entities implementing the initiatives, from private commercial organisations to research centres, intergovernmental organisations, and coalitions of multiple stakeholders.
5. *Replicability, Scalability, and Evidence:* Case studies were assessed based on their potential for replication or scalability in similar contexts. Additionally, the existence of robust documentation, publications, and reviews was a key factor in evaluating the credibility and depth of evidence surrounding each case.

Finally, the report advocates for the democratisation of science and policymaking in the era of AI. It provides a set of six recommendations aimed at fostering more inclusive and democratised forms of AI, aligned with the principles of the SDG Agenda and the strategic framework of the Italy's National

Sustainable Development Strategy[4], which is based on the 5Ps: People, Planet, Prosperity, Peace, and Partnership. The report concludes with the *AI Hype-Reality Gap Model*, a practical and critical assessment tool designed to help stakeholders, particularly policymakers, evaluate AI technologies by identifying overhyped claims, assessing realistic expectations, making informed decisions, and promoting accountability and transparency to ensure a responsible, democratic, and equitable deployment process.

Outline of Chapters

1. **Chapter 1**: Provides an introduction to AI, tracing its historical evolution, defining its key techniques, differentiating between predictive and generative AI, and providing a critical examination of AI technical capabilities.
2. **Chapter 2**: It positions AI within the broader field of international development. It critically examines the "AI for Good" discourse and evaluates AI's role in the Sustainable Development Agenda, with a specific focus on initiatives by the AI4SDG think tank. It further explores AI's use in anticipatory humanitarian action through predictive analytics and discusses its ethical implications in these settings. Finally, it reviews principal AI governance frameworks developed by governmental, global, and multilateral institutions.
3. **Chapter 3**: This chapter delves into AI's applications in advancing three critical SDGs: SDG 4 (Quality Education), SDG 10 (Reduced Inequality), and SDG 13 (Climate Action). Through an analysis of debates, trends, and case studies, the chapter provides insights into how AI is being employed to address these goals. It also identifies critical ethical concerns and challenges in leveraging AI for sustainable development.
4. **Chapter 4**: This chapter critically examines the societal impacts of AI across five key areas: corporate influence on digital development; bias, discrimination, and opacity in AI systems; surveillance and privacy concerns; environmental sustainability challenges; and digital colonialism.
5. **Chapter 5**: The final chapter provides actionable recommendations to guide policymakers and stakeholders, including institutions such as the Ministry of Foreign Affairs of Italy (MAECI) and the Italian Agency for Development and Cooperation (AICS).

[4] Ministero dell'Ambiente e della Sicurezza Energetica, (2022). *Strategia Nazionale per lo Sviluppo Sostenibile 2023: Strategia e allegati.* [pdf] Available at: https://www.mase.gov.it/sites/default/files/archivio/allegati/sviluppo_sostenibile/ALL1_SNSvS_2023_Strategia_e_allegati.pdf [Accessed on 18 November 2024]

CHAPTER 1
Artificial Intelligence: An Introduction

This chapter attempts to explore the journey of Artificial Intelligence (AI) from the Industrial Age to its current state, charting its evolution through a landscape filled with both groundbreaking achievements and profound controversies. We explore key milestones, define AI, and demystify its core techniques, providing a nuanced overview and understanding for both novices and those seeking to enrich their grasp of this contentious domain.

SECTION 1.1
From the March of Intellect to the ImageNet and AlexNet: A Brief History

The history of AI is a tapestry woven with innovation, scepticism, controversies, ambiguities, and pivotal moments that have set the stage for the contemporary landscape of intelligent systems. An 1829 caricature from the "March of Intellect" series by William Heath is a satirical depiction of the technological and intellectual aspirations of nineteenth-century England. It humorously illustrates various futuristic concepts, such as rapid transportation systems, extensive tunnel networks, and elaborate flying machines, all signifying the period's quest for technological acceleration, progress and its concerns about the repercussions of industrialization and innovation.

Initiated as a campaign in England during the Industrial Revolution, the March of Intellect, or "March of Mind", demanded the amelioration of society's ills through programmes of public education for the lower classes (Herzog: 2000, as cited in Pasquinelli: 2023). The March of Intellect was part of the so called "Machinery Question", established by David Ricardo whose thesis was the following: 'while it was true that new machinery would cheapen commodity prices, nonetheless the working class would not benefit from this, since wages would be reduced by the competition among workers which is caused by technological unemployment' (Pasquinelli 2023). The Machinery Question, as Pasquinelli (2023) states, was therefore a complex phenomenon intertwining popular culture, political propaganda, scientific contestation, and social control through education.

During the Industrial Age, the polymath Charles Babbage designed the first computing machine, celebrated as the precursor of modern computers, known as the "Difference Engine". Babbage's Difference Engine aimed to 'automate the calculation of logarithms and sell error-free logarithmic tables, which were crucial in astronomy and for maintaining British hegemony in maritime trade and [its] aggressive colonial expansion' (Pasquinelli 2023: 52-55).

However, it was Ada Lovelace, celebrated as the 'first woman programmer of history', who assisted Babbage in designing the Analytical Engine, considered the first steam-powered general-purpose computer. Even though the Analytical Engine was never realised, Lovelace's virtual program, understood as a set of instructions to be executed by machine, is considered the first example of present-day algorithms (Pasquinelli 2023: 69). Ada Lovelace had envisioned the machine's capacity to transcend mere numerical calculations, something that Babbage failed to see, proposing the concept of general computing and establishing herself as an early pioneer in technology. Babbage-Lovelace collaboration is marked also by relationships of power. According to Pasquinelli (2023), Babbage asked Lovelace to publish her notes on Analytical Engine anonymously, something that she resisted, thus becoming an exemplary figure of technical curiosity and emancipation in an academic and scientific world dominated by men.

Like any other revolution, the Industrial Revolution also triggered social discontent and unrest. In the beginning of the Industrial Revolution, a community of textile workers feared that the newly introduced cost-saving machines were taking their jobs. These machines were now doing the work, cheaply and less artfully, that had for generations formed the foundation of their lives and their communities (Merchant 2023). The unrest initiated by them came to be known as the Luddites movement (1811-1816). What Luddites were resisting was not simply the automatization of their work but their own reduction and automatization (McQuillan 2023: 131). The uprising of two centuries past bore the seeds of a conflict that remains central to our contemporary relationship with technology. In many respects, the resolution of that historical struggle is still instrumental in shaping the future we are moving towards.

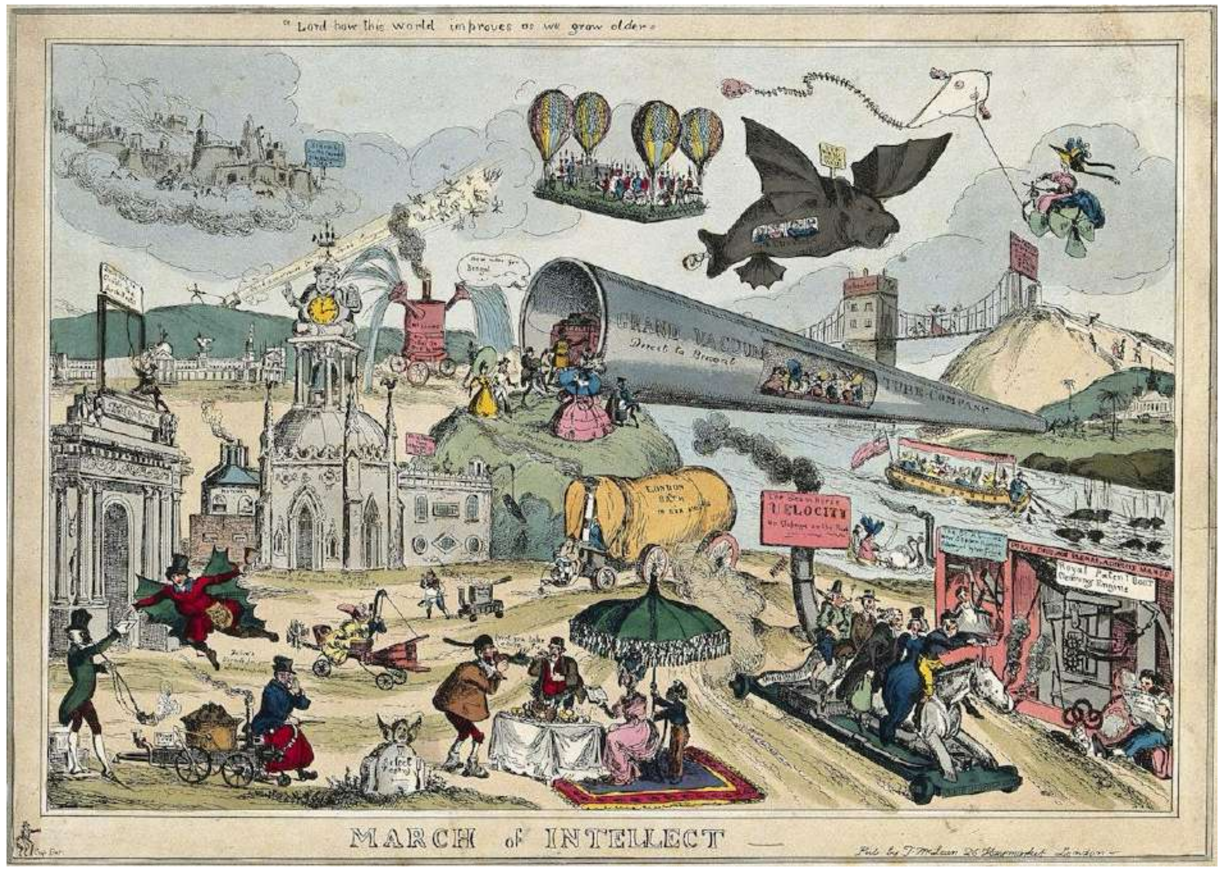

Figure 1.1. William Heath, *The March of Intellect*, 1829

As we move on, the period during and post World War II, becomes pivotal for computational technologies. For instance, IBM's role in providing to the Third Reich tabulating technology necessary for operating the census proved instrumental, with Dehomag punching up to one million cards daily. The introduction of

the "Dehomag D-11" tabulator, an advanced piece of technology, allowed for the efficient processing of extensive data, including individuals' ancestry, religious beliefs, and assets. This facilitated the classification and persecution of Jews, as the data contributed to their eventual concentration, deportation, and extermination[5].

The earliest contributions to AI were made in the mid-20th century by Alan Turing, a British logician, mathematician and pioneering figure in computing. During World War II, Turing was a leading cryptanalyst at the Government Code and Cypher School in Bletchley Park. He cracked the Enigma code, a type of enciphering machine used by the German armed forces to send messages securely. In London in 1947, Turing delivered what is believed to be the earliest public lecture to mention computer intelligence, stating, "What we want is a machine that can learn from experience," and adding that the "possibility of letting the machine alter its own instructions provides the mechanism for this."[6]. Turing's work, however, was not without controversy. For instance, Turing himself reiterated a hierarchical and authoritarian mode of thinking, during a 1947 lecture. As Pasquinelli (2023) highlights, Turing envisioned the Automatic Computing Engine (ACE), one of the first digital computers, as a centralised apparatus that orchestrated its operations as a hierarchy of master and servant roles:

> Roughly speaking those who work in connection with the ACE will be divided into its masters and its servants. Its masters will plan out instruction tables for it, thinking up deeper and deeper ways of using it. Its servants will feed it with cards as it calls for them. They will put right any parts that go wrong. They will assemble data that it requires. [...] As time goes on the calculator itself will take over the functions both of masters and servants. [...] One might for instance provide curve followers to enable data to be taken direct from curves instead of having girls read off values and punch them on cards[7].

Turing's projection of a centralised digital apparatus operating on a master-servant hierarchical division of labour, as highlighted by Pasquinelli (2023), can be seen as prescient of the modern dynamics between humans and technology. However, Turing's vision about 'calculators replacing masters and servants' is contravened by today's AI systems. Today's AI technologies rely on the labour of millions of invisible 'servants' from the Global South, as documented by Mary Gray and Siddharth Suri[8]. These 'ghost workers' perform tasks such as data labelling, content moderation, and verification, among other forms of digital piecework. These tasks are essential for training, maintaining, and keeping AI systems operational.

[5] For more see: Black E., 2001. *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation*. NYC: Three Rivers Press; William S., 1998. *Population Statistics, the Holocaust, and the Nuremberg Trials,* Population and Development Review; Emerson P, 1995. *Building IBM: Shaping an Industry and its Technology*. Massachusetts: MIT Press; Friedrich K., *The Dehomag D11 Tabulator - A Milestone in the History of Data Processing.*

[6] Copeland J., 2000. Alan Turing and the origins of AI. [online] Available at: https://www.alanturing.net/turing_archive/pages/reference%20articles/what_is_ai/What%20is%20AI03.html [Accessed on 20 March 2024]

[7] Turing A., 1947. *Lecture on the Automatic Computing Engine*, in *The Essential Turing,* ed. B. Jack Copeland, London: Clarendon Press*,* 2004.

[8] M. Gray, S. Suri, 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston, New York: Houghton Mifflin Harcourt

Perhaps the first academic project to focus expressly on "Artificial Intelligence" was a research proposal[9] written by John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (IBM), and Claude Shannon (Bell Telephone Laboratories) in 1955. In the proposal, the authors state the following:

> We propose [...] a study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Looking back, the research proposal was both pertinent and visionary, acknowledging contemporary computing constraints while foreseeing an escalation in computing power that would mitigate these limitations, the development of neural networks capable of recognizing and predicting data patterns to tackle complex issues, and the potential for machines to self-improve. The Dartmouth Conference in 1956 is often considered the official birth of AI as a field. This conference set the stage for the development of AI, marking the transition from theoretical discussions to practical research and development.

Finally, in the late 1950s and 1960s, we can see two currents of AI emerging. According to Stuart and Hubert Dreyfus 1988 essay "Making a Mind versus Modelling the Brain", the AI field was outlined into two lineages: symbolic and connectionist AI. The symbolic AI was associated with the Dartmouth Conference and is characterised by its use of formal logic systems, rule-based systems for knowledge representation. In other words, AI relied on inserting human knowledge and behavioural rules into computer codes[10]—rule-by-rule, line-by-line[11]. Connectionist AI, on the other hand, is the lineage of artificial neural networks pioneered by Frank Rosenblatt's invention of the 'perceptron' in 1957, which unfolded into convolutional neural networks in the late 1980s and, eventually, launched the deep learning architecture that has prevailed since the 2010s (Pasquinelli 2023). It should be noted that during the 1950s, AI was part of a broader discourse and practice surrounding *self-organisation theories* and triggered particular interest of the US military. Situated in the Cold War era, "the US military expressed interest in the logic of self-organisation as an alternative, more efficient, means of computation" (Pasquinelli 2023: 146). For instance, in May 1959, the conference on 'Self-Organising Systems' was organised by the Information System Branch of the US Office of Naval Research in collaboration with Illinois Institute for Technology.

[9] McCarthy J., et. al, 1955. *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* [online] Available at: https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html [Accessed on 20 March 2024].
[10] Flynn Sh., 2020. *The Difference Between Symbolic AI and Connectionist AI*. [online] Available at: https://blog.re-work.co/the-difference-between-symbolic-ai-and-connectionist-ai/ [Accessed on 21 March 2024]
[11] Friel K., 2023. *A Look Back on the Dartmouth Summer Research Project on Artificial Intelligence.* [online] Available at: https://www.thedartmouth.com/article/2023/05/a-look-back-on-the-dartmouth-summer-research-project-on-artificial-intelligence [Accessed on 21 March 2024].

This event brought together cyberneticians from both the connectionist and symbolic AI camps (Pasquinelli 2023: 146).

However, the field of AI underwent a transformative shift with the introduction of the paradigmatic deep learning dataset known as ImageNet. AI researcher Fei-Fei Li initiated work on ImageNet in 2006-2007, also drawing inspiration from the contributions of Princeton professor Christiane Fellbaum, who was among the creators of WordNet. But how can you build a large dataset, annotated and labelled, which includes concrete objects like strawberries and cars, and abstract concepts like love? As explained in the interview with Quartz[12], Li's first idea was to hire undergraduate students for $10 an hour to manually find images and add them to the dataset, but:

> [...] back-of-the-napkin maths quickly made Li realise that at the undergrads' rate of collecting images it would take 90 years to complete. Undergrads were time-consuming, algorithms were flawed, and the team didn't have money—Li said the project failed to win any of the federal grants she applied for, receiving comments on proposals that it was shameful Princeton would research this topic, and that the only strength of the proposal was that Li was a woman.

A solution was finally found. ImageNet, which consists of more than 14 million labelled images, each of which is tagged, belonging to more than 20,000 categories, was made possible by the efforts of thousands of invisible workers who were recruited through Amazon's Mechanical Turk platform. Crowdworkers who made ImageNet possible received payment for each task they finished, which sometimes was as little as a few cents[13]. In addition, as McQuillan (2023) states, nothing of contribution of these workers is acknowledged or granted any agency; rather they are characterised, where they are mentioned at all, as interchangeable sets of eyeballs[14].

As part of their efforts 'to measure the progress of computer vision for large scale image'[15]— ImageNet project has, since 2010, hosted an annual software contest, the ImageNet Large Scale Visual Recognition Challenge, where software programs compete to correctly classify and detect objects and scenes. The 30th September 2012 competition, a Convolutional Neural Network (CNN) called AlexNet won the ImageNet 2012 challenge, 'outperforming all the prior competitors and won the challenge by reducing the top-5 error from 26% to 15.3%. The second place top-5 error rate, which was not a CNN variation, was around 26.2%'[16]. AlexNet was developed by the SuperVision group, comprised of Alex Krizhevsky,

[12] Gershgorn D., 2017. The data that transformed AI research—and possibly the world. [ionline] Available at: https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world [Accessed on 21 March 2024]
[13] Newman A., 2019. *I Found Work on an Amazon Website. I Made 97 Cents an Hour.* [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed on 21 March 2024].
[14] For a critical view of the Imagenet, please see also: Denton E., Hanna A., Amironesei R., 2021. *On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet, Big data & Society,* 8(2). [doi] Available at: https://doi.org/10.1177/20539517211035955 [Accessed on 21 March 2024]
[15] Imagenet. [online] Available at: https://image-net.org/challenges/LSVRC/ [Accessed on 21 March 2024]
[16] Das Sh., 2017. *CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more…* [online] Available at: https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 [Accessed on 21 March 2024]

Geoffrey Hinton, and Ilya Sutskever[17]. Its remarkable performance in ImageNet's annual competition not only marked a turning point for computer vision but also triggered the widespread adoption and advancement of deep learning techniques across various domains. This achievement set a new standard for AI research and practice, as well as catalysed a wave of innovation that continues to shape the technological landscape today. Next, we will explore some of the primary concepts of AI and attempt to demystify some of its core operations.

SECTION 1.2

Defining AI

AI can be broadly defined as the branch of computer science dedicated to creating systems capable of performing tasks that typically require human intelligence. This encompasses a wide range of capabilities, from simple algorithms that automate repetitive tasks to complex systems that can learn, adapt, and make or recommend decisions. There are also great amounts of definitions about AI. For instance, The Alan Turing Institute defines AI as:

> The design and study of machines that can perform tasks that would previously have required human (or other biological) brainpower to accomplish. AI is a broad field that incorporates many different aspects of intelligence, such as reasoning, making decisions, learning from mistakes, communicating, solving problems, and moving around the physical world[18].

A 2019 recommendation of the Organisation for Economic Co-operation and Development (OECD) defined AI as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments'[19]. However, this definition was updated in November 2013, which has been incorporated in EU's AI rulebook known as "EU AI Act"[20]. The guidelines of the OECD define AI systems as following:

> [...] a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments[21].

UNICEF's definition, part of their policy guidance, on AI is rather more nuanced, stating that:

[17] For more information see: Krizhevsky A., Sutskever I., Hinton G, 2012. *ImageNet Classification with Deep Convolutional Neural Networks*. [pdf] Available at: https://proceedings.neurips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf [Accessed on 21 March 2024].
[18] The Alan Turing Institute. *Data Science and AI Glossary*. [online] Available at: https://www.turing.ac.uk/news/data-science-and-ai-glossary [Accessed on 15 March 2024]
[19] OECD, 2019. *Recommendation of the Council on Artificial Intelligence*. [online] Available at: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 [Accessed 15 March 2024]
[20] It should be mentioned that also the Council of Europe's Committee on Artificial Intelligence (CAI) adopted the same definition in their "Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law", as per 18 December 2023. Available at: https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043 [Accessed on 15 March 2024]
[21] Russell S., Perset K., Grobelnik M., 2023. *Updates to the OECD's definition of an AI system explained.* [online] Available at: https://oecd.ai/en/wonk/ai-system-definition-update [Accessed on 15 March 2024]

> AI refers to machine-based systems that can, given a set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment, either directly or indirectly. Often, they appear to operate autonomously, and can adapt their behaviour by learning about the context[22].

According to Holmes and Porayska-Pomsta (2022), UNICEF's definition does not depend on data, although it does accommodate data-driven AI techniques such as artificial neural networks and deep learning; it also includes rule-based or symbolic AI and any new paradigm of AI that might emerge in future years; and, lastly, it highlights that AI systems necessarily depend on human objectives and sometimes "appear to operate autonomously", rather than assuming that they do operate autonomously, which is key given the critical role of humans at all stages of the AI development pipeline.

In the realm of AI, distinctions are often made to understand the various capabilities and functions of AI systems. Predominantly, these can be classified into two broad categories:

- Narrow AI, known also as "weak AI", is the most common today. Narrow AI systems are specialised to perform a singular task, such as driving a car, playing chess, recognizing speech or text, predicting weather, spam filtering, virtual assistants, recommendations systems, etc.. Through continuous learning from their environment, they gradually master their designated tasks in real-time, yet they lack the capability to perform tasks beyond the specific single-task environment for which they are designed.
- General AI, known also as a "strong AI" or Artificial General Intelligence (AGI) which refers to an intelligent system with comprehensive or complete knowledge and cognitive computing capabilities, capable of rivalling human thinking. Why General AIl is seen as the next milestone in the AI field, some recognise it as 'the transhuman intelligence of fictional AI'[23]. General AI represents a future aspiration of the field, aiming to create machines that can generalise learning and apply it broadly, a stark contrast to the specialised and focused capabilities of today's Narrow AI systems.

However, Narayanan's framework[24] offers more nuanced and insightful, though not exhaustive, approach to categorising AI technologies into three distinct types, each reflecting a different aspect of AI's impact and development distinguished types:

1. AI technologies that represent "genuine, rapid technological progress" which includes "content identification (e.g. Shazam reverse image search, etc.), face recognition, medical diagnosis from scans, speech to text, deep-fakes, etc.."

[22] UNICEF, 2021. *Policy guidance on AI for children.* [online] Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Accessed on 15 March 2024] https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf [Accessed on 15 March 2024]
[23] McQuillan D., 2019. *Non-Fascist AI.* [online] Available at: https://osf.io/preprints/socarxiv/b64sw/download [Accessed on 17 March 2024]
[24] Narayanan A., 2019. *How to recognize AI snake oil.* [pdf] Available at: https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf [Accessed on 17 march 2024]

2. AI technologies that are "far from perfect, but improving" encompass spam detection, detection of copyrighted materials, hate speech detection, etc..
3. AI technologies that are "fundamentally dubious", involve predicting criminal recidivism, predicting job performance, predictive police, predicting at-risk kids, etc..

The first category of AI focuses on 'perception'; the second category emphasises 'automating judgement'; whereas the third category aims to 'predict social outcomes' as its primary operation axis. But where does generative AI stand into this categorisation? Given its versatile nature, generative AI, with its capacity to create new, unseen content, can touch upon all of the three categories. It belongs to the first category due to its innovative contributions to technology's perception capabilities, while also touching upon the concerns and potentials of the other two categories, illustrating the many-sided impact of generative AI across different domains.

SECTION 1.3

AI Techniques and Technologies

At the heart of AI lie several core techniques that empower machines to learn from data, make decisions, and perform tasks that would typically require human intelligence, such as speech and image recognition or language translation. While these techniques are great in number and too specialised to explore in depth here, we will provide a brief overview of some of the core AI techniques, as well as some AI technologies. Given the extensive array of techniques, each highly specialised, a comprehensive exploration within this document is not feasible. Nevertheless, this paper will offer a concise overview of some fundamental AI techniques and technologies. The categorization of AI techniques and technologies presented herein follows the framework established by Miao et al. (2021).

1.3.1 | Machine Learning

Machine learning (ML) is a subset of AI that gives computers the ability to learn and improve from experience without being explicitly programmed for every task. Machine-learning-based computational approaches made possible many advances in the field of AI, such as: face recognition, autonomous vehicles, natural language processing, and more. For example, by analysing thousands of emails, a machine learning system can learn to distinguish between spam and non-spam messages. Essentially, it's about enabling machines to gain insights and make predictions based on their analysis of data, thereby performing tasks that would typically require human intelligence. There are three main ML approaches (Miao et al 2021):

1. *Supervised learning* involves data that has already been labelled – such as many thousands of photographs of people that have been labelled by humans. The supervised learning links the data to the labels, to build a model that can be applied to similar data – for example, to automatically identify people in new photographs.
2. In *unsupervised learning*, the AI is provided with even larger amounts of data, but this time the data has not been categorised or labelled. The unsupervised learning aims to uncover hidden patterns in the data, clusters that can be used to classify new data. For example, it may automatically identify letters and numbers in handwriting by looking for patterns in thousands of examples.
3. *Reinforcement learning* involves continuously improving the model based on feedback – in other words, this is machine learning in the sense that the learning is ongoing. The AI is provided with some initial data from which it derives a model, which is assessed as correct or incorrect and rewarded or punished accordingly. The AI uses this reinforcement to update its model and then it tries again, thus developing iteratively (learning and evolving) over time. For example, if an autonomous car avoids a collision, the model that enabled it to do so is rewarded (reinforced), enhancing its ability to avoid collisions in the future[25].

1.3.2 | Artificial Neural Networks

Artificial neural networks (ANNs) are a type of machine learning inspired by the way the human brain works. Think of them as a web of interconnected nodes or "neurons" that work together to solve problems, much like how our brain's neurons fire signals to each other to process information. Each node in this network processes small pieces of information, and these nodes are layered to form the network. By passing data through these layers, ANNs can learn to recognize patterns, make predictions, and even make decisions based on the input they receive. In more technical terms ANNs:

> [...] each comprise three types of interconnected layers of artificial neurons: an input layer, one or more hidden intermediary computational layers, and an output layer that delivers the result. During the ML process, weightings given to the connections between the neurons are adjusted in a process of reinforcement learning and 'back propagation', which allows the ANN to compute outputs for new data[26].

A prominent example of Artificial Neural Networks (ANNs) is Google's AlphaGo. In 2016, this advanced AI system made headlines when it triumphed over the world's top Go player, demonstrating the potential of ANNs in mastering complex tasks. Other examples which employ ANNs are facial recognition technology used in smartphones or security systems, identifying individuals' faces with great accuracy.

1.3.3 | Deep Learning

Deep Learning is a sophisticated subset of ML that mimics the workings of the human brain in processing data and creating patterns for use in decision making. It's called "deep" because it makes use of deep ANNs—layers upon layers of interconnected nodes, or artificial neurons, much like the neural networks in our brains. These layers can learn to recognize complex features in data, from the simplest elements in the early layers to highly complex patterns in the deeper layers. For instance, in image recognition, the initial layers might identify edges and colours, while deeper layers recognize more complex elements like shapes and finally specific objects like cars or faces. Some of the emerging models in deep learning include:

- *Deep neural networks* (DNN), which find effective mathematical operations to turn an input into the required output;

[25] Miao F., Holmes W., Huang R., Zhang H., 2021. *AI and education: Guidance for policy-makers*. [online] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000376709 [Accessed on 18 March 2024]
[26] Ibid.

- *Recurrent neural networks* (RNN), which allow data to flow in any direction, can process sequences of inputs, and are used for applications such as language modelling;
- *Convolutional neural networks* (CNN), which process data that come in the form of multiple arrays, such as using three two-dimensional images to enable three-dimensional computer vision[27].

Another approach to deep learning is the Generative Adversarial Networks (GANs). GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated)[28]. The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples[29]. This technique is used widely in image manipulation. For instance, a GAN trained on photographs has generated images of people who look real but do not exist[30].

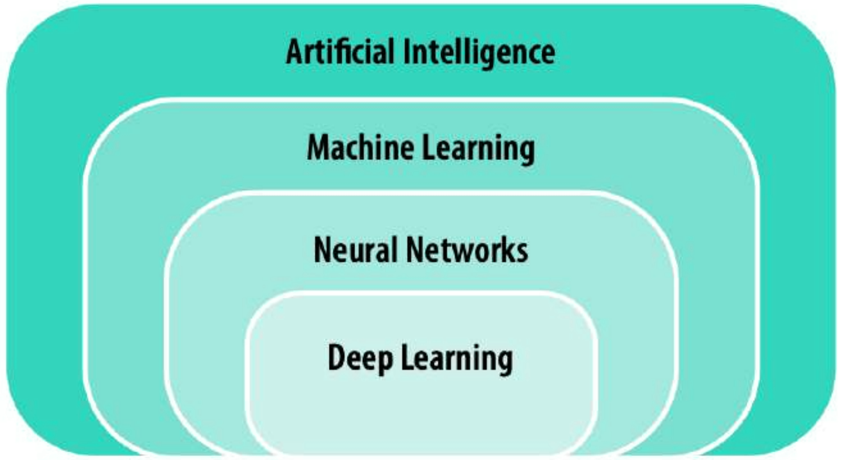

Figure 1.2. The relationship between Artificial Intelligence, Machine Learning, Neural Networks and Deep Learning © UNESCO "AI and education: Guidance for policy-makers"

1.3.4 | AI Technologies

The array of AI techniques outlined above serves as the foundation for numerous AI technologies that are now widely accessible. These technologies, as explored by Miao et al. (2021), have become integral to various sectors, and include the following:

[27] Ibid.
[28] Brownie J., 2019. A Gentle Introduction to Generative Adversarial Networks. Available at: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/ [Accessed on 17 March 2024]
[29] Ibid.
[30] This Person Does Not Exist, (n.d). [online] Available at: https://thispersondoesnotexist.com/ [Accessed on 17 March 2024]

- *Natural language processing (NLP)*: The use of AI to automatically interpret texts, including semantic analysis (as used in legal services and translation), and generate texts (as in auto-journalism).
- *Speech recognition:* The application of NLP to spoken words, including smartphones, AI personal assistants, and conversational bots in banking services.
- *Image recognition and processing*: The use of AI for facial recognition (e.g. for electronic passports); handwriting recognition (e.g. for automated postal sorting); image manipulation (e.g. for deep-fakes); and autonomous vehicles.
- *Autonomous agents*: The use of AI in computer game avatars, malicious software bots, virtual companions, smart robots, and autonomous warfare.
- *Affect detection*: The use of AI to analyse sentiment in text, behaviour and faces.
- *Data mining for prediction*: The use of AI for medical diagnoses, weather forecasting, business projections, smart cities, financial predictions, and fraud detection.
- *Artificial creativity*: The use of AI in systems that can create new photographs, music, artwork, or stories.

1.3.5 | Predictive and Generative AI

Considering that this research focuses on predictive and generative AI, it may be useful to briefly explore the differences between these two AI typologies.

The term generative AI refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data[31]. Belonging to unsupervised machine learning, generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics. Generative AI can generate multimodal contents, including but not limited to text, audio, image, video, and even three-dimensional models[32]. Some examples include ChatGPT for text, Midjourney or DAL-E for images, and DeepBrain or Sora for videos.

On the other hand, predictive AI forecasts future events or outcomes by analysing vast amounts of historical data trends to assign probability weights to the models. At its core, predictive AI relies on algorithms and statistical models to process and analyse vast amounts of data. Machine learning, a subset of AI discussed above, plays a crucial role in predictive AI, allowing the predictive systems to improve their accuracy over time as they process more data. In the international development predictive AI has sparked interest, deploying it in various contexts and sectors: from more general forecasting of weather conditions to more specific poverty prediction through satellite imagery[33].

These techniques and technologies do not operate in a vacuum but are deployed within larger and more complex social, cultural, economic, and political configurations. The world's complexity is also what

[31] Stefan Feuerriegel S., HartmannJ., Janiesch C, Zschech P., 2023. *Generative AI.* [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s12599-023-00834-7.pdf [Accessed on 17 March 2024]

[32] Ibid.
[33] Jean N., et al., 2016. *Combining satellite imagery and machine learning to predict poverty.* [online] Available at: https://www.science.org/doi/10.1126/science.aaf7894 [Accessed on 17 March 2024]

renders these techniques and technologies—sophisticated at first sight—fragile and brittle. We will attempt next to briefly capture some of the main critical aspects of AI capabilities.

SECTION 1.4

Critical Examination of AI Capabilities

As argued by McQuillan (2023) AI is more than a set of machine learning techniques; because "AI is never separate from the assembly of institutional arrangements that need to be in place for it to make an impact in society. AI is this layered and interdependent arrangement of technology, institutions and ideology, termed by him as 'apparatus" (McQuillan 2023: 1).

What McQuillan aims to do is to de-separate technical capabilities of AI from the real world. That is, to position the AI capabilities within the everyday life of society. It is precisely this displacement–from tech laboratories to the streets, schools, and offices of the city–that AI becomes fragile and brittle. The "wow factor" that we receive from AI capabilities when they recognise faces or win Go games, can be quickly turned into new kinds of failures with real harms in people's lives. For instance, the first pedestrian killed by a self-driving car was crossing the road while laboriously pushing a bicycle laden with their shopping bags. The high-end Volvo car being used by Uber as a test vehicle didn't recognise the person, a 50 years old Elaine Herzberg. Or, when on October 2022, Sebastian Galassi 26-year-old Italian university student, was making a Glovo[34] delivery, his bike was hit by a Land Rover. The young student died the next day. The day after his death, his cell phone received a message which read:

> "*We regret to inform you that your account has been terminated due to non-compliance with the terms and conditions. Glovo's goal is to offer an optimal experience to its couriers, partners and customers. In order to maintain a healthy and fair platform, it is sometimes necessary to take action when one of these users does not behave appropriately* ".

Tragic examples, like these, underscore one of the many issues with AI capabilities: their operations and outputs rely on correlations rather than causation. When they operate outside of their training data, their vulnerabilities can lead to nonsensical outcomes in the best-case scenarios or harmful events in the worst-case scenario.

[34] Glovo is a home delivery company, based in Spain, which uses autonomous couriers, called riders. It also has a presence in Italy. It has recently been purchased by Delivery Hero.

[35] Leetaru K., 2018. *Does AI Truly Learn And Why We Need to Stop Overhyping Deep Learning*. [online] Available at: https://www.forbes.com/sites/kalevleetaru/2018/12/15/does-ai-truly-learn-and-why-we-need-to-stop-overhyping-deep-learning/ [Accessed on 17 March 2024].

[36] Morozov E., 2023. *The problem with artificial intelligence? It's neither artificial nor intelligent*
. [online] Available at: https://www.theguardian.com/commentisfree/2023/mar/30/artificial-intelligence-chatgpt-human-mind [Accessed on 17 March 2024]

[37] Simonite T., 2021. *This Researcher says AI is Neither Intelligent nor Artificial.* [online] Available at: https://www.wired.com/story/researcher-says-ai-not-artificial-intelligent/ [Accessed on 17 March 2024]

There is also a tendency to promote a narrative that equates AI capabilities with human intelligence. This should be seen more as a part of the corporate sector's overselling hype rather than reality[35]. In fact, a growing number of scholars are arguing that AI is "neither artificial nor intelligent"[36][37]. AI's capabilities are dependent on human intelligence and labour, emanating from the human mind and body, in at least two aspects. First, AI operates thanks to the 'general intelligence' that it has captured from artists, musicians, programmers, writers, researchers, scientists, etc.. This vast array of creative work produced by humans has been fed to machine learning. And, through a combination of various learning techniques, mathematical operations, optimizations, and statistical estimations, it aims to produce a satisfactory output. Thus, we can say that the true 'intelligence' of AI is, in fact, human intelligence. The second aspect concerns human labour; the significant effort invested in developing, training, and maintaining AI technological capabilities is also considerable. For instance, to reduce the toxicity of AI chatbots like ChatGPT, OpenAI, the company who owns ChatGPT, outsourced Kenyan labourers earning less than $2 per hour[38]. Even the 'artificiality' of AI can be questioned considering the substantial amount of human labour committed to keeping AI technologies operational when they fall short.

The world we inhabit is a messy one. Is a world characterised by a polycrisis—multiple crises occurring simultaneously. If technologies like AI can fall short in simple tasks, such as misidentifying a photo of "an Israeli soldier holding down a young Palestinian while the boy's family tries to remove the soldier" as "People sitting on top of a bench together," (Katz, 2020, as cited in McQuillan 2023) what should we expect when these technologies and systems are deployed in development or humanitarian contexts? In such settings, situations can be messier and even more complex to navigate.

In the next section we will situate AI in the development and humanitarian contexts by exploring some of the main debates and trends both at its discourse and applicability, as well as governance.

CHAPTER 2
Situating Artificial Intelligence in Development

Is "AI for Good" merely a continuation of the "Data for Good" initiative, which contributed to the 'datafication' of development? How much of this effort is genuinely aimed at achieving good and helping the most vulnerable populations of the world, and how much is simply leveraging these terms for brand enhancement and market expansion? This chapter aims to situate AI within the expansive field of international development. First, we explore how AI emerged from the Digital for Development paradigm, also known as ICT4D 3.0, in tandem with the rise of Web 3.0 and other frontier technologies. Second, we delve into the 'AI for Good' discourse by capturing the main debates, trends and tensions, while critically analysing what genuinely constitutes 'good' and what merely serves as marketing tactics of 'washing'. Third, we assess the role, contributions, and impact of AI within the Sustainable Development Agenda, with a particular focus on the initiatives of the AI4SDG think tank. Fourth, we shift to humanitarian contexts to discuss the use of AI in anticipatory humanitarian action through predictive analytics, and we also consider the main ethical implications of employing AI predictive analytics in such settings. Finally, we outline the principal AI governance frameworks developed by governmental institutions and agencies as well as global and multilateral forums.

SECTION 2.1
AI and the Digital for Development Paradigm

In order to situate AI within the development sector, we should first trace the evolution of the Information and Communications Technologies for Development (ICT4D). To do this, we propose Richard Heeks' classification of ICT4D periods/paradigms (Heeks 2019), as follows:

1. ***Pre-digital paradigm***, also referred to as ICT4D 0.0, spans from the mid-1940s to the mid-1990s and predates modern technologies. More accurately, this paradigm should not be associated with ICT per se but rather with Information Technology (IT) only. The Internet was an expensive technology that required extensive infrastructure, hence only a few countries could afford it.
2. ***The ICT4D paradigm***, which includes: ICT4D 1.0 covering the period between mid-1990s until mid-/late-2000s and it is centred around the concepts of internet-connectedness and telecentres; and ICT4D 2.0 which encompasses the period between mid-/late-2000s until the present times and it represents, perhaps, the peak of deployments of ICTs in development. This period is accompanied also by the rise of Web 2.0..
3. ***Digital-for-development paradigm***, also referred to as ICT4D 3.0, emerged in late 2019. It is associated with reproduction, diffusion, mutation, and intensification of the dominant mode of the competitive markets, and hierarchical controls associated with capitalism and with traditional state-citizen relations. Simultaneously, the paradigm is also associated with growing examples for an alternative economics and an alternative politics (Heeks 2019).

[38] Perrigo B., 2023. *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. [online] Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed on 23 March 2024]

In Part 1 of the article "ICT4D 3.0? Part 1—The components of an emerging "digital for-for-development" paradigm", published in 2019, Heeks identifies three generations of digital infrastructure for development: 1) the first is largely based around *mobile phones*; 2) the second, is based around the *internet including Web 2.0 applications*; 3) the third, is based around *ubiquitous computing model of sensors, embedded processing and near-universal connectivity and widespread use of smart applications*.

While Heeks does not explicitly mention it, we can extend his discussion by noting that the third digital infrastructure points to the emerging Web 3.0, referred to as the 'semantic web.' According to Fan et al. (2013), Web 3.0 is marked by four main characteristics:

a) *Openness:* Enable users to access various platforms utilising only one account.
b) *Data privacy:* Strengthened via the decentralised structure of blockchain technology that protects user data ownership and eliminates reliance on third-party management platforms.
c) *Cooperation:* Facilitated through token incentives designed to reward content creators for their contributions and foster a more equitable platform.
d) *Interoperability:* Offers greater latitude in managing personal activities and engagements across varying digital environments.

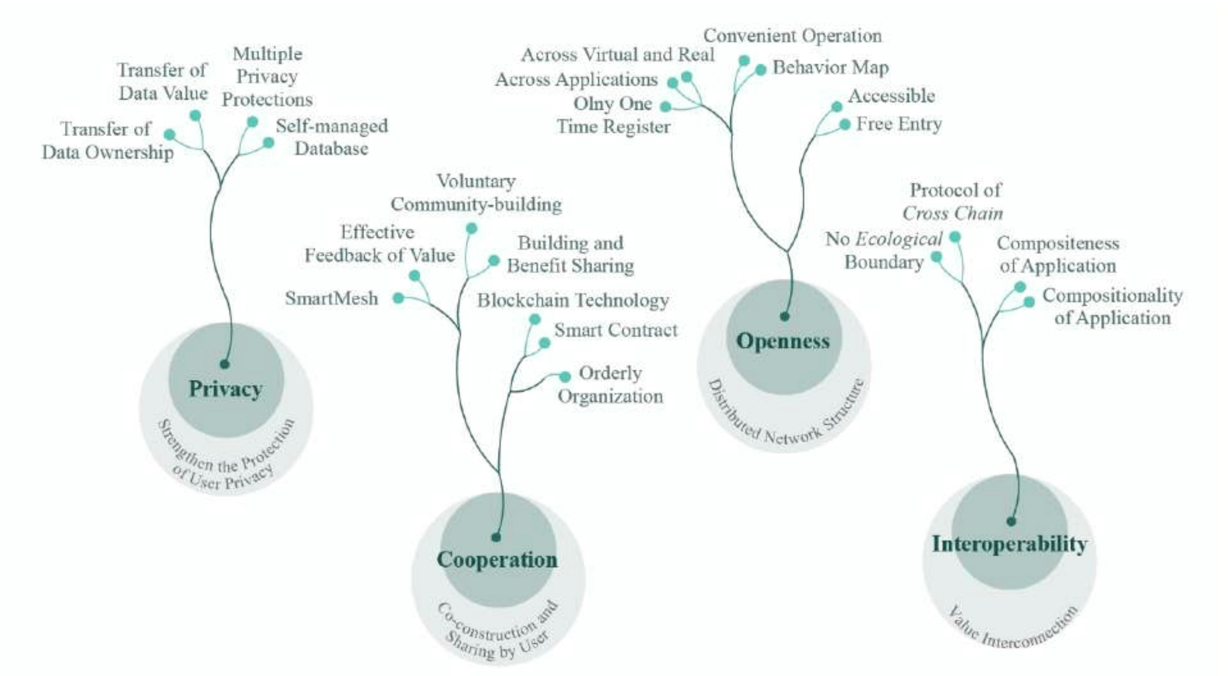

Figure 2.1. © Yuqing Fan et al., 2023, *The current opportunities and challenges of Web 3.0*

Web 3.0's promises have been high since its emergence, around late 2016, particularly by its brand of 'frontier technologies'. Frontier technologies refer to cutting-edge and advanced technological innovations that are at the forefront of scientific and technological progress. Following World Intellectual Property Organisation's categorisation, we can identify three overarching types of frontier technologies[39]:

1. *Digital technologies* such as: The Metaverse, Augmented Reality (AR), Virtual Reality (VR), Blockchain, Artificial Intelligence (AI), Big Data, Quantum Computing, etc.
2. *Physical technologies* which include: autonomous driving, 3D printing, hardware innovations such as robotics or 5G technology, and more.
3. *Biological technologies* such as: bioprinting, organoids, genetic engineering, human augmentation and the brain-computer-interface, etc.

We can thus position AI as part of Web 3.0-based digital frontier technologies within the Digital for Development (D4D) paradigm.

However, Web 3.0 and its frontier technologies arrived at a crucial moment for international development—the initiation of the Sustainable Development Agenda through its 17 Sustainable Development Goals (SDGs), which replaced Millenium Development Goals (MDGs) in 2015. Like with MDGs through Web 2.0, international development actors have once again placed their emphasis, hopes, and enthusiasm on the innovation potential of frontier technologies. For instance, the UN's World Economic and Social Survey (2018) recognises the breakthrough role that frontier technologies may play in sustainable development, yet it also acknowledges the policy and ethical challenges that accompany the regulation and governance of these technologies.

The United Nations Conference on Trade and Development's (UNCTAD) "Technology and Innovation Report" (2023) which covers seventeen frontier and green technologies, is conceptualised around the concept of 'green innovation'. It urges developing countries to "catch up, reduce poverty, and at the same time help mitigate climate change and set the world on a more sustainable course" (UNCTAD 2023: 1). The report shows also how the knowledge landscape and patents are dominated by the United States and China with a "a combined 30 per cent share of global publications and almost 70 per cent of patents" (UNCTAD 2023: 3). Additionally, the report indicated that "the most mature technology is AI", and "most patents for this technology were applied in 2014" (UNCTAD 2023: 4).

[39] World Intellectual Property Organisation, n/d. *Frontier Technologies Fact-sheet.* [pdf] Available at: https://www.wipo.int/about-ip/en/frontier_technologies/pdf/frontier-tech-6th-factsheet.pdf [Accessed on 29 March 2024]

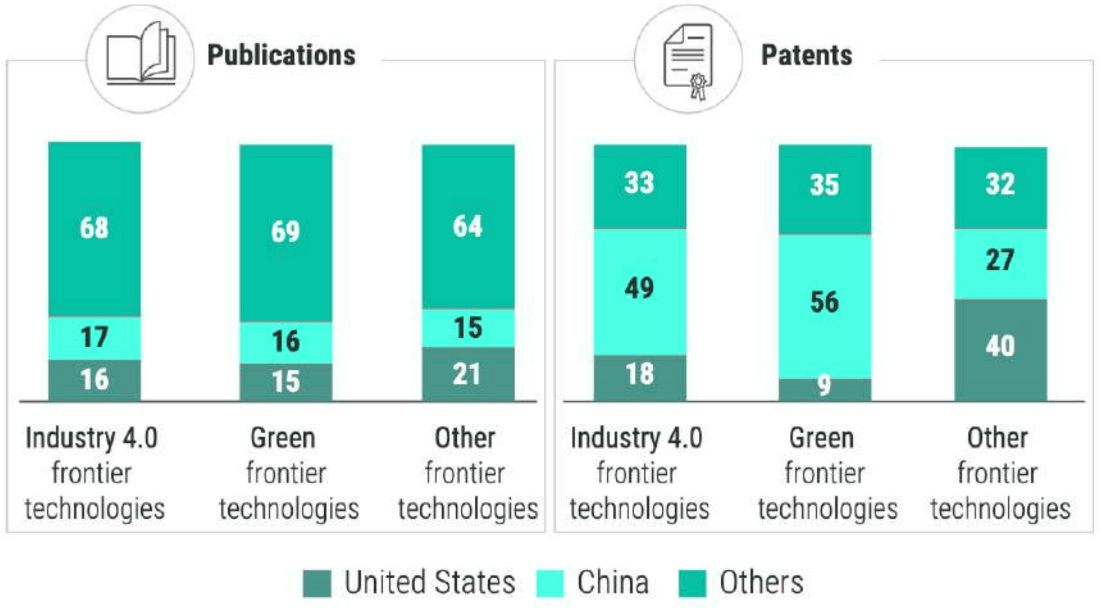

Figure 2.2. © *Country share of publications and patents, by frontier technology (percentage).* UNCTAD calculations based on data from Scopus and PatSeer.

A first reflection emerging from the UNCTAD report is that the expected leapfrogging of developing countries to catch up with frontier technologies may be unequal, given the hegemonic dominance of the United States and China in both knowledge production and patent ownership. Secondly, we observe how green frontier technologies are positioned as 'technical solutions' to broader, structural, and complex problems such as the climate crisis. Lastly, and in contrast to UNCTAD's report, we can argue that frontier technologies represent another wave of technological innovations seeking "market expansion and converting unused capacity into business assets on the premise that new technology is the gateway to hope" (Nederveen [2001] 2010: 173).

Nonetheless, in practice too, international development organisations and agencies have begun to embrace frontier technologies. UNICEF has launched its Blockchain programme[40] to explore its potential positive impacts on children and young people globally, as well as Drone programme[41] addressing transport, connectivity, and enhanced emergency preparedness. Similarly, the UNDP has "cautiously adopted some of the best proposals offered by AI" in line with their Digital Strategy[42]. The Centre for Humanitarian Data has created a catalogue of predictive models in the humanitarian sector to bolster transparency and knowledge sharing.[43] Additionally, the United Nations (UN) has established the AI for Sustainable Development Goals (AI4SDG) Think Tank, a global observatory that collects "AI projects and proposals that impact UN Sustainable Development Goals, both positively and negatively."[44]

These initiatives underscore the strong appeal of digital frontier technologies, especially AI, within international development. They exemplify how the Digital for Development (D4D) paradigm has enthusiastically adopted the new wave of AI and other technologies under the "AI for Good" banner. However, this raises critical questions: Is "AI for Good" merely a continuation of the "Data for Good" initiative, which contributed to the 'datafication' of development? How much of this effort is genuinely aimed at achieving good and helping the most vulnerable populations of the world, and how much is simply leveraging these terms for brand enhancement and market expansion?

We will explore these questions in the following section by capturing some of the main trends, debates and tensions within the "AI for Good" agenda by examining the main trends, debates, and tensions within the "AI for Good" discourse, aiming to discern its true impact and intentions.

SECTION 2.2

'AI for Good' — An Empty Signifier?

'AI for Good' is a direct extension of previous 'for Good' initiatives such as 'Tech for Good', 'Data for Good', 'ICT4D', and even 'Data Science for Good'. These are more than catchy slogans; they are established fields of research and practice with their own discursive constructions. These fields include actors such as UN agencies, particularly the International Telecommunications Union (ITU), the private sector, academia, and non-profit organizations. A recent survey of the field found over 1000 published papers on AI4SG topics (Shi, Wang, and Fang 2020, p. 1), growing from 18 papers in 2008 to 246 papers in 2019 (Shi, Wang, and Fang 2020, p. 5, as cited in Holzmeyer 2021).

[40] UNICEF (n.d.). *Blockchain*. [online] Available at: https://www.unicef.org/innovation/blockchain [Accessed on 29 March 2024]
[41] UNICEF (n.d.). *Drones*. [online] Available at: https://www.unicef.org/innovation/drones [Accessed on 29 March 2024]
[42] United Nations Development Programme (UNDP) (n.d.). *Making AI Work for Us*. [online] Available at: https://feature.undp.org/making-ai-work-for-us/ [Accessed on 29 March 2024]
[43] Centre for Humanitarian Data (n.d.). *Catalogue for Predictive Models in the Humanitarian Sector*. [online] Available at: https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/ [Accessed on 29 March 2024]
[44] AI for SDGs Academy (n.d.). [online] Available at: https://ai-for-sdgs.academy/ [Accessed on 29 March 2024]

In 2016, the 'AI for Good Foundation' was launched with the aim of driving "forward technological solutions that measure and advance the UN's Sustainable Development Goals."[45]. In 2017, the ITU's first AI for Good Global Summit took place with a goal "to identify practical applications of AI to advance the United Nations Sustainable Development Goals and scale those solutions for global impact."[46]. The private sector quickly tapped into the AI for Good discourse. The following year, in 2018, Google launched the AI for Social Good programme which consisted of "a group of researchers, engineers, volunteers, and other people across Google with a shared focus on positive social impact [...] to drive positive change in underserved communities [...]."[47] Other private tech companies such as Facebook, IBM, and Intel have pages on "AI for social good"[48], and Microsoft has one about "AI for good."[49]

This timeline of events reflects the rapid institutionalization and mainstreaming of the 'AI for Good' discourse, underscoring both its global appeal and its adaptability across sectors. However, the quick adoption by private tech companies raises critical questions about the underlying motivations and power dynamics at play. While the alignment with the UN's Sustainable Development Goals suggests a focus on collective welfare, the corporate emphasis on social impact initiatives may also serve to reinforce brand image and deflect scrutiny from ethical or regulatory concerns.

Moore (2019) rightly asks, "which good and for whom"? For Moore, AI systems are inherently political, embedded within the power structures that shape their development and deployment (Moore 2019: 2). This critique draws attention to the fact that 'good' cannot be universalized; it depends on whose needs and values are prioritized, often sidelining marginalized communities. Furthermore, Moore argues that the AI for Good label "distracts from the larger world in which AI exists" (Moore 2019: 4), emphasizing that framing technological solutions as inherently benevolent can obscure the systemic inequalities they may perpetuate. Moore provocatively suggests using the term "for not bad" instead, reflecting the need to avoid technology's potential harms rather than relying on vague claims of goodness (Moore 2019: 5). This rephrasing challenges the field to grapple with the ethical compromises involved in AI development and urges stakeholders to address its inherent risks transparently.

Whereas, Green (2019) highlights another critical gap: the lack of a rigorous methodology within computer science for assessing the relationship between algorithmic interventions and their long-term social impacts. According to Green (2019: 3):

[45] AI4Good (n.d.). *About Us*. [online] Available at: https://ai4good.org/about-us/ [Accessed on 29 March 2024]
[46] International Telecommunication Union (ITU) (n.d.). *About AI for Good*. [online] Available at: https://aiforgood.itu.int/about-ai-for-good/#:~:text=The%20goal%20of%20AI%20for,United%20Nations%20platform%20on%20AI [Accessed on 29 March 2024]
[47] Google AI (n.d.). *Why AI*. [online] Available at: https://ai.google/why-ai/ [Accessed on 29 March 2024]
[48] Facebook (n.d.). *AI for India*. [online] Available at: https://fbaiforindia.splashthat.com [Accessed on 29 March 2024]; Intel (n.d.). *AI for Social Good*. [online] Available at: https://www.intel.ai/ai4socialgood/ [Accessed on 29 March 2024]; IBM (n.d.). *AI for Social Good: Advantage Reports*. [online] Available at: https://www.ibm.com/watson/advantage-reports/ai-social-good.html [Accessed on 29 March 2024]
[49] Microsoft (n.d.). *AI for Good*. [online] Available at: https://www.microsoft.com/en-us/ai/ai-for-good [Accessed on 29 March 2024]

> Pursuing social good without considering the long-term impacts can lead to great harm, however: what may seem good in an immediate, narrow sense can be actively harmful in a broader sense. In other words, the dichotomy between the idealized perfect and the incremental good is a false one: it is only through debating and refining our imagined conditions of the perfect society—an essential component of politics—that we can conceive of and evaluate potential incremental goods. Because there is a multiplicity of imagined perfects, which in turn suggest an even larger multiplicity of incremental goods, any incremental good must be evaluated based on what type of society it promotes in both the short and long term.

What does 'AI for Good' then stand for? Namely, some of the most critical problems that international development aims to tackle/solve are politically, socially, and economically produced problems. Consequently, it is at the level of political, social, and economic intervention that the most impactful changes must occur. Moore's critique highlights how these challenges cannot be separated from the global structures of power that produce them, while Green emphasizes the necessity of creating robust methodologies to measure incremental progress in addressing these systemic issues. The way international development addresses its own interventions within the framework of Rights–the so-called 'Economic, Social and Cultural rights'[50]–is crucial for transforming the lives of those most in need. In practical terms, 'the Good' should stand about ways how international actors ensure the rights to adequate food, housing, universal education, healthcare, water and sanitation, decent work, etc..

At the same time, we may go further to suggest that AI perhaps is being instrumentalized for the depoliticization of the aforementioned issues and their reframing as 'technological fixes.' This aligns with Holzmeyer's (2021) critique that "the label for social good" directs "attention away from the cultural, economic and political power commanded by Big Tech companies and the harmful externalities inherent in burgeoning AI systems." Holzmeyer's analysis reminds us that behind the optimistic rhetoric lies the concentration of power in the hands of corporations whose priorities may not align with global equity. Their role in framing these issues as solvable by technology risks oversimplifying complex systemic problems and obfuscating the need for broader structural reforms.

'AI for Good,' as it currently stands, often functions as an empty signifier, lacking a clear definition, orientation, or historical grounding. To assess the true 'goodness' of 'AI for Social Good'—and to transform 'AI for Good' from an empty signifier into one with concrete meaning and evaluative rigor—we must establish robust benchmarks for evaluating the performance and societal impact of such technologies. Aligning these initiatives with frameworks like the Universal Declaration of Human Rights could help ground them in universally recognized principles. Similarly, the Principles of Digital Development offer a methodological approach to ensure that technological projects are contextually appropriate and ethically sound. Nevertheless, the most practical and comprehensive starting point for these benchmarks is likely the UN's Agenda for Sustainable Development, which provides a well-defined structure through its 17 goals, 169 targets, and 247 indicators. This alignment would allow 'AI for Good' to evolve into a meaningful framework, capable of addressing complex global challenges with both accountability and measurable outcomes.

[50] Office of the High Commissioner for Human Rights (OHCHR) (n.d.). *Economic, Social and Cultural Rights*. [online] Available at: https://www.ohchr.org/en/human-rights/economic-social-cultural-rights [Accessed on 29 March 2024]

SECTION 2.3

AI and the Sustainable Development Goals

The difficulty to assess what constitutes AI for Good is highlighted, among many others, also by Cowls et. al. (2021) in their article "A definition, benchmark and database of AI for social good initiatives". One reason for this vagueness in the definition of AI for Good is "the lack of a reliable benchmark with which to assess its success" (Cowls et al., 2021, pg.1). This is why the authors propose to use SDGs as a benchmark to evaluate AI for Good initiatives which they name it as 'AIxSDGs', and providing three main reasons:

> First, the SDGs offer clear, well defined and shareable boundaries to identify positively what is socially good AI (what should be done, as opposed to what should be avoided), although they should not be understood as indicating what is not socially good AI. Second, the SDGs are internationally agreed goals for development, and have begun informing relevant policies worldwide, so they raise fewer questions about relativity and cultural dependency of values. Although they are of course improvable, they are nonetheless the closest thing we have to a humanity-wide consensus on what ought to be done to promote positive social change and the conservation of our natural environment. Third, the existing body of research on SDGs already includes studies and metrics on how to measure progress in attaining each of the 17 SDGs, and the 169 associated targets defined in the 2030 Agenda for Sustainable Development. These metrics can be applied to evaluate the impact of AI×SDGs projects (2021: 3).

To use the proposed UN SDGs as a benchmark for assessing AI for Good initiatives, the authors conducted an international survey of AIxSDGs, which ran from July 2018 to November 2020. This survey identified 108 projects in English, Spanish, and French that matched the predefined five criteria: alignment with SDGs; use of some actual form of AI; real-life projects tested in the field for at least six months; documented positive impact; and no or minimal evidence of counter-indications or negative side effects.

The survey shows that all SDGs are already being addressed by at least one AI-driven project, with projects operating across five continents. However, the distribution of projects across the SDGs is uneven. For instance, "SDG 3 ('Good Health and Well-Being') leads the way, while SDGs 5 ('Gender Equality'), 16 ('Peace, Justice and Strong Institutions'), and 17 ('Partnerships for the Goals') appear to be addressed by fewer than five projects" (2021: 6).

In many cases, the impact of these projects was 'only' local or at an early stage, raising questions about how best to—or indeed in each case whether to — 'scale up' existing solutions to apply them at regional or even global levels (Cowls et al., p. 4). The question of 'scale' is highlighted in another report[51] published in 2018 by McKinsey Global Institute. In it, the authors admit that "for now AI capabilities are being tested and deployed, and they already show promise across a range of domains" (Chui et al., p. 42). However, they emphasise "the need to focus on scaling up AI solutions and overcoming the bottlenecks and market failures", suggesting three main areas in which stakeholders could make a meaningful contribution to further the use of AI for the benefit of society (Chui et al., p. 42-43):

[51] Chui M., et. al. 2018. *Notes from the Frontier: Applying AI for Social Good.* McKinsey Global Institute. [pdf] Available at: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf [Accessed on 13 May 2024]

1. Improving data accessibility for social impact cases which will require active participation of data collectors and generators
2. Overcoming AI Talent shortages for solving technical challenges and implementing AI solutions
3. Addressing last-mile implementation challenges for deploying and sustaining AI models for social good, which will require ongoing technical support, organisational change management, and infrastructure improvements.

McKinsey's report, characterised by a techno-optimistic view and an AI accelerationist agenda, continues to push for rapid advancements and scaling of AI technologies. In its most recent report[52] published in May 2024, McKinsey Global Institute, has discovered a total of about 600 use cases, more than a threefold increase from 2018[53]. Even in this report, the 'scaling' of AI is seen as a challenge, stating that "72 percent of respondents to our expert survey, most efforts to deploy AI for social good to date have focused on research and innovation rather than adoption and scaling" (Bankhwal M., et. al., p.19). Following the trends of other reports and articles, SDG 3 ('Good Health and Well-Being') continues to lead as the most targeted with 165 AI use-cases, followed by SDG 16 ('Peace, Justice and Strong Institutions') with 55 use-cases.

Despite the high promises for AI technologies to 'help to address the major challenges, both social and environmental, facing humanity today' (Cowls et al., pg. 5) and the potential to 'catalyze progress on these pressing social issues' (Bankhwal M., et. al., pg. 1), the situation in reality seems more fragile. For instance, The Sustainable Development Goals Report 2022[54] highlights the cascading and interlinked crises such as COVID-19, climate change, and conflicts have negatively impacted all Sustainable Development Goals (SDGs), reversing years of progress in eradicating poverty and hunger, improving health and education, and providing basic services. A year later, in 2023, the SDG Report[55] provided alarming numbers about the state of the 17 Goals: "Of the approximately 140 targets that can be evaluated, half of them show moderate or severe deviations from the desired trajectory. Furthermore, more than 30 percent of these targets have experienced no progress or, even worse, regression below the 2015 baseline." When it comes to the climate crisis, 2023 was the warmest year in the 174-years observational record[56]. U.N. High

[52] Bankhwal M., et. al., 2024. *AI for social good: Improving lives and protecting the planet.* McKinsey Global Institute. [pdf] Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/ai-for-social-good#/download/%2F~%2Fmedia%2Fmckinsey%2Fbusiness%20functions%2Fquantumblack%2Four%20insights%2Fai%20for%20social%20good%2F2024%2Fai-for-social-good-improving-lives-and-protecting-the-planet-v2.pdf%3FshouldIndex%3Dfalse [Accessed on 20 May 2024]

[53] In a 2018 report by McKinsey Global institute the number of high-potential use cases of AI for Good projects were only 170. For other use-cases of AIxSDGs, see: Ferili S., et. al. 2021. *Artificial Intelligence for Sustainable Development.* Available only in Italian language: https://www.cnr.it/sites/default/files/public/media/attivita/editoria/VOLUME%20FULL%2014%20digital%20LIGHT.pdf [Accessed on 13 May 2024]

[54] United Nations (2022). *The Sustainable Development Goals Report 2022*. [pdf] Available at: https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf [Accessed on 15 May 2024]

[55] United Nations (2023). *The Sustainable Development Goals Report 2023: Special Edition*. [pdf] Available at: https://desapublications.un.org/publications/sustainable-development-goals-report-2023-special-edition [Accessed on 15 May 2024]

[56] World Meteorological Organization (WMO), (2023). *Climate Change Indicators Reached Record Levels in 2023*. [online] Available at: https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-

Commissioner for Human Rights Volker Türk has recently painted a frightening picture of a world where proliferating conflicts are devastating the lives of millions of civilians, stating that "a quarter of humanity is caught in 55 global conflicts"[57], representing "the largest number of violent conflicts since 1946" (SDG Report 2022, pg. 58).

In light of the profound and interlinked polycrises undermining the 2030 Agenda for Sustainable Development, global peace, and environment, and given the uneven impact and scalability issues of current AI for Good initiatives, we must ask: What tangible actions can AI take to effectively navigate this complex landscape and genuinely contribute to overcoming these global predicaments and gridlocks?

Box 2.1. AI4SDGs Think Tank

The AI4SDG Think Tank[58] is an initiative designed to align artificial intelligence innovations with the Sustainable Development Goals (SDGs) set by the United Nations. Its primary mission is to harness AI technologies to promote sustainable development across various sectors globally. The Think Tank acts as both a repository and an analytical engine, evaluating AI projects and proposals for their impact on achieving the SDGs.

Key aspects of the AI4SDG Think Tank include:

- **Global Repository**: Collecting and showcasing a wide range of AI projects that impact the SDGs, both positively and negatively.
- **Analytical Engine**: Providing detailed evaluations of AI initiatives to determine their effectiveness and scalability in addressing specific SDGs.
- **Promotional Role**: Advocating for the positive use of AI to advance sustainable development, while also addressing potential negative impacts.
- **Educational Resource**: Offering insights and data to researchers, policymakers, and the public on how AI can be leveraged for social good.

As of 08 May 2024, the repository of the project counts 16,430 cases, which can be accessed online at AI for SDGs Academy (https://ai-for-sdgs.academy/). However, it is important to note that the projects are unevenly distributed across the SDGs.

The top three SDGs with the highest concentration of projects are:

- SDG 3 (Good Health and Well-being) with 3,221 projects;
- SDG 12 (Responsible Consumption and Production) with 3,075 projects;
- and SDG 9 (Industry, Innovation, and Infrastructure) with 2,724 projects.

Conversely, the SDGs with the least concentration of projects are:

- SDG 17 (Partnerships for the Goals) with 91 projects,
- SDG 6 (Clean Water and Sanitation) with 246 projects,
- and SDG 14 (Life Below Water) with 248 projects.

Interestingly, and in line with global trends[59], most of the projects are based in the United States (187 projects), followed by China (41 projects) and the United Kingdom (31 projects).

This geographical imbalance is recorded by a comprehensive survey of over 1000 AI4SGa applications between 2008 and 2020. Shi et al. (2010: 45) found that the overwhelming majority of projects were US-based, while the few projects that focused on Africa were run by US researchers. As authors of the research warn: "a field dominated by people from a certain culture will at best fail to recognize some key problems and considerations in the culture that the research is applied to, and at worst form an exclusive and biassed climate within the field".

SECTION 2.4

AI in Humanitarian Contexts

If deep learning is one of the most interesting and dynamic aspects of AI, humanitarian relief is one of the most dynamic and important aspects of the development field. The first uses a variety of mathematical minimisations and statistical techniques to enable machines to learn from vast amounts of data and provide a most plausible output or prediction, while the latter employs human's sense-making, reasoning, meaning processing and empathy to make decisions.

Humanitarian assistance is designed to alleviate human suffering, assist in recovery from disasters and conflicts, and provide support to vulnerable groups. Humanitarian operations take place in complex and messy environments characterised by unpredictable conditions, scarce resources, diverse cultural contexts, destruction of key infrastructure, lack of reliable data, ongoing wars, and other crises.

To address these challenges, humanitarian actors have always sought to use new technologies and novel methods to increase their efficiency during relief efforts. The emergence of Web 2.0 has brought many opportunities for humanitarian actors. For instance, the Kenyan-based crowdmapping tool "Ushahidi" played a key role during the devastating 2010 Haiti earthquake[60]. Similarly, UNICEF's RapidPro, an open-source software that allows the user to easily build and scale mobile- based applications from anywhere in the world, has been used in different humanitarian contexts[61] with positive results.

wmo#:~:text=2023%20was%20the%20warmest%20year%20in%20the%20174%2Dyear%20observational,above%20the%201850%E2%80%931900%20average [Accessed on 15 May 2024]

[57] Al Jazeera ,(2023). *UN's Volker Turk: A Quarter of Humanity is Caught in 55 Global Conflicts*. Talk to Al Jazeera, 22 December 2023. [online] Available at: https://www.aljazeera.com/program/talk-to-al-jazeera/2023/12/22/uns-volker-turk-a-quarter-of-humanity-is-caught-in-55-global-conflicts [Accessed on 15 May 2024]

[58] AI for SDGs Academy (n.d.). [online] Available at: https://ai-for-sdgs.academy/ [Accessed on 15 May 2024]

[59] AI Index Report 2024, published by Stanford University's Human-Centred AI, highlights the U.S.'s domination in the field of AI: "In 2023, 61 notable AI models originated from the U.S.-based institutions, far outpacing the European Union's 21 and China's 15.". The full report is accessible online: https://aiindex.stanford.edu/report/ [Accessed on 13 May 2024].

[60] Ushahidi blog, 2010. *Crisis Mapping Haiti: Some Final Reflections.* [online] Available at: https://www.ushahidi.com/about/blog/crisis-mapping-haiti-some-final-reflections/ [Accessed on 14 May 2024]; Ushahidi blog, 2012. *Haiti and the Power of Crowdsourcing*. [online] Available at: https://www.ushahidi.com/about/blog/haiti-and-the-power-of-crowdsourcing/ [Accessed on 14 May 2024]

[61] Elhra, 2015. *Text messages used to deliver humanitarian aid*. [online] Available at: https://www.elrha.org/news-and-blogs/text-messages-used-deliver-humanitarian-aid/ [Accessed on 14 May 2024]

If in the early 2010s 'humanitarian innovation' was the buzzword in humanitarianism, this is now replaced by AI, particularly with the mainstreaming of generative AI. Actors in the development sector in general, and the humanitarian field in particular, tend to compete in adopting new ways of 'doing business,' which in turn may increase funding opportunities. This obsessive love of novelty is what the scholar Tom Scott-Smith referred to as 'humanitarian neophilia'[62]:

> [Humanitarian neophilia] can be used in a positive as well as a negative sense, describing people who are quick to adapt to new technologies as well as those who have an uncritical desire for the latest gadgets, those who are creative and innovative as well as those who fail to learn from the past. When applied to humanitarianism, I use this term to embrace all these features, but also, as will become clear, to describe an ideology that combines New Left and New Right with techno-utopian fervour. 'Humanitarian neophilia' [...] designates a distinctive approach to aid, which combines an optimistic faith in the possibilities of technology with a commitment to the expansion of markets (Scott Smith, 2016: pg. 2).

Following Scott-Smith's argument, we can claim that today we are seeing the emergence of a new neophilia in the humanitarian field, under the name of 'AI neophilia'. As the humanitarian field continually seeks to deploy new technological innovations for people most in need, AI technologies are constantly seeking to expand their applicability in various fields, contexts and circumstances.

2.4.1 | *Overview of AI's Use in Humanitarian Contexts*

The use of AI, including machine learning and other advanced statistics, are not novel in the humanitarian sector. From assisting displaced persons through automated chatbots to leveraging AI-supported disaster mapping for emergency response, narrow applications of AI have been utilised in humanitarian response for several years.[63]

However, the landscape is rapidly evolving. The data revolution that has seen the exponential growth of data sets relevant to development and humanitarianism is, perhaps, the main catalyst of the proliferation of AI in the humanitarian field (Pizzi, Romanoff, Engelhardt, 2021, pg. 151). This surge in data availability, coupled with increasing availability and accessibility of AI technologies to the general public, is poised to support humanitarian efforts. Simultaneously, the deployment of AI in humanitarian contexts comes with risks, contradictions, ethical dilemmas and challenges. Humanitarian actors, therefore, will need to navigate the deployment of a complex technology like AI in complex contexts such as the humanitarian sector.

AI is applied in the humanitarian sector in multiple ways, making it difficult to draw aggregate inferences or make broad claims about its overall efficacy. Instead, to facilitate meaningful debate and understanding, it is crucial to be specific about the various AI applications within the humanitarian field.

[62] Scott-Smith T., 2016. *Humanitarian neophilia: the 'innovation turn' and its implications. Third World Quarterly*, *37*(12), 2229–2251. [doi] Available at: https://doi.org/10.1080/01436597.2016.1176856 [Accessed on 14 May 2024]

[63] UN OCHA, Centre for Humdata, 2024. *Briefing note on Artificial Intelligence and Humanitarian Sector*. [pdf] Available at: https://centre.humdata.org/note-briefing-note-on-artificial-intelligence-and-the-humanitarian-sector/ [Accessed on 24 May 20234]

Many of these applications remain speculative, and some raise ethical, operational, or contextual concerns.

Therefore, to facilitate a comprehensive understanding of AI applications in the humanitarian sector, it is beneficial to classify these tools and applications into a detailed classification that comprises several distinct yet overlapping categories, each representing a unique aspect of AI utilisation in humanitarian efforts. This classification does not aim to be a state-of-the-art taxonomy of AI in the humanitarian field—a rather unexplored area that requires further research—but rather a starting point that helps in identifying the specific AI technologies employed in humanitarian contexts. It provides a structured approach to exploring the diverse landscape of AI in humanitarian contexts, thereby facilitating a meaningful debate and understanding around AI in the humanitarian sector. In the next tables below, we present some of the classifications of AI use in the humanitarian sector.

Table 2.1. Classification of Predictive AI Use in the Humanitarian Sector

1. Predictive AI: using techniques such as neural networks, regression analysis and decision trees, predictive AI analysis historical and real-time data to forecast future events, enabling proactive responses to humanitarian crises. Predictive analytics can also support decision making by examining data or content to answer the question "What should be done?" or "What can we do to make ___ happen?".[64]

Use Cases	Brief Description with Examples
Disaster Prediction and Warning Systems	Predictive AI models predict natural disasters like hurricanes and floods. Examples: - **InaSAFE**[65] is an open source platform developed in Indonesia that produces realistic natural hazard impact scenarios for better planning, preparedness and response activities. Impact-based Forecasting, which is a functionality within the platform, will generate an estimate of districts, number of people and houses likely to be affected in the event of a flood with a certain return period. - **510 Typhon Model**[66] is a predictive model to identify high priority areas in the wake of a natural disaster. The model uses a large number of factors to assess this including but not limited to socio-economic factors, housing types, impact data and administrative boundaries.
Disease Outbreak Prediction	Predictive AI analyses global or region-specific health data to forecast disease outbreaks. Examples:

[64] UN OCHA, n.d. Predictive Analytics. [online] Available at: https://centre.humdata.org/glossary-2/predictive-analytics/ [Accessed on 24 May 2024]

[65] Inasafe (n.d.). *Inasafe Home*. [online Available at: https://inasafe.org/home/index.html [Accessed: 24 May 2024]

[66] 510 Global (2021). *Impact-Based Forecasting for Typhoons in the Philippines*. [online] Available at: https://510.global/2021/10/impact-based-forecasting-for-typhoons-in-the-philippines/ [Accessed: 24 May 2024]

	- **BlueDot**[67] is a private tech intelligence company, specialised in infectious disease intelligence, leveraging advanced data analytics and machine learning to predict and track the spread of infectious diseases. The platform integrates diverse data sources, including global airline ticketing data, climate data, and news reports, to identify potential outbreaks early and provide actionable insights. This enables public health officials, governments, and organisations to make informed decisions for mitigating risks and protecting populations from emerging health threats. - **Cholera Artificial Learning Model (CALM)**[68] is a predictive model that forecasts the exact number (with an error margin of 4.787 cases per 10,000) cholera cases in any given Yemeni governorate will experience for multiple time intervals ranging from 2 weeks to 2 months.
Food and Famine Prediction	Predictive AI models analyse various data points such as weather patterns, crop yields, market prices, and socio-economic factors to forecast potential food shortages and famines. Example: - **Famine Early Warning Systems Network (FEWSNET)**[69] is an early warning and analysis of food insecurity - livelihood based model, providing evidence-based warning information and analysis of food insecurity and its drivers worldwide. A key component of monitoring is satellite based remote sensing. Mixed methods analysis (quantitative and qualitative analysis) enables FEWSNET to build reliable scenarios and Integrated Food Classification (IPC).
Conflict Prediction and Prevention	Predictive AI analyses political, social, and economic data to predict the likelihood of conflicts or violence in certain regions, enabling early intervention and conflict prevention measures. Example: - **Violence Early-Warning System (ViEWS)**[70] is a cutting-edge conflict prediction system that generates monthly forecasts for violent conflicts across the world up to three years in advance. The model can predict armed conflict involving states and rebel groups, armed conflict between non-state actors, and violence against civilians. More importantly the model can disaggregate results into three different levels of analysis: national level, subnational level and actor level.

[67] BlueDot (n.d.). *Home*. [online] Available at: https://bluedot.global/ [Accessed: 24 May 2024]
[68] Lambert GA iGEM Team (2018). *CALM Model*. [online] Available at: https://2018.igem.org/Team:Lambert_GA/CALM_MODEL#target6 [Accessed: 24 May 2024]
[69] FEWS NET (n.d.). *Home*. [online] Available at: https://fews.net/ [Accessed on 24 May 2024]
[70] Uppsala University (n.d.). *VIEWS: Violence Early-Warning System*. [online] Available at: https://www.uu.se/en/department/peace-and-conflict-research/research/views/ [Accessed: 24 May 2024]

	- **CrisisWatch**[71] is a global conflict tracker, an early warning tool designed to help prevent deadly violence. It keeps decision-makers up-to-date with developments in over 70 conflicts and crises every month, identifying trends and alerting them to risks of escalation and opportunities to advance peace.
Migration and Displacement Forecasting	Predictive AI models forecast migration patterns and displacement due to conflicts, environmental changes, or economic factors, aiding in planning and resource allocation for affected populations. Example: - **Project Jetson**[72], developed by UNHCR's Innovation Service, is a machine learning-based experiment that provides predictions on the movement(s) of displaced people, combining data science, statistical processes, designing-thinking techniques, and qualitative research.
Environmental Degradation Monitoring	AI models predict environmental degradation, such as deforestation or desertification, by analysing satellite imagery and other environmental data, helping to mitigate adverse effects on communities. Example: - **Global Forest Watch**[73] offers cutting-edge data, technology, and tools to empower individuals and organisations globally to protect and manage forests more effectively. By providing real-time information on forest changes, deforestation alerts, and historical data, the platform helps users monitor forest health, track illegal activities, and support sustainable forest management efforts.

Table 2.2. Classification of Generative AI Use in the Humanitarian Sector

2. Generative AI: employing techniques such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Natural Language Generation (NLG) to create new content based on existing data. The content generated can be text, image, audio, video, synthetic data or code generation. Generative AI is used in the humanitarian field in multiple ways, such as: synthesise data, generate multilingual reports, simulate disaster scenarios, and produce creative content for awareness.

[71] Crisis Group (n.d.). *CrisisWatch*. [online] Available at: https://www.crisisgroup.org/crisiswatch [Accessed: 24 May 2024]
[72] UNHCR (n.d.). *Jetson: Predictive Analytics Tool*. [online] Available at: https://jetson.unhcr.org/ [Accessed on 24 May 2024]
[73] Global Forest Watch (n.d.). *Home*. [online] Available at: https://www.globalforestwatch.org/ [Accessed on 24 May 2024]

Use Cases	Brief Description with Examples
Natural Language Processing (NLP) / Natural Language Understanding (NLU)	Generative AI helps with generating translations and text for real-time communication in multiple languages during humanitarian operations. Example: - **LacunaFund**[74] language datasets create openly accessible text and speech resources that fuel natural language processing technologies in diverse languages across low- and middle-income contexts globally. - **The Internal Displacement Event Tagging Extraction and Clustering Tool (IDETECT)**[75] works by automating many of the steps that the monitoring team takes from identifying information on internal displacement and entering it in our database. First, IDETECT reads the world's news, as well as UN and NGO reports, and filters it keeping only articles that are likely to relate to internal displacement. Next, it begins analysing the text and extracts key bits of information including: type of displacement, cause, location, etc.. The team of monitoring experts reviews IDETECT's outputs and validates the suggestions or corrects them, and eventually ends in the database.
Scenario Simulation	Generative AI can assist in creating detailed simulations of disaster scenarios to aid in planning and preparedness. Example: - No real-world examples have been found that were deployed by humanitarian organisations or in humanitarian contexts.
Miscellaneous	Generative AI can streamline a wide range of administrative and operational tasks within humanitarian organisations. Some use cases examples include[76]: - **Generic use**: generate program updates, fundraising documents, donor reports, video call notes, job descriptions, review code, candidate fit, summarise research, impact reports, manage notes, organise themes in qualitative data - **Enhancing internal efficiency:** create templates, talking points, process routine work, note-taking - **Reporting:** match SDGs with outcomes, respond to audits, tackle reporting fatigue, automatic report tagging - **Communications:** design visual elements, edit content, create presentations, manage social media content

[74] Lacuna Fund (n.d.). *Language Datasets*. [online] Available at: https://lacunafund.org/datasets/language/ [Accessed on 24 May 2024]
[75] Internal Displacement Monitoring Centre (IDMC) (n.d.). *Monitoring Tools: Monitoring Platform*. [online] Available at: https://www.internal-displacement.org/monitoring-tools/monitoring-platform/ [Accessed on 24 May 2024]

[76] Motalebi N., Verity A., 2023. *Generative AI for Humanitarians.* [pdf] Available at: https://reliefweb.int/attachments/fe1baf61-33a6-4f55-955d-b977e8a76937/Generative%20AI%20for%20Humanitarians%20-%20September%202023.pdf [Accessed on 24 May 2024]

Table 2.3. Classification of Assistive AI Use in the Humanitarian Sector

3. Assistive AI: are designed to assist humanitarian organisations and responders in various tasks, ranging from data analysis to communications and crisis mapping, utilising techniques such as machine learning, knowledge representation and reasoning, computer vision and NLP.

Use Cases	Brief Description with Examples
Chatbots and Automated Helplines	AI-powered chatbots[77] and virtual assistants are perhaps the most used assistive AI. They provide real-time information and support to affected populations and aid workers. Additionally, they handle queries, offer guidance, provide emotional support, and direct users to relevant resources. Examples: - **The Emergency Telecommunications Chatbot (ETC)**[78] is the World Food Programme-led chatbot piloted in 2021, in response to high call volume in their call centre in Libya, and to optimise their service offering. TC Chatbot can use machine learning to communicate in a multilingual framework (e.g. English and Arabic). While ETC Chatbot functionality allows to store and learn from data, in practice it is limited in its ability to mitigate data protection risks. - **ParentText**[79] is a chatbot parenting intervention delivered through popular communication channels users trust: WhatsApp, Facebook Messenger, Telegram and SMS. The chatbot delivers evidence-based parenting support from Parenting for Lifelong Health, UNICEF, Helping Adolescents Thrive, and other parenting interventions that aim to reduce violence against children and improve child wellbeing. ParentText can easily be adapted to national parenting guidelines for local cultures and contexts. ParentText can be delivered as a standalone intervention or in a hybrid format, augmented within the PLH solutions ecosystem including WhatsApp support groups or in-person sessions.

[77] International Federation of Red Cross and Red Crescent Societies, 2023. *Chatbots in Humanitarian Contexts*. [pdf] Available at: https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623_CEA_Chatbots.pdf [Accessed on 24 May 2024]
[78] Emergency Telecommunications Cluster (ETC) (n.d.). *Introducing the New ETC Chatbot: AI-powered Buddy in Times of Crisis*. [online] Available at: https://www.etcluster.org/blog/introducing-new-etc-chatbot-ai-powered-buddy-times-crisis#:~:text=The%20ETC%20Chatbot%20has%20the,actions%2C%20enhancing%20resilience%20and%20empowerment [Accessed on 24 May 2024]
[79] Global Parenting Initiative (n.d.). *ParentText*. [online] Available at: https://globalparenting.org/parenttext [Accessed on 24 May 2024]

Crisis Mapping and Situational Awareness	AI tools aggregate and analyse data from social media, satellite images, and other sources to create real-time maps of disaster-affected areas. Example; - **fAIr**[80] is an open AI-assisted mapping service developed by the Humanitarian OpenStreetMap Team (HOT) that aims to improve the efficiency and accuracy of mapping efforts for humanitarian purposes. The service uses AI models, specifically computer vision techniques, to detect objects such as buildings, roads, waterways, and trees from satellite and UAV imagery. https://www.hotosm.org/updates/how-we-measure-the-effects-of-ai-assisted-mapping/
Language Translation Services	AI-powered translation tools help bridge language barriers during humanitarian crises, providing real-time translation for communication and documentation. Examples: - **Translators Without Borders**[81] has developed a chatbot in three languages to improve COVID-19 understanding in northeast Nigeria. The chatbot will allow people to message questions in their language and receive answers immediately and conversationally. - **Other AI-based translation softwares** are now widely available and used in humanitarian contexts. Examples include Google Translate, Bing, Microsoft Translator, DeepL, Reverso, Systran Translate, Amazon Translate, ChatGPT-4.0, ChatSonic, GPT-3 Playground, and YouChat. However, due to the monocultural nature of these tools, they should be used with caution and are best suited for non-sensitive, ad-hoc communication purposes. At present, they cannot replace or substitute human cultural mediators and professional translators.

Table 2.4. Classification of Optimisation AI Use in the Humanitarian Sector

4. Optimisation AI: is employed in the humanitarian field to help improve decision-making, resource allocation, and operational efficiency. By using a range of techniques such as linear programming (LP), genetic algorithms (GAs), simulated Annealing (SA), particle swarm optimisation (PSO), reinforcement learning (RL), Bayesian optimisation—optimisation AI aims to provide the best possible solutions to complex problems, ensuring effective responses in crisis situations.

[80] Humanitarian OpenStreetMap Team (HOTOSM) (n.d.). *FAIR-DEV: Data for Equitable and Inclusive Development*. [online] Available at: https://fair-dev.hotosm.org/ [Accessed on 24 May 2024]
[81] Translators Without Borders (n.d.). *Chatbot Release in Northeast Nigeria*. [online] Available at: https://translatorswithoutborders.org/chatbot-release-northeast-nigeria/ [Accessed: 24 May 2024]

Use Cases	Brief Description with Examples
Resource Allocation	AI-driven models optimise the distribution of essential resources such as food, shelter, and medical supplies based on real-time data and predictive analytics. Example: - **Hunger Map LIVE**[82] is developed by the World Food Programme which leverages AI, machine learning, and data analytics to track and predict the severity of hunger in near real-time across over 90 countries. By integrating various data streams—such as food security, weather, population, conflict, and economic indicators—HungerMap provides a comprehensive, up-to-date overview of food insecurity, helping WFP and other humanitarian organisations optimise the allocation of resources, improve response times, and make more informed decisions to combat hunger efficiently.
Supply Chain Management / Transportation and Logistics	AI optimises supply chain logistics to ensure timely and efficient delivery of aid, potentially reducing waste and improving responsiveness. Examples: - **Zipline**[83] is a pioneering company that utilises drone technology and optimization AI to optimise routes and deliver critical medical supplies, such as blood products and vaccines, to remote and hard-to-reach areas. One country that has tested implementation of optimization AI in its vaccine supply chain management is Rwanda[84]. - **WeRobotics**[85] is a social impact organisation that, in 2019, in a cross-sector consortium led by Dominican Republic Flying Labs tested localised approaches to medical drone delivery in the mountainous inland areas of the Dominican Republic. The aim was to enable clinicians and logisticians in the country's health care system to operate UAV delivery systems without requiring expert or vendor- maintained operations, and to understand how such systems might be operated routinely in the future. More than 100 flights over a

[82] World Food Programme (WFP) (n.d.). *HungerMap*. [online] Available at: https://hungermap.wfp.org/ [Accessed on 24 May 2024]
[83] Zipline (n.d.). *Fly Zipline*. [online] Available at: https://www.flyzipline.com/ [Accessed on 24 May 2034]
[84] Addy A., 2023. *Artificial Intelligence in the Supply Chain Management for Vaccine Distribution in the West African Healthcare Sector with a focus on Ghana.* [online] Available at: https://www.researchgate.net/publication/375603718_Artificial_Intelligence_in_the_Supply_Chain_Management_for_Vaccine_Distribution_in_the_West_African_Healthcare_Sector_with_a_focus_on_Ghana [Accessed on 24 May 2024]
[85] WeRobotics (n.d.). *Watch These Cargo Drones Bring Essential Health Services to Remote Communities.* [online] Available at: https://werobotics.org/blog/watch-these-cargo-drones-bring-essential-health-services-to-remote-communities/ [Accessed on 24 May 2024]

	six-week period transported a variety of medicines and lab samples, with clinicians as the primary operators. The key innovation was embedding an autonomous delivery model within the country's existing health care system, operated and supported by local actors as part of their everyday activities.

Table 2.5. Classification of Facial Recognition AI Use in the Humanitarian Sector

5. Facial Recognition AI (e.g. deep learning, etc.): is one of the prominent–and the most controversial– applications of AI, utilising various techniques and algorithms such as feature-based and holistic methods, eigenfaces, local binary patterns, and deep learning. These techniques are employed to analyse and compare facial features for identification and verification purposes.

Use Cases	Brief Description with Examples
Biometric identification	Facial recognition AI is used for biometric identification, allowing for verification of individuals based on their facial features. Biometric systems are categorised into two primary categories: ***foundational systems*** which supply general identification for many official uses (e.g. national IDs); and ***functional systems*** which are introduced in response to a demand for a particular service or transaction (e.g. voter IDs, health records, etc.)[86]. Example: - **SCOPE**[87] is World Food Programme's web-based platform that acts as a central repository for WFP beneficiary data, enabling the registration of a unique identity for a beneficiary and then operating to enable staff to authenticate beneficiaires in a one- to-many comparison to biometric profiles stored in a centralised database. - **Biometric Identity Management System (BIMS)**[88] is UNHCR's biometric system that enhances refugee assistance by securely managing biometric data for identification and verification purposes.To biometrically enrol refugees, BIMS captures both fingerprint and iris images as well as a facial photograph. This data is stored in a centralised, consolidated database, ensuring a globally unique identity for every refugee.
Severe Malnutrition Detection	Facial recognition AI is applied in the detection of severe malnutrition, assisting healthcare professionals in identifying and treating individuals at risk. Example: - **Methods for Extremely Rapid Observation of Nutritional Status (MERON)**[89] is a facial recognition tool to detect malnutrition using photographs. It allows for a non-invasive, time efficient, and tamper-proof approach to assessing the malnutrition status of an individual by using a facial recognition and processing algorithm.

[86] Gelb A., Clark J., 2013. *Identification for Development: The Biometric Revolution*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2226594 [Accessed on 24 May 2024]
[87] World Food Programme (WFP) (n.d.). *SCOPE Cash Accounts User Manual*. [online] Available at: https://usermanual.scope.wfp.org/cash-accounts/content/common_topics/introduction/1_introduction.htm & https://scope.wfp.org/login/?next=/ [Accessed on 24 May 2024]
[88] UNHCR (n.d.). *Biometric Identity Management System*. [online] Available at: https://www.unhcr.org/media/biometric-identity-management-system & https://www.unhcr.org/blogs/unhcrs-biometric-tools-in-2023/ [Accessed on 24 May 2024]

The classification presented here serves as a preliminary framework for understanding the diverse applications of AI in humanitarian contexts. However, it also underscores the need for further research in this area, particularly in evaluating the independent evidence of whether AI technologies and examples are genuinely supporting those most in need. While the examples provided are descriptive and showcase the use of various AI types in humanitarian cases, it is essential to recognize that they represent only a fraction of how the humanitarian field is adopting AI technologies. Due to constraints of length, a comprehensive analytical evaluation of each example is beyond the scope of this classification in particular, and this research report in general. Instead, they serve as illustrations of the evolving landscape where AI is being leveraged to address humanitarian challenges.

While AI represents the new neophilia in the humanitarian field, it doesn't come without its risks and contradictions. Next, we will explore the risks, challenges, and ethical dilemmas associated with the adoption of AI in humanitarian contexts, shedding light on the complexities and considerations necessary for the responsible deployment of these technologies in aid efforts.

2.4.2 | Risks and harms of AI in humanitarian contexts

As highlighted in the previous section, the adoption of AI in humanitarian contexts is substantial. International organisations, NGOs, activists, and scholars have produced numerous publications addressing the risks and ethics of AI in the humanitarian sector. Additionally, there exists a significant body of empirical research shedding light on the adverse impacts of AI systems, particularly on groups with protected characteristics (McQuillan, 2022; Benjamin, 2019; Crawford, 2021; Katz, 2020; Lambrecht & Tucker 2018; Eubanks 2018).

AI systems represent both powerfulness and brittleness. They are powerful in analyzing large amounts of data and drawing correlations, as long as they stay within the range of data parameters they have been trained on. However, they perform much worse when applied beyond their training distribution (Bengio et al. 2019). For instance, there have been cases where self-driving car has collided with a tow truck at the side of the road because its training data didn't include a statistically significant representation of tow trucks (Charrington 2019, as cited in McQuillan 2022). Deep learning-based algorithms hardly capture the effective potential for correlation between phenomena, and often struggle to attribute causation (Knight 2019). This leads, for example, to misclassification of objects within an image (e.g., misdiagnosis or

[89] American Institutes for Research (AIR) (n.d.). *Methods for Extremely Rapid Observation of Nutritional Status (MERON)*. [online] Available at: https://www.air.org/project/methods-extremely-rapid-observation-nutritional-status-meron [Accessed on 24 May 2024]

mistargeting) or misidentification of individuals that could lead to more severe (legal or humane) consequences (Coppi, et al. 2021). As Mcquillan (2022) argues, even when neural networks seem to understand the world around them, it turns out to be a show understanding, wherein 'to these systems, an image of people escaping a flood may look like "people on a beach" and a crashing airplane like "airplane on a tarmac"' (Katz 2020, as cited in McQuillan 2022).

This inherent limitation of AI systems can (re)produce a number of risks and harms that may impact the lives of the most vulnerable people. Some of the main categories of risks and harms of AI systems in the humanitarian aid can be summarised as follows:

- **Data Privacy:** The collection, storage, and analysis of vast amounts of data – including biometric data – in humanitarian operations raise significant concerns regarding data privacy[90]. The misuse or unauthorised access to sensitive information can jeopardise the security and confidentiality of individuals affected by crises. As Bridges (2017) notes, poor mothers and marginalised populations have a "weak version" of privacy rights. Interestingly, most sweeping digital decision-making tools are tested in what could be called "low rights environments" where there are few expectations of political accountability and transparency; systems first designed for the poor will eventually be used on everyone (Eubanks: 2018). UN's 2020 report warned about the use of 'smart border' technologies which can be unfair and regularly breach human rights[91], a warning that was echoed also by a joint statement of civil society organisations in relation to EU AI Act[92]. In 2022, the International Committee of Red Cross whose servers hosting personal data belonging to more than 515,000 people worldwide were hacked in a sophisticated cyber attack[93]. When it comes to the consent of people to share their own personal data, Wright and Verity (2020) argue that 'data consent' within the humanitarian context is meaningless because 'when people need humanitarian aid, they are almost certainly going to give their consent over the use of their data, [thus] making consent a meaningless concept' (Wright, Verity: 2020, pg 28). The question of data privacy becomes a fundamental right to be protected and safeguarded— particularly in humanitarian contexts where stakes are high and decisions urgent—but equally challenging with fast-paced evolving and opaque AI systems whose appetite for data is insatiable.
- **Biases:** AI systems trained on biassed datasets can perpetuate and exacerbate existing inequalities, particularly among marginalised communities. For instance, in the healthcare sector, ML-based systems have been praised for identifying early signs of various cancers based on radiographic screening. However, a study published in May 2024 has evidenced how an FDA-approved AI algorithm was more likely to wrongly indicate the presence of cancer in Black women compared to white, Hispanic and Asian women[94]. Similarly, another study published in October 2023 has shown how Large Language Models propagate race-based medicine[95]. When it comes to facial recognition and gender, Scheuerman et al (2019) have empirically illustrated how these systems are non-binary and transgender bias[96]. As we have seen above, generative AI, chatbots and facial recognition technologies are being deployed in humanitarian contexts. The problem of quality of datasets, used in training, which often may include biassed and contaminated data is accompanied by another—equally important—problem, that is, just like most innovations, even chatbots have western bias in their design (Madianou 2021) reproducing colonial matrix of domination and exploitation[97].
- **Opaqueness:** With the term 'opaqueness' we refer to the lack of transparency or clarity in AI algorithms and decision-making processes. This is sometimes referred to as 'black box'. When AI systems operate in an opaque manner, it means that their inner workings are not easily understandable or accessible to human observers. Burrell identifies "three distinct forms of opacity include: (1) opacity as intentional corporate or institutional self-protection and concealment and, along with it, the possibility for knowing deception; (2) opacity stemming from the current state of affairs where writing (and reading) code is a specialist skill and; (3) an opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation" (Burrell 2016, as cited in Coppi et al. 2021). This is particularly the case regarding the use of the proxy means test (PMT), which selects recipients through a mathematical formula, based on a survey of household assets. The PMT methodology is now widely used in both humanitarian aid and social protection programmes, despite well-known issues. There are already good examples of the PMT failing in a humanitarian context. In a voucher programme in Burkina Faso, an evaluation showed that the PMT used was no more effective than a random distribution in identifying the most food insecure households[98]. Qualitative research into the perceptions of refugees of the PMT targeting mechanism used in Lebanon showed that refugees do not understand how beneficiaries were selected. Even the field staff administering the assistance in Lebanon would sometimes tell people that the programme uses random selection, or would not know what the selection criteria are, adding to the frustration of refugees[99].
- **Exploitation:** AI systems deployed in humanitarian contexts may inadvertently contribute to the exploitation of vulnerable populations. Whether through the manipulation of personal data for

[90] Welsh T., 2019. *Biometrics disagreement leads to food aid suspension in Yemen.* [onl;ine] Available at: https://www.devex.com/news/biometrics-disagreement-leads-to-food-aid-suspension-in-yemen-95164 [Accessed on 01 June 20234]

[91] Fallon K., 2020. *UN warns of impact of smart borders on refugees: 'Data collection isn't apolitical'*. [online] Available at: https://www.theguardian.com/global-development/2020/nov/11/un-warns-of-impact-of-smart-borders-on-refugees-data-collection-isnt-apolitical [Accessed on 01 June 2024].

[92] AccessNow, (2024). *Joint statement – A dangerous precedent: how the EU AI Act fails migrants and people on the move.* [online] Available at: https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/ [Accessed on 01 June 2024]

[93] ICRC, 2022. *Cyber attack on ICRC: What we know*. [online] Available at: https://www.icrc.org/en/document/cyber-attack-icrc-what-we-know%E2%80%8B [Accessed on 01 June 2024]

[94] Nguyen D., et. al., 2024. *Patient Characteristics Impact Performance of AI Algorithm in Interpreting Negative Screening Digital Breast Tomosynthesis Studies*. [online] Available at: https://pubs.rsna.org/doi/10.1148/radiol.232286 [Accessed on 01 June 2024]

[95] Omiye I. J., et. al., 2023. *Large language models propagate race based medicine*. [online] Available at: https://www.nature.com/articles/s41746-023-00939-z [Accessed on 01 June 2024]

[96] Scheuerman M., et. al., 29019. *How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services*. [online] Available at: https://dl.acm.org/doi/10.1145/3359246 [Accessed on 01 June 2024)

[97] Drahl C., 2023. *AI was asked to create images of Black African docs treating white kids. How'd it go?* [online] Available at: https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it- [Accessed on 01 June 2024]

[98] Development Pathways blog, 2018. *Targeting humanitarian aid: something to be left to opaque algorithms?* [online] https://www.developmentpathways.co.uk/blog/targeting-humanitarian-aid-something-to-be-left-to-opaque-alogorithms/ [Accessed on 01 June 2024]

[99] Idem.

commercial gain or the reinforcement of harmful stereotypes, safeguarding against exploitation is paramount in AI-driven humanitarian initiatives. “Data colonialism” and “digital colonialism” have become popular metaphors for academics, policymakers, and advocacy organisations looking to critique harmful AI practices (Crawford, et al. 2019). Colonialism is commonly invoked to elucidate the extractive and exploitative dynamics characterising the relationship between technology companies and individuals, employed for diverse political objectives. For instance, the amount of human labour invested to develop and train AI technologies is immensely vast. Take, for example, one of the first deep learning dataset known as ImageNet, which consists of more than 14 million labelled images, each of which is tagged, belonging to more than 20,000 categories. This was made possible by the efforts of thousands of anonymous workers who were recruited through Amazon’s Mechanical Turk platform. This platform gave rise to ‘crowdwork’: the practice of dividing large volumes of time-consuming tasks into smaller ones that can be quickly completed by millions of people worldwide. Crowdworkers who made ImageNet possible received payment for each task they finished, which sometimes was as little as a few cents[100]. Beyond the precarious working conditions of the workers and their ill-paid jobs, the worker’s mental health was at stake, suffering from recurring visions after reading a graphic description of a man having sex with a dog in the presence of a young child[101]. Although the scope of exploitation could be further expanded to examine how the AI industry utilizes natural resources, this aspect will be explored in detail in Chapter 4 of this report.

In examining the risks and harms posed by the integration of AI in humanitarian contexts, it becomes evident that fundamental principles of humanitarian action—such as humanity, neutrality, impartiality, and independence—are at risk. The challenges presented by AI systems, ranging from data privacy concerns to biases and opaqueness in decision-making processes, directly impact the ability of humanitarian organisations to uphold these principles. As we transition to the next section, which explores AI governance frameworks, it is imperative to consider how these principles can be safeguarded in the face of evolving technological landscapes and the ethical dilemmas they entail, including the more radical voices advocating for the abolition of AI[102].

SECTION 2.5

AI Governance: Current and Emerging Developments by Governments and International Organizations

The rapid advancement of AI technologies has brought with it new-facing challenges such as infringement of intellectual property, misinformation, privacy violations, bias and discrimination, job displacement,

[100] Newman A., 2019. *I Found Work on an Amazon Website. I Made 97 Cents an Hour.* [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed on 01 June w024]
[101] Perrigo B., 2023. *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. [online] Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed on 01 June 2024]
[102] On this theme, see: Earl C. C., 2021. *Towards an Abolitionist AI: the role of Historically Black Colleges and Universities*. [pdf] Available at: https://arxiv.org/abs/2101.02011 [Accessed on 01 June 2024]; STOP LAPDS Spying Coalition, 2020. *The Algorithmic Ecology: An Abolitionist Tool for Organizing Against Algorithms*. [online] Available at: https://freerads.org/2020/03/02/the-algorithmic-ecology-an-abolitionist-tool-for-organizing-against-algorithms/ [Accessed on 01 June 2024]; *Algorithmic Sabotage Research Group, 2024. Theorizing “Algorithmic Sabotage”*. [pdf] Available at: https://cryptpad.fr/file/# [Accessed on 01 June 2024].

security vulnerabilities, and ethical concerns. This has prompted governments and international organisations to establish frameworks for AI governance. To this date, OECD’s Policy Observatory has listed over 1000 AI policy initiatives from over 69 countries, territories and the EU.

Apart from governments and institutional organisations, scholars have also proposed new ways to regulate AI. For instance, Roio (2018) has pushed forward the thesis of ‘Algorithmic Sovereignty’. In contrast, Christiaens (2024) has called for the nationalisation of AI, which entails 'the establishment of state ownership over AI development without state control.' Floridi and Cowls (2019) proposed a unified framework of five principles for AI in society which includes: *beneficence*, *non-maleficence*, *autonomy*, *justice,* and *explicability.* While these proposals provide fresh perspectives, more 'standard'[103] approaches by governments and international organisations focus on establishing comprehensive regulatory frameworks and ethical guidelines. The following is an overview of the existing and latest developments in AI governance by various governmental bodies and international organisations. Given the extensive landscape of AI governance, it is impossible to cover all initiatives comprehensively. Therefore, we will focus on key countries around the world, providing descriptive information on their significant AI regulations.

2.5.1 | Governmental AI Governance Initiatives

According to Stanford University’s AI Index Report[104], as of now, 75 national AI strategies have been unveiled. The peak year was 2019, with 24 strategies released. In 2023, eight new strategies were introduced by countries in the Middle East, Africa, and the Caribbean, highlighting the global expansion of AI policymaking discourse.

Figure 2.3. © *Countries with a national strategy on AI, 2023.* AI Index, 2024

[103] With respect to 'standards', we should highlight the critical examination by Solow-Niederman, who argues that standards are neither neutral nor objective. Solow-Niederman recognizes that “standards are crafted by actors who make normative choices in particular institutional contexts, subject to political and economic incentives and constraints [...]”. Solow-Niederman A., 2024. *Can AI Standards Have Politics?* [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4714812 [Accessed on 01 June 2024]
[104] Human-Centred Artificial Intelligence, University of Stanford, 2024. AI Index Report 2024. [pdf] Available at: https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf [Accessed on 10 June 2024].

Stanford University's AI Index Report analysed legislation containing "artificial intelligence" in 128 countries from 2016 to 2023, out of which 32 countries have enacted at least one AI-related bill. In total, these countries have passed 148 AI-related bills. The number of AI-related bills passed in 2023 significantly exceeded the total passed in 2016, with 39 bills passed in the previous year.

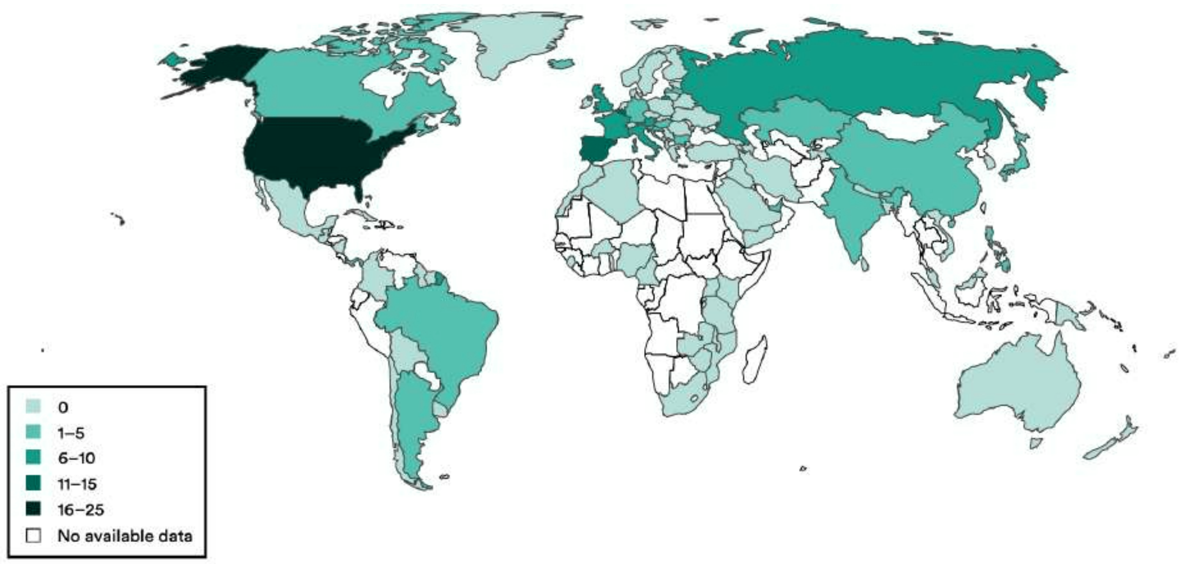

Figure 2.4. © *Number of AI-related bills passed into law by country, 2016-2023.* AI Index, 2024

Another interesting figure from the AI Index Report is the global analysis of AI legslitation by their primary subject matter:

> [...] in 2023 the distribution of primary topics among passed bills broadened significantly, encompassing a diverse range of policy areas. Specifically, two bills were passed in each of the following categories: armed forces and national security; civil rights and liberties, minority issues; commerce; education; labor and employment; science, technology, and communication. This diversity indicates that AI policy concerns are increasingly spanning various sectors (AI Index Report 2024, pg. 381).

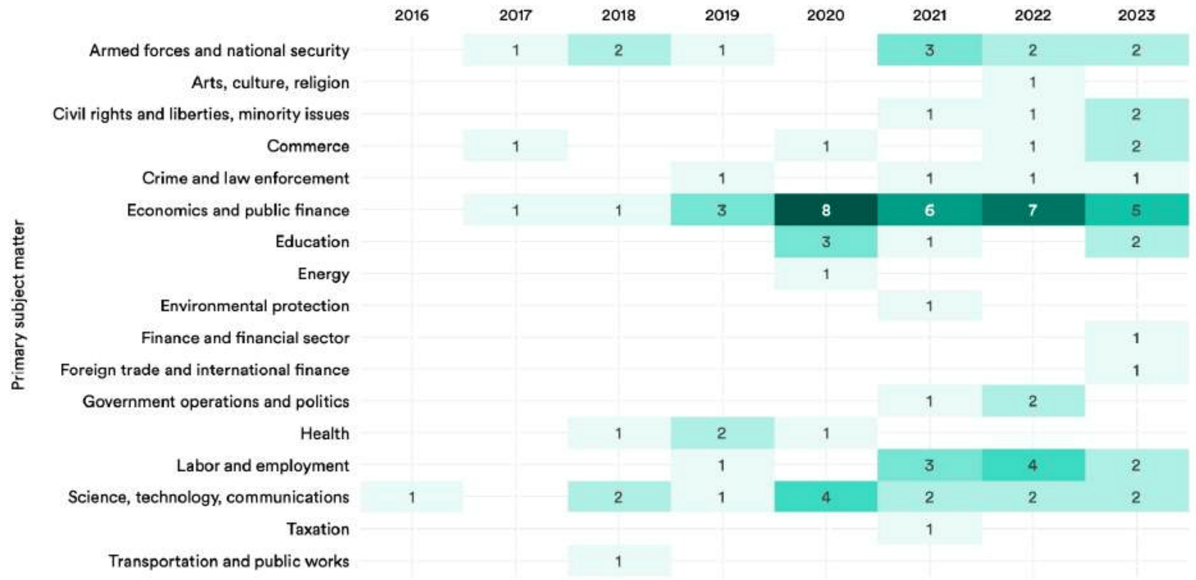

Figure 2.4. © *Number of AI-related bills passed into law in select countries by primary subject matter, 2016-2023.* AI Index, 2024

Given this extensive and evolving landscape of AI legislation, we will now focus on some key governance and regulation acts around the world.

Table 2.6. Key AI Governance and Regulation Acts Around the World

Geographical area	Key AI Governance and Regulation Acts
Europe	While numerous countries across Europe have adopted national strategies on AI, two significant regulatory frameworks stand out, reflecting the efforts of both the European Union (EU) and the Council of Europe (CoE) in AI governance. Both frameworks represent the pioneering efforts of European bodies to establish robust and comprehensive AI governance, setting standards that could influence global AI regulatory practices: ◆ **The European Union artificial Intelligence Act[105]:** Approved on 21 May 2024, the EU AI Act is the first of its kind to regulate general-purpose AI models. It imposes obligations on developers, deployers, and various other participants in the AI value chain (i.e., supply chain) for both high-risk and low-risk AI systems, while also prohibiting certain AI systems. The Act categorises AI systems based on their risk levels—unacceptable risk, high risk, and limited risk—and imposes corresponding regulatory requirements. The high-risk category includes AI systems used in critical

[105] European Union, (n.d.). *Artificial Intelligence Act*. [online] Available at: https://artificialintelligenceact.eu/ [Accessed on 01 June 2024]

	areas such as healthcare, law enforcement, and transportation. ◆ **The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law[106]:** Adopted on 17 May 2024 by the Committee of Ministers of the Council of Europe, this is the first-ever internationally legally binding treaty on AI. Unlike the EU AI Act, which focuses on regulatory obligations within the European Union, the Council of Europe's treaty sets out a legal framework that covers the entire lifecycle of AI systems and addresses the risks they may pose while promoting responsible innovation. This convention, open to non-European countries as well, adopts a risk-based approach to the design, development, use, and decommissioning of AI systems, requiring careful consideration of any potential negative consequences of using AI systems.
United States and Canada	Except for the two framework regulations, the United Kingdom prioritises a flexible framework over comprehensive regulation and emphasises sector-specific laws underpinned by five core principles. These are safety, security and robustness, appropriate transparency and explainability, fairness, accountability and governance, and contestability and redress. However, one of the key documents is AI Regulation White Paper unveiled by the UK Government on August 2023 titled **"A pro innovation approach to AI"[107]** and its written response on February 2024 to the feedback it received as part of its consultation on the White Paper **"A pro innovation approach to AI: a government response"[108]**. Both papers indicate that the UK does not intend to enact horizontal AI regulation in the near future. Instead, the White Paper and the Response support a "principles-based framework" for existing sector-specific regulators to interpret and apply to the development and use of AI within their domains. At present, there is no comprehensive federal legislation or regulation in the US that governs the development of AI or explicitly prohibits or restricts its use. However, AI regulation follows a decentralised approach in the US, resulting in a patchwork of regulations across various sectors, some of which are summarised as follows: ◆ **National AI Initiative Act of 2020[109]:** This Act established a coordinated program across the federal government to accelerate AI research and development. It focuses on expanding AI research and development and created the National Artificial Intelligence Initiative Office and promoting public-private partnerships. ◆ **The White House Executive Order on AI "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"[110]:** It was signed in

[106] Council of Europe, (n.d.). *The Framework Convention on Artificial Intelligence*. [online] Available at: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence [Accessed: 01 June 2024]
[107] UK Government, (2023). *AI Regulation: A Pro-Innovation Approach – White Paper*. [online] Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [Accessed on 01 June 2024]
[108] UK Government, (2023). *A Pro-Innovation Approach to AI Regulation: Government Response*. [online] Available at: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response [Accessed on 01 June 2024]
[109] United States Congress, (2020). *Congressional Report on AI Regulation*. [pdf] Available at: https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210 [Accessed on 01 June 2024]
[110] White House, (2023). *Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*. [online] Available at: https://www.whitehouse.gov/briefing-room/statements-

	October 2023, and focuses on federal agencies and developers of foundation models, mandates the development of federal standards, and requires developers of the most powerful AI systems to share safety tests results and other critical information with the U.S. government. The Executive Order also calls on the Department of Commerce to issue guidance for content authentication and watermarking to label AI-generated content. ◆ **The White House Blueprint for an AI Bill of Rights[111]:** Asserts guidance around equitable access and use of AI systems, and provides five principles and associated practices to help guide the design, use and deployment of "automated systems" including safe and effective systems; algorithmic discrimination and protection; data privacy; notice and explanation; and human alternatives, consideration and fallbacks. Canada is expected to regulate AI at the federal level, through the **Artificial Intelligence and Data Act (AIDA)**, which forms part of Bill C-27, a bill that also includes the Consumer Privacy Protection Act (an update to the current federal privacy law) and the Personal Information and Data Protection Tribunal Act.
Asia	China: ◆ **New Generation AI Development Plan**[112]: Announced in 2017, this plan outlines China's strategy to become a global leader in AI by 2030. It includes measures for promoting AI research, fostering industry development, and addressing ethical and legal challenges. ◆ **Beijing AI Principles[113]:** Were released on May 28, 2019 by the Beijing Academy of Artificial Intelligence (BAAI), an outline to guide the research and development, implementation, and governance of AI. India: ◆ **National AI Strategy[114]:** India's National Institution for Transforming India released a national AI strategy in 2018, focusing on leveraging AI for inclusive growth. The strategy emphasises AI applications in agriculture, healthcare, education, smart cities, and infrastructure. ◆ **The Principles for Responsible AI[115]:** Published in February 2021, the Principles serve as India's roadmap for the creation of an ethical and responsible AI ecosystem across sectors.

releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [Accessed on 01 June 2024]
[111] White House ,(n.d.). *AI Bill of Rights*. [online] Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [Accessed on 01 June 2024]
[112] Stanford University (2017). *Full Translation: China's New Generation Artificial Intelligence Development Plan (2017)*. [online] Available at: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ [Accessed on 01 June 2024]
[113] Beijing Academy of Artificial Intelligence, (2019). *Beijing AI Principles*. [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s11623-019-1183-6.pdf [Accessed on 01 June 2024]
[114] NITI Aayog (2023). *National Strategy for Artificial Intelligence*. [pdf] Available at: https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf [Accessed on 01 June 2024]
[115] NITI Aayog, (2021). *Responsible AI*. [pdf] Available at: https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf [Accessed on 01 June 2024]

	Japan: ◆ **AI Regulatory Policy**[116] **:** the Ministry of Economy, Trade and Industry (METI) published Japan's AI regulatory policy in January 2022. This policy is not legally binding but is expected to support and induce voluntary efforts in good practices by companies developing and deploying AI systems.
Africa	The regulation of AI in Africa is gradually taking shape. Although no African country has enacted specific AI legislation, various governments and institutions are making strides. These efforts are evident in developing national AI strategies and policies, often standalone or integrated into broader data or digital strategies (Tsebee, Oloyede: 2024, pg. 11). Out of 55 African countries, five have adopted a specific National AI Strategy: **Algeria**[117], **Benin**[118], **Ghana**[119], **Mauritius**[120], and **Senegal**[121]. Additionally, there are also countries working on developing AI strategies. For example, in **Nigeria**, the Minister of Communication and Digital Economy has disclosed efforts to develop a National AI Strategy[122]. The African Union—made up of 55 member nations—is preparing an ambitious AI policy that envisions an Africa-centric path for the development and regulation of this emerging technology. On February 29, the African Union Development Agency published a policy draft that lays out a **blueprint of AI regulations for African nations**[123].
South America	Brazil: ◆ **The Brazilian National Strategy for Artificial Intelligence (EBIA)**[124]**:** launched in April 2021, which leverages the country's strengths in technology, innovation, and entrepreneurship to address Brazil's

[116] Ministry of Economy, Trade and Industry (METI), (2021). *AI Implementation in Society*. [pdf] Available at: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf [Accessed on 01 June 2024]
[117] Algeria Press Service (APS), (n.d.). *National Artificial Intelligence Strategy 2020–2030 Presented*. [online] Available at: https://www.aps.dz/en/health-science-technology/37591-higher-education-national-artificial-intelligence-strategy-2020-2030-presented [Accessed on 01 June 2024]
[118] Benin Ministry of Digitalisation, (2024). *Benin National Artificial Intelligence and Big Data Strategy*. [pdf] Available at: https://www.d4daccess.eu/sites/default/files/knowledge_base_products_files/national-artificial-intelligence-and-big-data-strategy-1682673348%20%281%29.pdf [Accessed on 01 June 2024]
[119] Ministry of Finance, Ghana, (2024). *Ghana Charts Collaborative Path to Catalyse AI for Africa's Development*. [online] Available at: https://mofep.gov.gh/news-and-events/2024-03-19/ghana-charts-collaborative-path-to-catalyse-ai-for-africas-development [Accessed on 01 June 2024]
[120] Central Information Board (CIB), (2024). *Mauritius AI Strategy*. [pdf] Available at: https://cib.govmu.org/Documents/Strategies/Mauritius%20AI%20Strategy%20(7).pdf [Accessed on 01 June 2024]
[121] Ministère de la Communication, des Télécommunications et de l'Économie Numérique (2023). *Stratégie Nationale et Feuille de Route du Sénégal sur l'Intelligence Artificielle (Version Résumée)*. [pdf] Available at: https://drive.google.com/file/d/1kXmzNmOayAHo4ESE2rGEh2AaUTZMA-u2/view [Accessed on 01 June 2024]
[122] Fakiya V., 2023. *Nigeria to create a National AI Strategy*. [online] Available at: https://techpoint.africa/2023/08/29/techpoint-digest-658/ [Accessed on 01 June 2024]
[123] African Union Development Agency, (2023). *Regulation and Responsible Adoption of AI for Africa Towards Achievement of AU Agenda 2063*. [pdf] Available at: https://onedrive.live.com/?authkey=%21AKJcwcnXeRGANKQ&id=14DDAD979C3656DF%2145404&cid=14DDAD979C3656DF [Accessed on 01 June 2024]
[124] OECD, (n.d.). *Brazilian AI National Strategy*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-27104 [Accessed on 01 June 2024]

	challenges and create new opportunities for citizens in the innovation ecosystem. The EBIA is based on the five principles defined by the OECD AI Principles, namely: (i) inclusive growth, sustainable development, and well-being; (ii) values centred on human beings and equity; (iii) transparency and explainability; (iv) robustness, security and protection; and (v) accountability. Argentina: ◆ **AI National Plan**[125]**:** The objective of Argentina's AI National Plan is to develop policies that contribute to sustainable growth and the improvement of equal opportunities through AI technologies, ultimately positioning the country as a regional AI leader. Chile: ◆ **AI National Policy**[126]: In October 2021, Chile presented its first National Policy on Artificial Intelligence and the Action Plan that will facilitate its adoption. The Action Plan aims to promote the use and development of AI and brings together 70 priority actions and 185 initiatives that will have an impact on the social and economic fields and nurturing talent. It includes a US$26 billion public investment.
Australia	Australia has not yet enacted any specific statutes or regulations that directly regulate AI. To date, Australia's response to AI includes the **AI Ethics Principles**[127] published in 2019, as well as **AI Action Plan**[128] which set out a vision for Australia to be a global leader in developing and adopting trusted, secure and responsible AI.

2.5.2 | Global and Multilateral AI Governance Efforts

The role of international organisations in shaping AI governance becomes increasingly crucial, particularly in establishing global standards, promoting ethical practices, and facilitating international cooperation to address the complex challenges posed by AI technologies. Some of the key initiatives, frameworks, strategies and significant contributions adopted by these organisations are outlined as follows.

1. **United Nations (UN)**
   - ***Principles for the Ethical Use of Artificial Intelligence in the United Nations System***[129]*:* Were developed through the High-level Committee on Programmes (HLCP) which approved the *Principles* at an intersessional meeting in July 2022. This set of ten

[125] OECD, (n.d.). *AI National Plan of Argentina*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26935 [Accessed on 01 June 2024]
[126] OECD, (n.d.). *AI National Policy, Chile*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24840 [Accessed on 01 June 2024]
[127] Department of Industry, Science and Resources, (n.d.). *Australia's AI Ethics Principles*. [online] Available at: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles [Accessed on 01 June 2024]
[128] OECD, (n.d.). *Australia's AI Action Plan*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26948 [Accessed on 01 June 2024]
[129] United Nations, (2022). *Principles for the Ethical Use of AI in the UN System*. [pdf] Available at: https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf [Accessed on 04 June 2024]

principles, grounded in ethics and human rights, aims to guide the use of artificial intelligence (AI) across all stages of an AI system lifecycle across United Nations system entities. It is intended to be read with other related policies and international law, and includes the following principles: do no harm; defined purpose, necessity and proportionality; safety and security; fairness and non-discrimination; sustainability; right to privacy, data protection and data governance; human autonomy and oversight; transparency and explainability; responsibility and accountability; and inclusion and participation.

- Apart from the Principles, various other UN agencies have developed numerous guidelines, principles, toolkits, and frameworks that contextualise AI technologies for UN agency-specific contexts. Some examples include: ***UNESCO***'s efforts in producing numerous publications[130] on AI and education, including the "Beijing Consensus on Artificial Intelligence in Education"[131]; ***UNICEF***'s publication "Policy guidance on AI for children"[132]; and ***UNDP***'s DEEP (Demystify and Democratise; Empower people; Explore and Experiment; Protect people) framework[133], which champions AI to accelerate progress towards sustainable development, whilst steadfastly promoting human rights.

2. **Organisation for Economic Co-operation and Development (OECD)**
   - ***Principles on AI[134]:*** Adopted in 2019 and updated in May 2024, these principles guide AI actors in their efforts to develop trustworthy AI and provide policymakers with recommendations for effective AI policies. Countries use the OECD AI Principles and related tools to shape policies and create AI risk frameworks, building a foundation for global interoperability between jurisdictions.
3. **World Economic Forum (WEF)**
   - ***AI Governance Alliance[135]:*** Is an initiative by the World Economic Forum's Centre for the Fourth Industrial Revolution, brings together top leaders from industry, government, academia, and civil society. With its 360 members and 250 organisations, The World Economic Forum's AI Governance Alliance aims to create policies, frameworks, and best practices that ensure AI technologies are developed and deployed ethically and responsibly.
4. **G20**
   - ***AI Principles[136]:*** In 2019, the G20 adopted human-centred AI principles based on the OECD guidelines. These principles emphasise inclusivity, transparency, and accountability in AI development and deployment.

[130] UNESCO, (n.d.). *Digital Education and Artificial Intelligence*. [online] Available at: https://www.unesco.org/en/digital-education/artificial-intelligence [Accessed on 04 June 2024]
[131] UNESCO, (2019). *Beijing Consensus on Artificial Intelligence in Education*. [pdf] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000368303 [Accessed on 04 June 2024]
[132] UNICEF, (2021). *Policy Guidance on AI for Children*. [pdf] Available at: https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf [Accessed on 04 June 2024].
[133] United Nations Development Programme (UNDP) (n.d.). *Artificial Intelligence at UNDP*. [online] Available at: https://www.undp.org/digital/ai [Accessed on 04 June 2024]
[134] OECD, (n.d.). *OECD AI Principles*. [online] Available at: https://oecd.ai/en/ai-principles [Accessed on 04 June 2024]
[135] World Economic Forum, (n.d.). *AI Governance Alliance*. [online] Available at: https://initiatives.weforum.org/ai-governance-alliance/home [Accessed on 04 June 2024]
[136] G20, (2019). *AI Principles*. [pdf] Available at: https://www.mofa.go.jp/mofaj/files/000486596.pdf [accessed on 04 June 2024]

5. **G7**
   - ***The Hiroshima AI Process[137]:*** Launched in 2023, the G7's Hiroshima AI Process aims to develop and promote common principles for trustworthy AI. This initiative focuses on enhancing international collaboration, addressing ethical considerations, and fostering innovation while ensuring AI's alignment with democratic values and human rights.

The landscape of AI governance is evolving rapidly as governments and international organisations strive to create frameworks that foster innovation while addressing ethical, legal, and social challenges. One thing that is clear from these efforts is a shared aim to promote human rights, democracy, and the development of trustworthy and ethical AI. However, whether these goals remain merely aspirational in documents or translate into real-world action remains to be seen. Civil society organisations play a fundamental role in evidencing human rights violations and voicing their concerns, thus keeping decision-makers and policymakers accountable. Although not included here, numerous civil society organisations at the international, national, regional, and local levels have produced a large body of research and publications. Continued collaboration, transparent decision-making, adoption of new forms of democratic deliberations, and adaptive regulatory approaches will be crucial in ensuring that AI technologies are developed and deployed responsibly, benefiting societies worldwide.

SECTION 2.6

Concluding Reflections

In this chapter, we have explored the intricate landscape of AI within the sphere of international development and humanitarianism. Beginning with the evolution of AI from the Digital for Development paradigm, we traced its emergence alongside Web 3.0 and other frontier technologies. This progression highlights how AI has become intertwined with modern development strategies, often touted as a panacea for complex global challenges.

The discourse on "AI for Good" revealed a spectrum of perspectives. On one hand, AI holds the promise of addressing critical issues such as healthcare, disaster response, the climate crisis, etc.. On the other, there is a legitimate concern that the term may sometimes serve as a veneer for brand enhancement and market expansion rather than genuine social impact. The critical examination of what constitutes "good" in this context underscores the necessity for transparency and ethical considerations in AI initiatives.

Our assessment of AI's role in the Sustainable Development Agenda, particularly through the lens of the AI4SDG think tank, illustrated the disparities in AI applications across different SDGs. While there is a concentration of projects targeting certain goals like Good Health and Well-being (SDG 3), others remain underrepresented, pointing to an imbalance that could exacerbate existing inequalities.

In humanitarian contexts, our development of a preliminary taxonomy of AI applications serves as a starting point for understanding the diverse ways AI is being leveraged to address complex challenges. AI's application in anticipatory action through predictive analytics showcases its capacity to enhance

[137] Ministry of Foreign Affairs of Japan, (2024). Hiroshima Process: International Guiding Principles for Organizations Developing Advanced AI Systems. [pdf] Available at: https://www.mofa.go.jp/files/100573471.pdf [Accessed on 04 June 2024]

preparedness and response. However, this also brings forth ethical implications, especially concerning data privacy, biases, and the potential for unintended harm. The deployment of AI in these sensitive settings necessitates rigorous ethical standards and a commitment to safeguarding the rights and dignity of affected populations.

Finally, the overview of AI governance frameworks developed by governments and international organizations highlights a global recognition of the need for regulation and ethical guidelines. Despite these efforts, the rapid pace of AI advancement often outstrips policy development, creating a gap that must be addressed through collaborative and adaptive governance models.

Reflecting on the integration of AI in development, it is evident that AI is not a silver bullet for the complex socio-economic and political challenges faced globally. All stakeholders must prioritise inclusive approaches that consider local contexts and needs, and are grounded in communities. This includes addressing the digital divide, ensuring equitable access to AI technologies, and involving communities in the different stages of AI lifecycle. Ethical considerations should be at the forefront, with robust frameworks to mitigate risks such as data privacy breaches, biases, and other unintended negative consequences.

CHAPTER 3

The Use of AI in SDGs 4 (Quality Education), 10 (Reduced Inequality), and 13 (Climate Action)

This chapter explores the use of Artificial Intelligence (AI) in addressing three critical Sustainable Development Goals (SDGs): SDG 4 (Quality Education), SDG 10 (Reduced Inequality), and SDG 13 (Climate Action). By examining the main debates and trends in these fields, as well as examples and real-life cases, this chapter aims to provide an overview of how AI technologies are being leveraged to address these goals and highlight some of the main critical areas concerning the ethics of AI.

SECTION 3.1

AI and SDG 4 (Quality Education)

The integration of technology into education is not new, nor is the concept of "personalization," which has become increasingly prevalent with the emergence of AI as an ‘inevitable technology’. The history of AI in education (AIED) and the idea of personalization can be traced back to the early 20th century, when psychologist Sydney Pressey introduced the first multiple-choice “teaching machine”—an early precursor to modern AIED—at the 1924 American Psychological Association (APA) meeting.

Pressery’s ‘automatic teacher’ would display a question followed by four possible answers. As Waters (2021: 38) notes:

> The test was fed into the machine on a sheet of paper just as one would load a piece of paper into a typewriter. The test taker had four keys with which to respond, and after selecting her answer, the machine would advance automatically to the next question, calculating the number of correct responses along the way. Alternatively, a lever in the back could change its operation slightly, and the machine would not move on to the next question until the test taker got it right, tabulating the number of tries on each question.

Pressey’s teaching machine emerged during the Progressive Era, a time when behaviourism—a branch of psychology that focuses on how people learn through their interactions with the environment—was consolidating itself as a “whole field of human adjustments” (Watson, 124, 11). This period also marked the rise of psychometrics and psychological testing, which “had become one of the predominant means by which psychology sought to ‘reduce education to a science,’ making its system more scrutable and its outcomes more quantifiable” (Watters, 2021, p. 36). Influenced by behaviourism, particularly Edward Thorndike’s "law of effect," Pressey's invention marked the first wave of behaviorally-driven educational technology.

Pressey's machine was based on the psychological principle of reinforcement, a core concept in behaviourism. Positive reinforcement, such as receiving a reward for a correct answer, was designed to encourage repeated behaviours, while negative reinforcement, like being unable to proceed until the right answer was given, aimed to correct errors and shape the "right kind" of learning behaviour. This principle is still fundamental in modern educational technologies.

One can observe a similar approach in today's AI-based apps. For instance, Duolingo, an AI-powered language learning app, structures its material into small, bite-sized lessons, allowing learners to progress to the next question only after providing the correct answer. Notably, Pressey added an optional feature to his machine that dispensed candy whenever a student answered correctly (Watters, 2021, p. 46), much like contemporary AIED and EdTech platforms that reinforce "the right kind of learning behaviour" through gamification and instant gratification, such as Duolingo's *streak rewards* or Kahoot's *point system* that motivates learners with immediate feedback and rewards.

B.F. Skinner, one of the leading figures in behaviourist psychology, was deeply invested in the idea of applying his theories of learning to education. In the 1950s, Skinner developed a "Teaching Machine" based on his principles of operant conditioning. This device allowed students to work through instructional material at their own pace, receiving immediate feedback on their responses. Correct answers were rewarded by advancing to the next problem, reinforcing the "right" behaviour, while incorrect answers were corrected immediately, aligning with Skinner's belief in the power of reinforcement to shape learning.

Skinner was optimistic that his machine could revolutionise education by making learning more efficient and individualised. He believed it would not only enhance student performance but also alleviate the burden on teachers by automating repetitive tasks like grading and instruction. However, despite Skinner's efforts to marketize the Teaching Machine, his vision faced several significant obstacles. As Watters (2021) points out, Skinner's machine ultimately failed to gain widespread adoption for a variety of reasons, including the rise of competing educational technologies such as computers in the 1960s and 1970s, which quickly overshadowed Skinner's relatively simplistic mechanical devices. While these developments played a significant role in the machine's decline, further exploration of these factors is beyond the scope of this discussion[138].

Today, AI's promise to revolutionise modern education through personalised learning, intelligent tutoring systems, and learning analytics is driving a surge of interest and investment among both hard policymakers, such as governments, and soft policymakers, including intergovernmental organisations like the the World Bank (WB), Organisation for Economic Co-operation and Development (OECD), United Nations (UN), European Union (EU), International Labour Organisation (ILO), etc..

While the future of AIED is uncertain, unpredictable and unknowable (Selwyn 2022), we shall next explore the use of AI to tackle the SDG (4): Quality Education.

[138] For in-depth exploration of Skinner's "Teaching Machine" failure, please see: Watters, A. (2021) *Teaching Machines: The History of Personalized Learning*. Cambridge, MA: MIT Press.

SECTION 3.1.1

Current Developments of AI in Education

The *Sustainable Development Goals Report 2024*[139] (hereinafter referred to as "the SDG Report") shows that by 2019, only 58% of students worldwide had reached minimum proficiency in reading. The SDG Report, drawing on recent data, also reveals significant declines in maths and reading scores, partly due to the impact of the COVID-19 pandemic. Many countries continue to face challenges such as inadequate educational infrastructure, teacher shortages, and insufficient training. While technology has expanded educational opportunities, it has also widened inequalities, particularly for marginalised and low-income communities.

Global upper secondary school completion rates have slowed, but Eastern and South-Eastern Asia has made remarkable progress. Girls generally have higher completion rates than boys in most regions, though boys still maintain a slight lead in Central and Southern Asia and sub-Saharan Africa. The decline in maths and reading proficiency, exacerbated by the pandemic, remains a major concern, and sustainability and climate change education are still underemphasized in most curricula.

On the other hand, while the SDG Report highlights persistent challenges in achieving SDG 4, emerging technologies, particularly AI, offer new promises to address these issues. The AI for SDG Observatory, explored in Chapter 2, presents a collection of 1,482 projects (as of 30 July 2024: https://www.ai-for-sdgs.academy/observatory#c4_Quality_Education), providing insights into how AI is being leveraged to enhance educational outcomes. This reflects a shift in the education sector. Historically, investors were hesitant to invest in education due to low returns, long investment cycles, fragmented markets, heavy regulation, and public hesitancy toward privatisation; however, with the emergence of EdTech, the sector is now evolving in line with other areas of the digital economy (Komljenovic et al., 2023). This is evidenced in numerous studies showing the sharp increase in newly established Edtech companies, with Venture Capital (VC) investment in Edtech increasing from $500 million in 2010 to more than $20 billion in 2021 (Williamson and Hogan 2020; Komljenovic, Sellar, and Birch 2021; see also: https://www.holoniq.com/notes/global-edtech-venture-capital-report-full-year-2021 as cited in Komljenovic et al., 2023).

AI in Education (AIED) is complex and applied in a variety of ways within educational settings. Therefore, making generalizable claims about its efficacy or safety is difficult (Holmes, 2023). According to Holmes (2023), there are at least twenty different types of AIED applications, each with unique impacts. Consequently, it is crucial to consider each application, or at least each type of application, separately, ensuring clarity about the specific variations of AIED being discussed.

Holmes and Tuomi (2022) provide a comprehensive taxonomy of AIED systems, categorising them into three distinct but overlapping groups: 1) student-focused; 2) teacher-focused; and 3) institution-focused

[139] United Nations, 2024. *The Sustainable Development Goals Report 2024*. [pdf] Available at: https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf [Accessed on 30 July 2024]

AIED. Each category serves different functions and has varying implications in the education sector.

The table below, derived from Holmes and Tuomi (2022), categorises AIED systems based on their focus and current development stage. These applications are marked as either speculative (*), *researched (**), or commercially available (***),* highlighting the varying degrees of implementation and maturity within AIED systems.

Table 3.1. A taxonomy of AIED systems

STUDENT-FOCUSED AIED	
Intelligent Tutoring Systems (ITS)	* * *
AI-assisted Apps (e.g., maths, text-to-speech, language learning)	* * *
AI-assisted Simulations (e.g., games-based learning, VR, AR)	* * *
AI to Support Learners with Disabilities	* * *
Automatic Essay Writing (AEW)	* * *
Chatbots	* * */* *
Automatic Formative Assessment (AFA)	* * */* *
Learning Network Orchestrators	* * */* *
Dialogue-based Tutoring Systems (DBTS)	* * *
Exploratory Learning Environments (ELE)	* *
AI-assisted Lifelong Learning Assistant	*
TEACHER-FOCUSED AIED	
Plagiarism detection	* * *
Smart Curation of Learning Materials	* * *
Classroom Monitoring	* * *
Automatic Summative Assessment	* * */* *
AI Teaching Assistant (including assessment assistant)	* * */*
Classroom Orchestration	* *
INSTITUTION-FOCUSED AIED	
Admissions (e.g., student selection)	* * *
Course-planning, Scheduling, Timetabling	* * *
School Security	* * *
Identifying *Dropouts* and *Students at risk*	* * *
e-Proctoring	* * *

Source: Holmes and Tuomi (2023, p.550).

To highlight the complexity of how these categories manifest in practice, we will now briefly explore each one—student-focused, teacher-focused, and institution-focused AIED—examining both their unique and overlapping applications and implications within the education sector.

Student-focused AIED

Student-focused AIED systems are designed to tailor educational experiences to the individual needs of students, contrasting significantly with traditional educational approaches. These systems use adaptive learning algorithms and intelligent tutoring systems (ITS) to provide personalised learning paths for students. As a result, this category has become the most extensively researched and attracts the majority of funding in the AIED field. For instance, platforms like *Knewton*[140] and *Aleks*[141] utilise AI to assess individual student performance in real-time and adjust the content accordingly, ensuring that each learner receives material suited to their current understanding. This allows for timely, individualised feedback, which enhances student engagement and learning outcomes (Cui et al., 2018).

In recent years, many student-focused AIED applications, especially intelligent tutoring systems, have been adapted to better support learners with disabilities (Holmes & Tuomi, 2022), or to diagnose learning disabilities like ADHD, dyslexia, and dysgraphia, helping to identify students' specific needs early in their educational journey (Barua et al., 2022, as cited in Holmes & Tuomi, 2022). For example, *Lexplore*[142] employs AI and eye-tracking technology to detect reading difficulties such as dyslexia by analysing students' eye movements during reading tasks, enabling early intervention.

Further, ITS are capable of adjusting the level of support and the type of instruction based on the specific requirements of each student, thereby promoting inclusivity in education (Khazanchi & Khazanchi, 2021). Other student-focused AIED includes dialogue-based tutoring systems. An example of such a system is *AutoTutor*[143], which engages students in conversational dialogues and adapts its feedback based on the learner's responses, promoting deeper understanding and personalised support (Graesser et al., 2005).

[140] Wiley, (n.d.). *Alta*. [online] Available at: https://www.wiley.com/en-it/education/alta [Accessed on 04 August 2024]
[141] ALEKS, (n.d.). *ALEKS: Assessment and Learning in Knowledge Spaces*. [online] Available at: https://www.aleks.com/?_s=2894576279172929 [Accessed on 04 August 2024]
[142] Lexplore, (n.d.). *Lexplore Assessment*. [online] Available at: https://lexplore.com/lexplore-assessment [Accessed on 04 August 2024]
[143] University of Memphis, (n.d.). *AutoTutor Project*. [online] Available at: https://www.memphis.edu/iis/projects/autotutor.php [Accessed on 04 August 2024]

AIED systems are increasingly penetrating humanitarian and development contexts too, extending their reach beyond traditional educational environments. For example, the recent partnership between the International Rescue Committee (IRC) and OpenAI aims to leverage AI to scale educational technology in crisis-affected areas by developing "aprendIA," an AI-driven chatbot platform that delivers personalised learning experiences to teachers and parents[144]. We should therefore ask whether deploying AI in these vulnerable contexts might inadvertently exacerbate existing inequalities, compromise privacy rights (see Bridges, 2017), or normalise surveillance under the guise of educational advancement.

Despite the promising potential of student-focused AIED systems, significant concerns have been raised regarding their impact on students' educational experiences and well-being. One major critique centres on issues of data privacy and surveillance. These systems often collect extensive personal data, including learning behaviours, emotional states, and even biometric information, raising questions about data security, consent, and the potential misuse of information (see Williamson, 2017). Moreover, the reliance on algorithms to personalise learning experiences can inadvertently reinforce existing biases, as these algorithms may reflect the prejudices present in their training data (see O'Neil, 2016). Ultimately, the overemphasis on technology might also diminish the vital human elements of teaching and learning—what Biesta (2011) refers to as *the socialisation* function of education—such as empathy, mentorship, and the nuanced understanding that educators bring to the classroom (Knox, 2020). These concerns—briefly summarised here—highlight the need for a critical examination of student-focused AIED to ensure the holistic development of learners and the essential role of teachers aren't hindered.

Teacher-focused AIED

Teacher-focused AIED systems are designed to support educators in various aspects of teaching, ranging from administrative tasks to instructional planning, 'smart' curation of learning materials, classroom monitoring, automatic grading, and feedback, etc.. However, as Holmes and Tuomi (2022) observe, student-focused AIED often overlaps with teacher-focused ones. This is the case, for instance, with intelligent tutoring systems (ITS); many of them incorporate teacher interfaces or dashboards that provide educators with information about individuals and groups of students. One innovative approach utilises augmented reality (AR) glasses worn by teachers to superimpose dashboard-like information above their students' heads as the students engage with an ITS (Holstein et al., 2018, as cited in Holmes & Tuomi, 2022). While this technology offers real-time insights and enhances interactive learning, it also raises significant privacy and ethical concerns, including the possibility that malicious actors could exploit AR systems to monitor student behaviours or manipulate the technology to cause physical harm (see Dwivedi et al., 2022; Nichols, 2022).

One prominent application of teacher-focused AIED is in automated grading and feedback systems. On one hand, there are *Automatic Assessment AI* tools that replace teachers in marking students

[144] International Rescue Committee (IRC), (2024). *OpenAi x International Rescue Committee: Leveraging AI to Scale Ed-Tech in Crisis Affected Settings.* [online] https://www.rescue.org/press-release/openai-x-international-rescue-committee-leveraging-ai-scale-ed-tech-crisis-affected [Accessed on 04 August 2024]

assignments and homework. An example of this is *Gradescope*[145], a commercially available tool that offers online and AI-assisted grading for higher education. Gradescope handles multiple types of assessments—including multiple-choice, short answer, and programming assignments—providing consistent and timely feedback to students (Singh et al., 2017). On the other hand, we are witnessing the emergence of *AI Teaching and Assessment Assistant* tools which support teachers—rather than replacing them—by augmenting teachers' expertise and skills (Holmes & Tuomi, 2022). For example, *Jill Watson*[146], developed at the Georgia Institute of Technology, is an AI teaching assistant that answers routine student inquiries in online forums. This reduces the administrative burden on instructors and allows them to focus on more complex student needs (Goel & Polepeddi, 2016).

Plagiarism detection is another area where teacher-focused AIED are being implemented. AI-driven tools like *Turnitin*[147] employ ML algorithms and natural language processing to compare student submissions against extensive databases of academic work, internet sources, and previously submitted papers. These systems assist educators by automatically identifying potential instances of plagiarism, highlighting matched text, and providing similarity reports.

While teacher-focused AIED systems promise numerous benefits in streamlining administrative tasks and enhancing instructional capabilities, they also raise significant concerns regarding the role of teachers and the nature of their work. Some of the tasks which are supposed to be 'pedagogical enjoyments'—such as syllabus design and lecture planning; interactive-based learning and assistant to students; feedbacking; etc.—may increasingly be delegated to AI systems. The increasing reliance on AI tools for such tasks can contribute to the alienation of teachers, reducing their roles to facilitators of algorithm-driven processes. This shift may lead to the precarization of teaching positions, as educational institutions might prioritise cost-saving automation over investing in human educators. Moreover, as Introna (2016) discusses in the context of plagiarism detection software like Turnitin, algorithmic quantification and prediction practices fundamentally alter how we perceive educational subjects and governance practices. The assemblage of algorithms, institutional expectations, and the valorization of measurable outputs leads to governance techniques that reshape teacher and student subjectivities. Teachers and students may internalise and regulate their practices around these algorithmic systems, potentially undermining their professional autonomy and pedagogical judgement.

Institution-focused AIED

Institution-focused AIED systems are designed to support educational institutions in managing and optimising their operations, policies, and strategic planning. These systems leverage AI to analyse large-scale institutional data, providing insights that can enhance administrative efficiency, resource allocation, course-planning and scheduling, school security, identifying and predicting student dropouts, students admissions, and overall institutional effectiveness. While not as extensively researched as student-

[145] Gradescope, (n.d.). *Gradescope: Simplifying Grading for Educators*. [online] Available at: https://www.gradescope.com/ [Accessed on 04 August 2024].
[146] GVU Center, Georgia Tech, (n.d.). *Virtual Teaching Assistant: Jill Watson*. [online] Available at: https://gvu.gatech.edu/research/projects/virtual-teaching-assistant-jill-watson [Accessed on 04 August 2024].
[147] Turnitin, (n.d.). *Turnitin: Integrity in Every Assignment*. [online] Available at: https://www.turnitin.com/ [Accessed on 04 August 2024]

focused AIED, institution-focused applications are gaining traction due to their potential to drive systemic improvements in education (Holmes & Tuomi, 2022).

One key application of institution-focused AIED is in predictive analytics for student success and retention. Educational institutions employ AI algorithms to analyse data on student performance, attendance, engagement, and other factors to identify students at risk of dropping out or underperforming. Systems like *Civitas Learning*[148] and *IBM Watson Education*[149] uses ML models to predict student outcomes and recommend timely interventions (Sclater et al., 2016).

Institution-focused AIED is being used for admissions and enrollment management, particualrl;y in many higher education institutions, mainly in the US. These AI-assisted admissions software help reduce costs while making the admissions system more equitable, by helping to remove unseen human biases (such as group think and racial and gender biases) that can impact decisions (Holmes & Tuomi, 2022). Some of the commercially available AIED tools that are increasingly being used in education include *Admityogi*[150], *Salesforce*[151], *Kira Talent*[152], etc..

However, these systems do not come without controversy (Pangburn 2019). One major critique centres on issues of data privacy and ethical use of student information. Collecting and analysing vast amounts of personal data can lead to potential misuse or breaches of sensitive information. There is also the risk of algorithmic bias, where AI systems may perpetuate existing inequalities or discriminate against certain groups of students if the training data reflects societal biases (see O'Neil, 2016; Buolamwini & Gebru, 2018; Benjamin 2019). For example, predictive models might inaccurately label students from marginalised backgrounds as high-risk, leading to stigmatisation or reduced opportunities. Additionally, when it comes to the 'efficacy of predictions', evidence suggests that AI models, even with large amounts of data, fail to produce reliable outcomes in complex human contexts. This is exemplified by the *Fragile Families Challenge*, a Princeton University study involving hundreds of AI and ML researchers, data scientists, and statisticians. The goal was to predict six life outcomes, including a child's grade point average, their perseverance in schoolwork, and whether a family would face eviction. Despite access to nearly 13,000 data points on over 4,000 families over a span of fifteen years, none of the teams succeeded in developing even moderately successful statistical models (Salganik et al., 2019).

[148] Civitas Learning, (n.d.). *Civitas Learning: Transforming Higher Education*. [online] Available at: https://www.civitaslearning.com/ [Accessed on 04 August 2024]

[149] IBM, (n.d.). *Watson Education Classroom*. [online] Available at: https://www.ibm.com/mysupport/s/topic/0TO50000000Qei8GAC/watson-education-classroom?language=en_US [Accessed on 04 August 2024]

[150] AdmitYogi, (n.d.). *AdmitYogi: Your Admissions Helper*. [online] Available at: https://admityogi.com [Accessed on 04 August 2024]

[151] Salesforce, (n.d.). *Education Recruitment and Admissions Software*. [online] Available at: https://www.salesforce.com/education/recruitment-admissions-software/ [Accessed on 04 August 2024]

[152] Kira Talent, (n.d.). *Kira Talent: Holistic Admissions Solution*. [online] Available at: https://www.kiratalent.com/ [Accessed on 04 August 2024]

Reliance on AI for institutional decision-making may also diminish the human judgement and contextual understanding that administrators bring to their roles. Important nuances and individual circumstances might be overlooked in favour of data-driven approaches, potentially leading to impersonal or unfair policies (Selwyn, 2016). There is also concern that such systems could contribute to the commodification of education, where decisions are driven more by efficiency metrics than by educational values. where decisions are driven more by efficiency metrics than by educational values. For instance, companies like *Othot* offer predictive models to educational institutions, stating: *"Our models are built from scratch for each customer to identify which students are most likely to enroll and how to influence that likelihood"*[153]. Such practices may encourage institutions to focus on applicants based on predicted enrollment probabilities and potential financial contributions, leading to admissions practices that prioritise institutional revenue over educational equity and diversity. This could potentially undermine principles of equal opportunity and access to education, which are central to SDG 4.

To further examine the practical applications and inherent complexities of AIED systems, we turn to a case study of Third Space Learning (TSL)—a Learning Network Orchestrator—that exemplifies both the potential benefits and the concerns associated with AIED.

[153] Othot, (n.d.) *Enrollment Management*. [onl;ine] Available at: https://www.othot.com/products/enrollment-management [Accessed on 04 August 2024]

3.1.2 | Case Study: Third Space Learning (TSL) - The Case of Learning Network Orchestrator

We have chosen to highlight one specific case of LNOs because they encourage hybrid learning experiences where students interact with human—not chatbot—tutors, and where online learning is blended with physical learning environments, while, at the same time, learning happens within the school environment, as is the case in our case study.

Learning network orchestrators (LNOs) are AI-driven platforms/tools designed to facilitate collaboration between students and educators/tutors, coordinating learning activities across networks. These systems match participants based on factors such as availability, subject domain, and expertise, promoting personalised educational experiences (UNESCO, 2021). By applying a range of AI techniques, such as NLP, ML, and social network analysis, to provide individualised recommendations, LNOs connect learners and teachers with peers and relevant resources, and facilitate self-directed learning, peer-to-peer learning, and collaboration (Holmes, 2023).

One notable example of LNOs is the *Third Space Learning*[154] (TSL), an online tutoring service, for primary and secondary education, that uses AI algorithms to match students with qualified tutors (Holmes, 2023). Tutors, located mainly in India and Sri Lanka, engage with UK-based students through real-time, interactive lessons delivered via the platform. This setup ensures that students receive targeted support in mathematics, tailored to their specific learning gaps and needs. By using data analytics, TSL can track each student's progress, ensuring continuous improvement and personalised learning paths. The platform's use of AI for student-tutor matching—based on factors such as time availability and learning objectives—demonstrates its role as an LNO.

Available only for schools in the UK and the US, TSL offers 22 tutoring programmes across primary and secondary levels. The lessons are created by former teachers and follow a structured pedagogical approach using scaffolded teaching slides. Each lesson follows the 'I do, we do, you do' approach, helping to build conceptual understanding in students. Tutors begin each session with a structured formative assessment to gauge how much support the pupil needs on a given topic, allowing for adaptive learning, where prior knowledge is reinforced, or advanced topics are tackled depending on student needs. This approach aligns well with TSL's goal of ensuring personalised and targeted learning.

[153] Othot, (n.d.) *Enrollment Management*. [onl;ine] Available at: https://www.othot.com/products/enrollment-management [Accessed on 04 August 2024]
[154] Third Space Learning (n.d.). *How It Works*. [online] Available at: https://thirdspacelearning.com/how-it-works/ [Accessed on 04 August 2024]

In order to personalise students' learning, each student is required to complete a diagnostic assessment before they begin the programme, indicating to teachers and tutors which of the lessons each student is secure with, and in which they have learning gaps[155]. What is particularly interesting about TSL is its flexibility in aligning tutoring lessons with the class teaching of a particular school. Furthermore, TSL external tutors receive extensive training on delivering the National Curriculum, with training modules designed by former teachers to ensure consistency and quality in the tutoring provided (TSL, n.d.).

According to an evaluation conducted by National Centre for Social Research (2022)[156], the programme was well-received by both school staff and students, with improvements reported in areas such as verbal reasoning, understanding of mathematical concepts, and student confidence in maths. The one-to-one, personalised nature of the tuition was identified as a key factor in driving these positive outcomes. However, the evaluation also highlighted some barriers to successful implementation, such as technical difficulties, lack of resources, and limited engagement from some school staff. These issues, coupled with disruptions caused by the COVID-19 pandemic, posed challenges to fully realising the programme's potential.

Discussion

While TSL has shown promise in addressing educational inequalities and providing tailored support to students, several key concerns merit critical attention.

One area of concern is the **surveillance** aspect inherent in AI-driven platforms like TSL. The system collects and processes a significant amount of data on students' performance, attendance, and engagement. Although this data is used to tailor instruction and track progress, it also raises issues around student privacy and the potential for excessive monitoring. According to TSL's "Data Protection & Privacy Policy"[157] the company "may also collect publicly available information about you or information about you we may acquire from service providers or educational information providers." This broad scope of data collection extends beyond what is directly shared by users, potentially increasing the risk of normalising surveillance within educational spaces. The continuous tracking of student behaviour through such platforms risks normalising a culture of surveillance within educational spaces, where students' every action is recorded and analysed. This could lead to unintended consequences, such as shaping student behaviour based on the anticipation of being watched, rather than fostering genuine learning motivation.

[155] See here: [online] https://thirdspacelearning.com/tutoring/ [Accessed on 04 August 2024]
[156] National Centre for Social Research, (2022). *Evaluation of Third Space Learning's Affordable Maths Tuition*. [pdf] Available at: https://whatworks-csc.org.uk/research-report/pilot-evaluation-of-affordable-maths-tuition/ Accessed on 04 August 2024]
[157] Third Space Learning, (n/d). *Data Protection & Privacy Policy.* [pdf] Available at: https://thirdspacelearning.com/data-protection-privacy-policy/ [Accessed on 04 August 2024]

Another critical issue is **data privacy**. TSL collects sensitive student data, including academic performance and personal identifiers, which are processed by tutors and AI algorithms. The platform collects a wide range of personal information, not only from students but also from teachers and school staff. For teachers, TSL gathers individual contact information, including names, school addresses, job titles, and even personal mobile numbers (if provided), along with details about their timetables and educational resources accessed via TSL's learning hub. This extensive data collection allows TSL to manage its services effectively but also opens the door to potential risks, such as unauthorised access or data misuse. For students, especially minors, the concerns become more pronounced. TSL collects personal data from children, including their name, age, gender, academic year, and information about any medical or learning needs. While this data is necessary for providing personalised one-on-one tuition, the fact that all lessons are recorded (including on-screen interactions and conversations between students and tutors) introduces significant surveillance concerns. These recordings are stored for safeguarding and performance monitoring but also contribute to a sense of constant surveillance. Despite anonymization efforts in post-session analysis, the continuous collection and storage of sensitive data may expose students to risks related to privacy violations. Additionally, the retention policy for pupil data states that TSL holds onto personal data for up to two academic years after a student ceases to be active. While this policy aligns with operational needs, the relatively long retention period could lead to discomfort among parents and students, especially considering the sensitive nature of the data.

While TSL complies with data protection regulations like the UK's General Data Protection Regulation (GDPR), however, the international data flow introduces complexities. Tutors based in India and Sri Lanka access UK or US student data, and although TSL uses safeguards like the UK-U.S. Data Bridge and standard contractual clauses, these cross-border transfers raise concerns about potential data breaches or unauthorised access. Additionally, the retention of student data for up to two years after they stop using the platform could lead to unease, especially given the sensitive nature of the information.

Furthermore, **pedagogical concerns** about TSL's approach have been raised. While the company's 'I do, we do, you do' approach provides a structured learning path, it may oversimplify complex educational processes. This model emphasises direct instruction and incremental student independence but may not provide sufficient space for critical thinking, creativity, or experiential learning—all of which are key components of deeper, autonomous learning. Additionally, while TSL claims to align with classroom teaching, it could potentially lead to a fragmented learning experience if students' online tutoring sessions are not seamlessly integrated with their broader educational goals. Although AI systems can provide quick feedback and adaptive learning, they often fail to engage students in open-ended questioning, debate, or collaborative problem-solving—elements that are essential for fostering educational functions beyond 'learnification' and 'qualification'. These additional functions, such as *socialisation* and *subjectification*, are crucial for education. As Biesta (2011: 21-22, as cited in Holmes & Tuomi, 2022) notes, *socialisation* refers to the way education helps individuals become part of specific social, cultural, and political orders, while *subjectification* enables students to become more autonomous and independent in their thinking and acting. This raises the question of whether AI-based tutoring may inadvertently promote a 'one-size-fits-all' model of education, potentially undermining the role of human educators in cultivating higher-order thinking skills.

Lastly, the issue of **equity** must also be considered. Although TSL aims to close the attainment gap for disadvantaged students, its success is dependent on the availability of digital infrastructure, access to technology, and digital literacy. In under-resourced environments, technical difficulties such as unstable internet connections or limited hardware access can severely impact the quality of the tutoring sessions. This may inadvertently widen the gap between students in well-funded schools and those in economically disadvantaged areas, despite TSL's mission of promoting educational equity. Moreover, schools may need to invest in teacher training about AI (known as 'AI Literacy') to ensure that educators understand AI-based learning and its ethical implications, an area that remains underdeveloped.

SECTION 3.2

AI and SDG 10 (Reduced Inequality)

In its recent *Inequality Inc.*[158] report, Oxfam International warns about the high concentration of financial and monopoly power by a handful of global corporations exacerbating economy-wide inequality. According to the report, "the wealth of the world's five richest billionaires has more than doubled since the start of this decade, while 60% of humanity has grown poorer" (Riddell et al., 2024, p. 8). The report highlights gender inequality, noting that "globally, men own US$105 trillion more wealth than women—the difference in wealth is equivalent to more than four times the size of the US economy" (Riddell et al., 2024, p. 9).

Similarly, the *SDG Report 2024* highlights how the global share of people living on less than half the median income has been declining due to social assistance programs, despite economic disruptions from the pandemic (United Nations, 2024, p. 28). In 2023, the *SDG Report* maintains, a tragic milestone occurred as it became the deadliest year on record for migrants, with 8,177 fatalities documented. The majority of deaths occurred on routes taken by migrants affected by crises. On a positive note, the *SDG Report* emphasises the importance of digital remittance services, which could help reach the target faster: "To achieve the SDG target, stakeholders could leverage digitalization to reduce costs, increase efficiency and improve remittance accessibility" (United Nations, 2024, p. 29).

Understanding the historical interplay between inequality, technology, and policy provides valuable context for examining the contemporary relationship between AI and SDG 10, and the impact it may have today. For instance, in the early 19th century, the United States established poorhouses as one of the first institutional responses to poverty. These institutions were not merely about providing relief but also about enforcing moral discipline and work ethics among the poor (Katz, 1986). The underlying belief was that poverty resulted from individual moral failings rather than systemic issues. As Eubanks (2018, p. 17) highlights, "Josiah Quincy III, scion of a wealthy and influential Unitarian family [...] suggested two clauses of paupers. The *impotent* poor, he wrote in 1821, were "wholly incapable of work, through old age, infancy, sickness or corporeal debility", while the *able* poor were just shirking".

By the late 19th and early 20th centuries, the eugenics movement gained prominence in the United States and Europe. Eugenics, rooted in pseudo-scientific beliefs about heredity and social Darwinism, advocated for the improvement of the human population and elimination of social ills through controlled breeding.

[158] Riddell, R., et al., (2024). *Inequality Inc. How corporate power divides our world and the need for a new era of public action.* [pdf] Available at: https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/2024-01/Davos%202024%20Report-%20English.pdf [Accessed on 14 October 2024]

Policies were implemented to encourage reproduction among those deemed "fit" and to restrict or prevent reproduction among those labelled "unfit," often targeting marginalised groups including people of colour, the poor, immigrants, and individuals with disabilities (see Bashford & Levine, 2010). Technological interventions played a significant role in eugenic practices. For example, the eugenics movement in the early 20th century led to the creation of some of the first extensive databases on the poor and other marginalised groups:

> From a Carnegie Institution-funded laboratory in Cold Spring Harbor, New York, and state eugenics records offices stretching from Vermont to California, social scientists fanned out across the United States to gather information about poor people's sex lives, intelligence, habits, and behavior. They filled out lengthy questionnaires, took photographs, inked fingerprints, measured heads, counted children, plotted family trees, and filled logbooks with descriptions like "imbecile," "feeble-minded," "harlot," and "dependent." (Eubanks, 2018, p. 22-23)

Just as eugenicists collected extensive personal data to categorise individuals into fixed hierarchical groups, contemporary AI technologies often rely on large datasets to profile and segment populations. Data brokers collect and sell vast amounts of personal information, which AI algorithms use to cluster individuals into categories for purposes like targeted advertising, credit scoring, predictive policing, and social services eligibility. For example, AI systems used in criminal justice have been criticised for perpetuating racial biases. Predictive policing algorithms may disproportionately target marginalised communities by relying on historical crime data that reflect systemic discrimination (Angwin et al., 2016). Similarly, in social services, algorithms used to assess risk or eligibility can unfairly classify individuals based on socio-economic factors, echoing the invasive scrutiny of the poor in past eugenic practices (Eubanks, 2018).

Dan McQuillan (2022) extends this critique by arguing that the mathematical models and statistical methods employed in AI inherently reproduce existing social hierarchies and exclusions. Additionally, McQuillan (2022) contends that AI's classification and prediction mechanisms echo the eugenicists' attempts to categorise and control populations. By assigning individuals to predetermined categories based on patterns detected in data, AI can limit opportunities and reinforce stereotypes. An illustrative example of this is the Dutch SyRI (System Risk Indication**)** case. Developed by the Dutch government, SyRI was an algorithmic system intended to detect welfare fraud by analysing data from various government agencies, including tax records, employment history, and housing information. The system aimed to identify individuals at risk of committing fraud by uncovering hidden patterns and correlations. However, SyRI predominantly targeted low-income neighbourhoods with high immigrant populations, effectively profiling and stigmatising marginalised communities without their knowledge (van Brakel, 2020). In 2020, a Dutch court ruled that SyRI violated human rights, particularly the right to privacy as protected by the European Convention on Human Rights (ECHR). The court emphasised that the system's lack of transparency and the potential for discrimination outweighed the government's interest in detecting fraud (Rechtbank Den Haag, 2020). The judgement highlighted concerns about automated decision-making processes lacking sufficient safeguards against bias and the infringement of individual rights.

A similar concern arises with the deployment of AI technologies initially designed for city cleanliness[159] being used to monitor and manage homeless populations. For example, AI-powered surveillance systems that detect litter and waste can be repurposed to identify and track homeless individuals in public spaces[160]. This practice risks stigmatising and marginalising vulnerable groups by framing homelessness as a nuisance or hazard to be eradicated, rather than addressing the root causes of poverty and housing insecurity.

Despite these concerns, AI promises a potential to contribute positively towards reducing inequality, aligning with the objectives of SDG 10. By leveraging human-centric approaches to AI, various organisations and initiatives are working to address social disparities and support marginalised communities. AI's potential to augment labour, rather than merely automate tasks, can lead to productivity improvements for lower-wage workers, enabling them to perform at higher levels without extensive education or experience[161]. One promising area is the use of AI in humanitarian contexts to predict and respond to crises more effectively. For instance, Save the Children has been exploring the use of predictive analytics to build better futures for displaced children[162]. By analysing patterns in data related to displacement, conflict, and environmental factors, AI models can help anticipate humanitarian needs and enable timely interventions.

A socially-grounded Africa-centric research and product lab, Lelapa, is working towards bridging the gap and the divide between current commercially available Large Language Models which can't understand African languages by developing language technologies that support underrepresented African languages[163]. Their work enables better access to information and services for communities often left out of technological advancements due to language barriers.

Suel et al. (2019) applied a deep learning model to street-level images to measure inequalities related to income, education, unemployment, health, housing, and crime across four major UK cities. This approach leveraged public imagery data from sources like Google Street View, paired with official government statistics, to assess and monitor disparities in urban environments with high spatial resolution. The study's results revealed that "the application of deep learning to street imagery better predicted inequalities in some outcomes (i.e. income, living environment) than others (i.e. crime, self-reported health)." (Suel et al., 2019). Additionally, the study demonstrated the model's adaptability across different cities, as

[159] EE Times Asia, (2021). *Using AI to Keep City Clean Makes Amsterdam GO SMART Award Winner.* [online] Available at: https://www.eetasia.com/using-ai-to-keep-city-clean-makes-amsterdam-2021-go-smart-award-winner/ [Accessed on 14 October 2024]
[160] Feathers S.. (2024). *Revealed: A Californian city is training AI to spot homeless encampments.* [online] Available at: https://www.theguardian.com/technology/2024/mar/25/san-jose-homelessness-ai-detection [Accessed on 14 October 2024]
[161] Ajay, Agrawal., Joshua, S., Gans., Avi, Goldfarb. (2023). *Do we want less automation?*. Science, 381, 155-158. Available from: 10.1126/science.adh9429
[162] Kaplan J., Morgan S.. (2018). *Predicting Displacement: Using predictive analytics to build a better future for displaced children.* [pdf] Available at: https://resourcecentre.savethechildren.net/pdf/predicting_displacement_report_-_save_the_children_mdi.pdf/ [Accessed on 14 October 2024]
[163] Tsanni A., 2023. *This company is building AI for African languages.* [online] Available at: https://www.technologyreview.com/2023/11/17/1083637/lelapa-ai-african-languages-vulavula/ [Accessed on 14 October 2024]

networks trained in London achieved comparable results when fine-tuned with minimal local data from Birmingham, Manchester, and Leeds. The authors argue that the use of AI and street imagery could significantly enhance traditional methods of urban surveillance, offering timely insights that could inform policy decisions aimed at reducing inequalities. They suggest that with improvements in data availability, especially in data-poor settings, this approach could help city governments better allocate resources and evaluate the effectiveness of interventions aimed at addressing urban inequalities. The study highlights the scalability of this method, suggesting its potential application in diverse urban contexts globally, which could greatly benefit regions with limited access to reliable socio-economic data.

Building on these examples, the research by Aiken et al. (2020) further illustrates how AI can be harnessed to target inequality in regions with limited data infrastructure. The study explores how ML, coupled with non-traditional data sources such as mobile phone metadata and satellite imagery, can enhance the targeting of humanitarian aid in low- and middle-income countries. For example, the analysis of Togo's Novissi program reveals that this ML approach reduces exclusion errors by 4-21% compared to geographic targeting methods, highlighting the potential of new data sources to enhance humanitarian response efforts in crisis situations[164].

However, while AI shows promise in addressing aspects of inequality, much of this potential remains in the realm of experimentation and theoretical application, with many initiatives still in the early stages of development. There is a growing body of evidence highlighting AI's potential benefits, yet robust independent assessments and long-term sustainability remain limited. In contrast, the harms associated with AI, particularly in the reinforcement of existing inequalities, are increasingly well-documented. These issues warrant a careful examination of AI's impact on inequality, especially as they relate to systems currently in use. In the following sections, two case studies will be explored: the use of AI-enhanced Proxy Means Testing approaches utilised by the World Bank, and the analysis of spatial apartheid through AI and satellite imagery by the Distributed AI Research Institute (DAIR). These case studies offer concrete examples of AI's mixed impact on inequality, illustrating both its potential benefits and risks.

3.2.1 | Case Study: The Study of Spatial Apartheid in South Africa

Although apartheid officially ended in 1994, South Africa continues to grapple with the lingering effects of spatial apartheid—physical divisions based on race and class that manifest in the country's urban and rural landscapes. These effects significantly contribute to deepening inequality by perpetuating systemic disparities in access to resources, opportunities, and essential services such as quality education, healthcare, and infrastructure. For instance, schools in townships and rural areas often lack adequate funding, qualified teachers, and learning materials, leading to lower educational outcomes compared to

[164] It should be noted that the paper acknowledges the limitations of mobile phone data, particularly regarding the exclusion of individuals without phone access, which may affect the targeting of humanitarian aid programs. This highlights a gap in understanding how to include these populations effectively. Furthermore, there is a need for further research on the social, political, and ethical implications of using mobile phone data for targeting, which is not fully explored in the paper.

schools in affluent, predominantly white areas.[165] Despite comprising the majority of the population, black households receive a smaller share of the national income, exacerbating wealth inequality and limiting economic advancement.[166]. As the World Bank (2022) notes, South Africa is the most unequal country in the world, with race being the largest contributor to this inequality—a contribution that is growing. According to International Monetary Fund (IMF) general unemployment in South Africa is one of the world's highest with more than 33 percent[167], and nearly 40% of Black South Africans were unemployed during the first three months of 2023, while the unemployment rate among White South Africans stood at 7.5%[168]. Spatial apartheid reinforces social divisions by maintaining physical barriers between communities—for example, townships are often located on the outskirts of cities, separated from affluent neighbourhoods by physical buffers—which leads to social isolation, limited interaction among different racial and economic groups, and restricted access to amenities[169].

To investigate the enduring effects of spatial apartheid, the DAIR researchers developed a comprehensive dataset—including 6,768 high-resolution satellite images covering all nine provinces, with 550 images labelled—by integrating multiple sources of geographical and demographic information. They utilised high-resolution satellite imagery from 2006 to 2017, provided by the South African National Space Agency, which offered detailed visual representations of the country's urban and rural landscapes. Alongside this, they incorporated land use data from the Enumeration Areas (EAs) dataset by Statistics South Africa, which classifies land based on its designated use, and geospatial data indicating the locations of all buildings across South Africa.

Central to their methodology was the application of AI and ML algorithms. The team employed AI to process and analyse the vast amounts of data, enabling the classification of neighbourhoods into categories such as wealthy areas, non-wealthy areas, and non-residential neighbourhoods based on visual characteristics identified in the satellite images[170]. Techniques like 'buffer algorithms' were used to

[165] Lawal S., (2024). *South Africa: 30 years after apartheid, what has changed?* [online] Available at: https://www.aljazeera.com/news/2024/4/27/south-africa-30-years-after-apartheid-what-has-changed [Accessed: on 17 October 2024].

[166] World Bank, (2022). *Inequality in Southern Africa: An Assessment of the Southern African Customs Union*. Washington, DC: World Bank. [pdf] Available at: https://documents1.worldbank.org/curated/en/099125003072240961/pdf/P1649270b73f1f0b5093fb0e644d33bc6f1.pdf [Accessed on 17 October 2024]

[167] International Monetary Fund, (2024). *Unemployment Rate - World Economic Outlook (WEO) Data Mapper - IMF Data Mapper*. [online] Available at: https://www.imf.org/external/datamapper/LUR@WEO/VNM/THA/SGP/PHL/MYS/IDN [Accessed on 17 October 2024].

[168] Statistics South Africa, 2023. *Quarterly Labour Force Survey: Quarter 1, 2023*. Available at: https://www.statssa.gov.za/publications/P0211/P02111stQuarter2023.pdf [Accessed on 17 October 2024].

[169] Sefala R., Gebru T., Mfupe L., Moorosi N., Klein R., (2021). *Constructing a Visual Dataset to Study the Effects of Spatial Apartheid in South Africa. Constructing a Visual Dataset to Study the Effects of Spatial Apartheid in South Africa', Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS).* [pdf] Available at: https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/07e1cd7dca89a1678042477183b7ac3f-Paper-round2.pdf [Accessed on 17 October 2024]

[170] Tsanni A., 2024. *How satellite images and AI could help fight spatial apartheid in South Africa* . [online] Available at: https://www.technologyreview.com/2024/01/19/1086837/satellite-images-ai-spatial-apartheid-south-africa/ [Accessed on 17 October 2024]

approximate building areas, and the involvement of local students in identifying townships added valuable on-the-ground insights to enhance the dataset's accuracy.

Despite limited resources, the use of AI helped in managing the scale and complexity of the data. The team faced challenges such as dealing with imbalanced datasets—where some regions had significantly more data than others—and ensuring accurate classification across diverse areas with varying visual features. Ensuring consistency and accuracy required meticulous validation processes and the development of algorithms capable of adapting to different regional characteristics.

A unique feature of the research was the active participation of people on the ground, particularly local students who brought invaluable firsthand knowledge of the areas being studied. The team recruited 10 graduate students from the University of Witwatersrand, each hailing from different provinces, to assist in identifying and labelling townships that were not distinctly classified in official datasets. These students leveraged their personal experiences and consulted with community members and their relatives to accurately distinguish between townships and suburbs, ensuring that the dataset reflected the nuanced realities of South African neighbourhoods.

The AI-enhanced dataset developed by DAIR has significant applications for researchers, policymakers, and advocacy groups. It enables a detailed analysis of urban development patterns, allowing for the study of how neighbourhoods have evolved and how townships have expanded over time. By revealing the spatial distribution of wealth and resources, the dataset supports informed decision-making aimed at addressing inequalities.

The data facilitates the equitable allocation of public services by providing evidence-based insights into which areas are underserved. For instance, the data has been shared with a South African policy think tank, Human Sciences Research Council (HSRC), to advise on budget allocations for HIV treatment programs, ensuring that resources reach the communities most in need (Tsanni 2024). The ability to pinpoint areas lacking in services is greatly enhanced by the AI's capacity to analyse complex spatial data efficiently.

Moreover, the dataset may open new possibilities for organisations working towards urban land justice by providing concrete, data-driven evidence of disparities. Advocacy groups can use these insights to support calls for policy changes and interventions aimed at reducing inequality. The application of AI in processing and analysing the dataset has been instrumental in uncovering subtle patterns of segregation and resource distribution that might otherwise remain hidden.

By prioritising a human-centric approach and utilising AI and ML as a supportive tool, DAIR's work not only sheds light on the persistent issue of spatial apartheid but also provides practical means to address it, aligning with the goals of reducing inequalities as outlined in SDG 10.

3.2.2 | Case Study: Proxy Means Testing

As global inequalities deepen and the demand for social safety nets grows, targeting mechanisms have become central to social protection strategies in developing countries. Proxy Means Testing (PMT) has emerged as a widely-used method, particularly in countries with large informal economies where direct income data is unreliable. The concept of PMT originated from the efforts in the 1980s and 1990s to improve the targeting of social programs, especially in Latin America. Countries like Chile and Mexico were early adopters of targeted social protection mechanisms, and their experiences were instrumental in shaping the development of PMT. As part of the Living Standards Measurement Study (LSMS), the World Bank formalised PMT as a tool to address the challenges of social protection programs in contexts where reliable income data was scarce. The first formal implementations of PMT occurred in Latin America during the 1990s, responding to rising social inequalities and the need for more efficient use of limited social assistance funds. Since then, PMT has been adopted by numerous countries across Africa, Asia, and Eastern Europe.

PMT is employed by governments and international organisations, such as the World Bank, to identify and provide support to the poorest households using non-income-based indicators, or proxies. The term "proxy means test" refers to the use of household or individual characteristics that are correlated with welfare levels—such as assets, housing conditions, education levels, and other observable factors—in a formal algorithm to estimate household income, welfare, or need[171]. For instance, a household that owns a refrigerator or vehicle may be considered wealthier than one without these assets. The PMT model inputs these characteristics into a statistical algorithm that estimates the household's overall welfare status. This method bypasses the need for direct income verification, which is often unreliable or impractical in informal economies.

PMT operates on the assumption that certain household traits, when combined, provide a reliable estimate of welfare. Governments have embraced this approach to allocate limited resources in programs like cash transfers, food assistance, and healthcare subsidies, ensuring that assistance reaches the neediest households.

Box 3.1 Illustrative Example of Proxy Means Testing (PMT) Implementation

In Country X, the government was faced with the challenge of identifying the poorest households for a new social protection program. Since many families in Country X worked in the informal economy, their income was often unreported, making it difficult to determine who needed the most help. To solve this, officials decided to use Proxy Means Testing (PMT), a method that relies on observable household characteristics rather than direct income data.

Sana, a mother of three living in a rural village, was one of the many people who were assessed for the program. When the government enumerator visited her home, they didn't ask for income details. Instead, they recorded information such as the number of people in her household, the materials used for the roof and walls, and whether Sana owned items like a refrigerator, TV, car or a bicycle. These factors, called proxies, were indicators of a household's overall welfare.

Once the data from Sana's household and others were collected, it was fed into a statistical model. The

[171] Grosh M.E. and Baker J.L., (1995). *Proxy means tests for targeting social programs: simulations and speculation*. Living Standards Measurement Study (LSMS) working paper, no. LSM 118. Washington, D.C.: The World Bank. Available at: http://documents.worldbank.org/curated/en/750401468776352539/Proxy-means-tests-for-targeting-social-programs-simulations-and-speculation [Accessedon 21 Oct. 2024].

model assigned different weights to each proxy based on how strongly it correlated with poverty in the country. For example, owning a refrigerator might decrease the likelihood of being classified as poor, while having a dirt floor might increase it. The model combined these weighted proxies into a score for each household. If Sana's score was low enough she qualified for the program, if the score was high she didn't qualify for the program remaining outside of the assistance scheme.

This PMT approach allowed Country X to identify the neediest households using a set of proxies, without having to rely on hard-to-obtain income data. By using an algorithm to assign weights to various factors, the government could make efficient and fair decisions about who received assistance.

Adaptations and Variations of PMT

Since its inception, PMT has seen numerous adaptations and refinements. Different countries implement PMT based on their specific contexts, resulting in varying outcomes. For example, the PMT method in Vietnam has been shown to have higher coverage and lower leakage rates compared to traditional income questionnaires, allowing for more transparent poverty estimation across regions[172]. In Malawi, systematic comparisons of different PMT models have been conducted to improve targeting efficiency, highlighting the costs and benefits of indicator-based targeting[173].

It is important to note that PMT is not a one-size-fits-all solution; rather, it comes in many variations depending on the country and the available data. The basic model involves selecting proxies that are most strongly correlated with poverty, but the choice of proxies and the way the model is implemented can vary considerably. For instance, the most common PMT approach uses *regression analysis* to predict household welfare based on a set of observable characteristics. Household surveys are conducted to identify the variables that best predict poverty, and these are then used to develop a formula. This formula assigns each household a score based on their reported proxies, and households below a certain threshold are classified as poor.

In some contexts, countries may assign different weights to the proxies based on their importance in predicting poverty. For example, in regions where access to education is a key determinant of economic mobility, education level might be weighted more heavily than asset ownership. These weights can vary not only by country but also by region within a country, providing a more tailored approach to targeting. This is known also as *weighted proxy models*. Some countries have moved towards *hybrid models* that combine PMT with other targeting approaches, such as community-based targeting or self-targeting. Additionally, some countries are exploring *dynamic PMT models* that can be updated more frequently as new data becomes available.

[172] Cuong, Nguyen., Dat, Tho, Tran. (2018). 4. *Proxy Means Tests to Identify the Income Poor: Application for the Case of Vietnam*:. Journal of Asian and African Studies. Available at: https://doi.org/10.1177/00219096177094 [Accessed on 21 Oct. 2024].
[173] Houssou I. N., (2013). *Operational Poverty Targeting by Proxy Means Tests: Models and Policy Simulations for Malawi*. 1st ed. Frankfurt: Peter Lang GmbH.

Proxy Mechanisms: Automating Poverty and Exclusion

Despite its growing popularity, PMT has been the subject of intense debate, with numerous studies critiquing its effectiveness and fairness, as well as highlighting political and social costs of such a system. Furthermore, there is little robust evidence in the literature on the PMT's effectiveness after implementation[174].

One major concern is that algorithms can have an adverse effect on vulnerable populations even without explicitly including protected characteristics. This often occurs when a model includes features that are correlated with these characteristics - known as 'proxies'[175]. A common instance of proxy discrimination is "redlining" in the financial industry. In the mid-20th century, rather than openly discriminating based on race in their underwriting and pricing practices, certain financial institutions utilised zip codes and neighbourhood boundaries as substitutes for race to restrict lending to areas predominantly inhabited by African Americans[176].

Similarly, PMT's reliance on static household characteristics also poses challenges in countries with dynamic and rapidly changing economies. A household's poverty status may fluctuate significantly due to shocks such as natural disasters or economic downturns, which PMT does not quickly capture. The fixed proxies used in the test may not reflect real-time changes in living conditions, leading to inappropriate assessments of eligibility for social programs: "These algorithms only capture a static picture of what people are going through at a single point in time — but this is not how people actually suffer" (Toh, 2023, as cited in Osseiran, Asher-Schapiro, and Farouk, 2023).

One of the major concerns about the methodology is the high rate of errors in both inclusion and exclusion. Studies have shown that PMT systems in various countries, including Bangladesh and Rwanda, can misclassify up to 55% of households (Kidd and Wylde, 2011: 8). This means that a significant portion of poor households may be excluded from benefits, while relatively better-off households might be included. Additionally, the inaccuracies inherent in PMT can negatively impact community cohesion by creating divisions between beneficiaries and non-beneficiaries. Studies have documented feelings of resentment, jealousy, and despair among those excluded from assistance programs, even when they are equally needy. In Mexico and Nicaragua, non-beneficiaries expressed anger and jealousy, which sometimes escalated into conflict (see Kidd & Wylde: 2011). In Indonesia, significant exclusion errors in the Program Keluarga Harapan led to tension and unrest, including incidents of stone-throwing and arson, making the work of program facilitators dangerous (Hannigan: 2010, as cited in Kidd & Wylde: 2011). As Kidd and Wylde note (2011) stigmatisation is another social cost associated with PMT. In Mexico, some individuals refused to admit their poverty during surveys due to shame, leading to their exclusion from assistance programs.

[174] Kidd, S. and Wylde, E., (2011: 4). *Targeting the Poorest: An assessment of the proxy means test methodology*. Canberra: Australian Agency for International Development (AusAID). [pdf] Available at: https://www.dfat.gov.au/sites/default/files/targeting-poorest.pdf [Accessed 21 October 2024].
[175] See: MIT Media Lab, (2024). *Discrimination by Proxy. AI Blindspot. [*online] Available at: https://aiblindspot.media.mit.edu/discrimination_by_proxy.html [Accessed 21 October 2024].
[176] See : BakerHostetler, (2022). *Proxy problems: solving for discrimination in algorithms*. [online] Available at: https://www.bhfs.com/insights/alerts-articles/2022/proxy-problems-solving-for-discrimination-in-algorithms [Accessed 21 October 2024].

On the other hand, politically, PMT can undermine broader support for social protection programs. Conventional economic theory suggests that programs excluding the middle class are less likely to receive significant public funding (Sen, 1995). Comparative analyses indicate that programs using PMT often have smaller budgets than more inclusive programs, such as universal old-age pensions. While many PMT-based programs operate with budgets of no more than 0.4% of a country's gross domestic product (GDP), universal programs can command budgets exceeding 1% (Kidd and Wylde: 2011). This implies that tightly targeted PMT programs may deliver smaller benefits to poor households compared to universal programs with broader political support. The exclusion of a significant portion of the population can lead to decreased political backing and sustainability challenges for PMT-based initiatives. In contrast, universal programs tend to garner wider public and political support, ensuring more substantial and stable funding, which ultimately benefits the poor more effectively.

Despite fundamental flaws and evidenced social and political costs, numerous international organisations and governmental agencies have called for deeper reflection and pause of PMT. For example, Australia's Aid Agency study (Kidd and Wylde: 2011) "show that the errors inherent in the PMT methodology are significant, there is so far no evidence that other forms of poverty targeting perform any better in developing countries with large informal sectors, weak administrative capacity and low fiscal space" (p. 32). International Labour Office's working paper (Kidd, Gelders, Bailey-Athias: 2017) concludes by qualifying PMT as "a lottery in which the poorer a household, the more lottery tickets it has." (p. 18). The paper notes also that the PMT is not compliant with a human rights approach "since it cannot ensure that most people, including the most vulnerable, are able to access social security, [and] through its 'exclusion by design,' it usually guarantees that the majority of those in need will miss out." (p. 18). As Kidd et al. (2017) argue, many developing country governments have been persuaded by PMT advocates to adopt the mechanism—often linked to the acceptance of loans—without being fully informed of its significant inaccuracies and arbitrary performance.

The lack of transparency in PMT processes, particularly regarding the weighting of indicators in the algorithm, has been a significant point of concern. As noted in the ILO Policy Note[177], "One aspect giving rise to concern is that information on how indicators in the unified scoring formula are weighed is not publicly available. The definition of eligibility criteria should be open for political debate and contestation" (2022: 6). This lack of openness restricts public oversight and limits opportunities for constructive debate on the fairness and accuracy of eligibility criteria, contributing to concerns around exclusion and fairness in the targeting process.

Notwithstanding years of critique, PMT continues to be implemented in many developing countries, frequently linked to international funding conditions and loans, even though its methodological flaws and potential for error are well-documented, as highlighted above.

[177] International Labour Organisation, (2022). *The Economic Assistance Programme in Albania: Challenges and Reform Trends.* [pdf] Available at: https://albania.un.org/sq/download/136372/237264 [Accessed on 27 October 2024].

2023 reports from Human Rights Watch[178] (HRW), Amnesty International[179], and news outlets[180] reveal that the adverse effects of PMT are ongoing and, in some cases, worsening due to the increased reliance on automated, algorithm-driven processes in resource allocation. These reports underline how PMT continues to automate poverty assessment in ways that often exclude vulnerable populations and entrench economic divides.

HRW Report "Automated Neglect" (2023) exposes the impact of PMT-based cash assistance systems in the MENA region (Mauritania, Morocco, Tunisia, Lebanon, Palestine, Egypt, Jordan and Iraq). Drawing on vast amount of interviews with applicants and beneficiaries, as well as government officials from MENA region, HRW highlights how the algorithm is driving decisions that deprive people's rights to social security:

> The problem is not merely that the algorithm relies on inaccurate and unreliable data about people's finances. Its formula also flattens the economic complexity of people's lives into a crude ranking that pits one household against another, fueling social tension and perceptions of unfairness. (p. 1)

Additionally, the report argues that the PMT algorithm reinforces gender-based discrimination:

> Its calculation of household size, one of the measures of vulnerability, only considers the number of Jordanian members in the household. This formula artificially shrinks the size of households headed by Jordanian women with noncitizen spouses and children because the law does not recognize their right to pass on citizenship to these family members on an equal basis with men, lowering their benefit payments or excluding them from the program entirely. Women in male-headed households are not spared from gendered design choices either: awarding Takaful payments to heads of households, who in Jordan are usually regarded as the husband or father, rather than individual adult
> members, heightens dependency on male family figures and unduly restricts women's access to benefits. (p. 4)

Similar problems were also found in Serbia, where the World Bank-funded Social Card registry introduced automation, incorporating a data-driven system into the process of determining eligibility for social

[178] Human Rights Watch, (2023). *Automated Neglect: How the World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights*. [pdf] Available at: https://www.hrw.org/sites/default/files/media_2023/11/thr_jordan0623%20web.pdf [Accessed on 30 October 2024]

[179] Amnesty International, (2023). *Trapped by Automation: Poverty discrimination in Serbia's welfare state.* [online] Available at: https://www.amnesty.org/en/latest/research/2023/12/trapped-by-automation-poverty-and-discrimination-in-serbias-welfare-state/ [Accessed on 30 October 2024]

[180] Osseiran N., Asher-Schapiro A., Farouk M., (2023). *In Middle East, poor excluded from welfare by 'faulty' algorithms.* [online] https://www.japantimes.co.jp/news/2023/10/05/world/society/middle-east-poor-algorithms-aid/ [Accessed on 30 October 2024]

assistance[181]. In the case of Serbia, people with protected characteristics, such as Roma and people with disabilities, are most suffering the effects of welfare automation:

> [...] data is often woefully low in quality, especially for individuals from marginalised groups whose documentation and records in government systems are often, for structural reasons, not up to date. The reliance on this inaccurate data drawn from databases that are not regularly updated or data that ignores the realities of an individual's complex economic situation, leads to an increased likelihood of mistakes, removing people from receiving benefits they have the right to access. [Additionally] the semi-automated layer has reduced the role of social workers in verifying the data and documents of applicants. Previously, social workers would conduct field visits and interviews to ascertain the complex realities of people's lives, however, the Social Card registry has reduced the economic realities of people living off informal workstreams and with hugely varying personal circumstances to, often outdated, data points.

Drawing on McQuillan (2021, p. 1), who argues that "AI is never separate from the assembly of institutional arrangements that need to be in place for it to make an impact in society," we observe that Serbia's Social Card registry exemplifies the "layered and interdependent arrangement of technology, institutions, and ideology" (McQuillan, 2021, p. 1). For instance, the structural discrimination against Roma people is reflected in the lack of quality data about this group, which is then fed into the (semi-)automated Social Card registry. In turn, the output of the algorithm further intensifies and amplifies discriminatory practices toward Roma people, undermining their right to equal access to social assistance.

Over more than two decades, POMT has shaped social assistance programs across diverse regions, from the Middle East and North Africa to Eastern Europe and Asia, often entrenching rather than alleviating social inequalities. Despite ongoing criticism and evidence of its flaws, PMT's rigid, algorithmic criteria frequently fail to account for the complex realities of poverty, particularly in informal economies. The system's reliance on proxies and outdated data compounds errors, leading to exclusion and intensifying discrimination against marginalised groups, such as the Roma community in Serbia and non-citizen families in Jordan. This exclusion is not merely an operational flaw; it also undermines multiple Sustainable Development Goals (SDGs). While PMT aims to streamline poverty alleviation efforts under SDG 1 (No Poverty) and SDG 10 (Reduced Inequalities), its limitations and blind spots contribute to negative outcomes in other areas, including SDG 5 (Gender Equality). For example, PMT systems that designate male heads of households as beneficiaries reinforce economic dependency for women and restrict their access to resources. By automating exclusion and reinforcing existing biases, PMT inadvertently sustains structural inequalities, emphasising the need for a universal, rights-based approach to social protection that truly aligns with the SDGs' mission to Leave No One Behind.

[181] Amnesty International, (2023). *Serbia: World Bank-funded digital welfare system exacerbating poverty, especially for Roma and people with disability.* [online] Available at: https://www.amnesty.org/en/latest/news/2023/12/serbia-world-bank-funded-digital-welfare-system-exacerbating-poverty-especially-for-roma-and-people-with-disabilities/#:~:text=Launched%20in%202022%20and%20aimed,determining%20eligibility%20for%20social%20assistance. [Accessed on 30 October 2024].

SECTION 3.3

AI and SDG 13 (Climate Action)

SDG 13 on Climate Action aims to create a climate-neutral world by mid-century and limit global warming to well below 2°C—ideally 1.5°C—relative to pre-industrial levels. It emphasises strengthening countries' climate resilience and adaptive capacity, particularly for least-developed nations. Alarmingly, 2023 was the hottest year in recorded history[182], with global temperatures reaching 1.45°C above pre-industrial levels—dangerously close to the 1.5°C limit set by the Paris Agreement. Furthermore, global greenhouse gas emissions hit a new peak in 2022, reaching 57.4 gigatons of CO2 equivalent, as reported in the United Nations Environment Programme's *Emissions Gap Report 2023*[183].

Europe saw one of its hottest summers in 2023, resulting in widespread destruction and over 47,000 heat-related deaths[184]. According to a study[185] by the United Nations Office for Disaster Risk Reduction (UNISDR), 90% of major disasters between 1995 and 2015 were climate-related, stemming from nearly 6,500 events such as floods, storms, and heatwaves. This trend persists, as evidenced by a 2023 Save the Children study[186], which estimated that at least 12,000 lives were lost globally in climate-related disasters in 2023—a 30% increase from 2022. Between 11–15 September 2024, record-breaking rainfall from Storm Boris triggered the 2024 Central European floods, impacting Austria, the Czech Republic, Poland, Romania, Slovakia, Germany, and Hungary and resulting in over 20 fatalities.

Italy is facing significant impacts from the climate crisis as well. Legambiente reports[187] that many glaciers in the Alps are now in critical condition, with rapid melting observed since the early 2000s. Together with Italian and European environmental organisations, Legambiente has surveyed 12 glaciers across Italy and nearby countries, documenting severe ice loss. The effects of climate change on glaciers were tragically highlighted by the collapse of the Marmolada glacier in July 2022, which triggered a massive avalanche of

[182] World Meteorological Organization (WMO), (2023). *Climate Change Indicators Reached Record Levels in 2023.* [online] Available at: https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-wmo#:~:text=The%20WMO%20report%20confirmed%20that,ten%2Dyear%20period%20on%20record [Accessed on 30 October 2024]

[183] UN Environmental Programme, (2023). *Broken Record: Temperatures hit new heights, yet world fails to cut emissions (again).* [pdf] Available at: https://wedocs.unep.org/bitstream/handle/20.500.11822/43922/EGR2023.pdf?sequence=3&isAllowed=y [Accessed on 30 October 2024]

[184] Yale Climate Connections, (2023). *Heat contributed to 47,000 deaths in Europe during summer 2023, study finds.* [online] Available at: https://yaleclimateconnections.org/2024/11/heat-contributed-to-47000-deaths-in-europe-during-summer-2023-study-finds/#:~:text=The%20summer%20of%202023%20was,47%2C000%20deaths%20across%20the%20continent. [Accessed on 30 October 2024]

[185] The United Nations Office for Disasters Risk Reduction, (2015). *The Human Cost of Weather Related Disasters.* [pdf] Available at: https://www.preventionweb.net/files/46796_cop21weatherdisastersreport2015.pdf [Accessed on 30 October 2024]

[186] Save the Children, (2023). *2023 in review: climate disasters claimed 12,000 lives globally in 2023.* [online] Available at: https://www.savethechildren.net/news/2023-review-climate-disasters-claimed-12000-lives-globally-2023 [Accessed on 30 October 2024].

[187] Legambiente, (2024). *L'Italia e l'Europa rischiano di rimanere senza ghiacciai.* [online] Available at: https://www.legambiente.it/comunicati-stampa/litalia-e-leuropa-rischiano-di-rimanere-senza-ghiacciai/ [Accessed on 01 November 2024].

snow, ice, and rocks, killing several hikers along its path[188]. Italy's exposure to climate hazards continues to escalate; in 2023, the country experienced 378 extreme weather events—an average of more than one per day—marking a 22% increase from 2022[189]. Lombardy, Emilia Romagna, and Tuscany were among the most impacted regions, with Emilia Romagna alone suffering 17 fatalities and €8.5 billion in certified damages from the 2023 floods[190].

Amid escalating climate challenges, AI has been promoted and seen as an indispensable asset in the pursuit of SDG 13—Climate Action. The following exploration looks into how AI is being harnessed across various sectors to support SDG 13, setting the stage for a comprehensive understanding of its potential and the imperative to navigate its complexities responsibly.

3.3.1 | Applications of AI for Climate Action

The application of AI in addressing climate change is diverse, involving predictive, generative, and assistive AI systems that support prediction, monitoring, mitigation, and adaptation efforts across global, national, and local levels. As the urgency of climate action grows, AI technologies are increasingly seen as pivotal in forecasting climate impacts, optimising resource allocation, managing environmental data, and guiding decision-making processes. However, while AI offers promising solutions, the swift adoption of these technologies risks reinforcing technological solutionism if it overlooks concrete challenges, such as data privacy, energy consumption, infrastructure ownership, and equitable access to AI systems. This section explores the primary applications of AI for climate action, structured across four key domains:

1. Climate Modeling and Prediction, and Disaster Forecasting
2. Monitoring and Conservation
3. Optimization in Energy and Resource Management
4. Urban Sustainability and Resilient City Planning

Climate Modelling and Prediction, and Disaster Forecasting

AI's capability to process large datasets and recognize complex patterns has significantly advanced climate modelling and disaster forecasting. ML algorithms enhance existing climate models by integrating diverse data sources—including historical climate records, satellite imagery, and real-time sensor data—into predictive models with potentially greater accuracy and adaptability[191]. For instance, neural networks are increasingly used to improve temperature, precipitation, and sea-level rise projections, critical for long-term climate policy and adaptation planning[192]. Beyond environmental projections, AI-driven models are also helping to assess climate impacts on the economy. Hsiang et al. (2017), for example, used climate data across six economic sectors to predict the effects of climate change on the U.S. economy based on short-term weather changes[193]. Their work demonstrates how integrating large datasets from economic and environmental fields with AI models can yield valuable insights into how climate change may reshape economies and affect resource distribution at national and local levels.

ML algorithms further enhance meteorological models by analysing extensive historical and real-time weather data, enabling more accurate predictions of climatic factors such as temperature, precipitation, and wind speed. For example, Chen et al. (2020) calculated daily evapotranspiration in the Northeast China Plain by comparing three models—deep neural networks, time convolutional neural networks, and short-term memory neural networks—against traditional methods like support vector machines, random models, and empirical equations[194]. However, despite a strong scientific consensus on the core aspects of climate change, accurately forecasting its outcomes remains challenging. This difficulty arises from the complexity of Earth system models and the inherent uncertainties associated with climate dynamics (Bonan and Doney: 2018, as cited in Chen, L., Chen, Z., Zhang, Y. *et al.*: 2023).

Rising sea levels, more frequent natural disasters, reduced crop production capacity, and biodiversity loss are all closely linked to climate change. Predictive AI is also being used for disaster forecasting (see also Chapter 2 of this research). AI models help predict the likelihood and impact of extreme weather events, such as hurricanes, floods, and wildfires, which are intensifying due to climate change. Projects like IBM's Green Horizons use AI to forecast pollution and weather patterns, allowing for proactive responses to environmental hazards. Similarly, the European Centre for Medium-Range Weather Forecasts (ECMWF) leverages ML for precise weather and climate event predictions, supporting disaster preparedness and emergency response efforts[195].

Monitoring and Conservation

AI plays a critical role in environmental monitoring and biodiversity conservation, helping to track and manage natural resources and ecosystems. Through satellite imagery and remote sensing data, AI systems can detect patterns and anomalies indicative of environmental degradation, such as deforestation, habitat loss, and ocean acidification. For example, the World Resources Institute's Global Forest Watch (GFW) platform uses ML to monitor forest cover changes and provide alerts on illegal logging activities, enabling

[188] Copernicus, (2022). *Collapse of the Marmolada glacier, Italy.* [online] Available at: https://www.copernicus.eu/it/node/11693 [Accessed on 01 November 2024]
[189] Carboni, K., (2023). *Nel 2023 la crisi del clima ha scatenato un estremo evento al giorno in Italia.* [online] Available at: https://www.wired.it/article/clima-crisi-2023-italia-evento-estremo-numeri/ [Accessed on 01 November 2024].
[190] Regione Emilia Romagna, (2024). *Alluvione, un anno dopo: il primo pensiero alle vittime. Il punto su quanto fatto dalla Regione*. [online] Available at: https://www.regione.emilia-romagna.it/notizie/2024/maggio/alluvione-un-anno-dopo [Accessed on 01 November 2024]
[191] Jones, N., (2017). *How machine learning could help to improve climate forecasting.* [pdf] Available at: https://doi.org/10.1038/548379a [Accessed on 01 November 2024]
[192] Rolnick, D., et al. (2019). *Tackling climate change with machine learning. Nature Climate Change*, 9(7), 498-504. [pdf] Available at: https://arxiv.org/pdf/1906.05433 [Accessed on 01 November 20234]
[193] Hsiang, S., Kopp, R.E., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D.J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K. and Houser, T., (2017). *Estimating economic damage from climate change in the United States*. *Science*, 356(6345), pp.1362-1369. [doi] Available at: https://www.science.org/doi/10.1126/science.aal4369 [Accessed on 01 November 2024]
[194] Chen, Z., Zhu, Z., Jiang, H., and Sun, S., (2020). *Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods*. *Journal of Hydrology*, 591, p.125286. [online] Available at: https://www.sciencedirect.com/science/article/abs/pii/S0022169420307460 [Accessed on 01 November 2024]
[195] Düben, P., (2020). *AI and machine learning at ECMWF.* [online] Available at: https://www.ecmwf.int/en/newsletter/163/news/ai-and-machine-learning-ecmwf [Accessed on 01 November 2024]

timely intervention[196]. This tool aids governments, NGOs, and indigenous communities in protecting forested areas and reducing illegal logging.

Similarly, Rainforest Connection (RFCx) utilises AI-driven technology to detect and respond to environmental threats, but it focuses on acoustic monitoring[197] rather than satellite imagery. Unlike Global Forest Watch, which relies on satellite data to create broad visual maps of deforestation, RFCx deploys solar-powered listening devices—known as "Guardians"—in trees to capture real-time audio data from the forest. These devices use ML algorithms to identify sounds associated with illegal activities, such as chainsaws and logging trucks, and alert local authorities or conservationists immediately, allowing for on-the-ground intervention.

Conservation efforts are increasingly supported by AI-driven technologies that enhance wildlife monitoring, population tracking, and habitat preservation. WildMe[198] is an innovative example, leveraging AI and ML to identify and catalogue individual animals through pattern recognition in photographs. By analysing unique features, such as the spots on a whale shark or the stripes on a zebra, WildMe's platform can track individual animals across different locations and times, helping researchers build comprehensive population databases and monitor animal movements. This data is crucial for understanding population health, migration patterns, and threats to species in real time. Unlike traditional tagging methods, WildMe's approach is non-invasive, relying on photos contributed by scientists, tourists, and citizen scientists, making conservation efforts more inclusive and scalable. Through partnerships with conservation organisations and wildlife researchers, WildMe empowers data-driven conservation strategies, aiding in the protection of endangered species and providing insights to support biodiversity conservation globally.

Optimization in Energy and Resource Management

Optimization in energy and resource management is another critical application of AI in climate action. Verma et al. (2024) note that AI's learning capabilities and rapid adaptation to parameter changes have made it widely applicable in smart controls for renewable energy systems, especially given its effectiveness in addressing uncertainties within energy models. The authors highlight AI's growing role in strengthening connectivity and decision-making within energy supply chains, particularly for renewable energy sources such as wind, solar, and hydrogen-based storage. As Rozite, Miller, and Oh (2023) further explain, AI is uniquely positioned to handle the vast complexity of modern power systems, which increasingly support multi-directional energy flows across distributed generators, grids, and users. This shift in the grid structure demands real-time analytical tools, and AI enables utilities to optimise electricity production, manage grid loads, and better align energy supply with fluctuating demand. For instance, Siemens leverages AI to balance grid loads, incorporating fluctuating renewable energy outputs and ensuring stability in energy supply[199]. And, Google's DeepMind has demonstrated how AI can refine wind power forecasts by up to 36 hours, allowing Google to sell renewable energy in advance and increase its financial value by 20%, while also enhancing overall grid efficiency (Rozite, Miller, and Oh: 2023). In 2019 Italy-based utility Enel began installing sensors on power lines to monitor vibration levels. ML algorithms allowed Enel to identify potential problems from the resulting data and discern what caused them. As a result, Enel has been able to reduce the number of power outages on these cables by 15%.

AI also enhances energy efficiency in buildings and industrial processes, playing a critical role in resource management. AI is being used in public buildings and research facilities with smart building management systems that monitor and control energy usage by dynamically adjusting heating, cooling, and lighting in response to occupancy patterns and external weather conditions. For instance, in manufacturing and public facilities, smart building systems utilise sensors and AI-powered data analytics to identify optimal energy use levels, adapt to real-time conditions, and reduce unnecessary consumption. These systems enable facilities to automatically adjust energy use during peak times and reduce waste during off-hours, which can significantly lower both emissions and operational costs[200]. Similarly, the World Economic Forum highlights how AI-driven solutions, such as those developed by BrainBox AI, optimise heating and cooling systems through continuous analysis of building data. By predicting energy needs and responding to changing conditions, these systems can cut energy costs by as much as 25% while reducing carbon emissions, illustrating the potential of AI to support sustainable energy use in building management[201].

Ultimately, the use of ML, Internet of Things (IoT) and sensor networks in precision agriculture reduces emissions by optimising water use, fertiliser application, and crop yield, supporting both food security and climate goals. These technologies enable precise monitoring and management of agricultural processes, leading to optimised resource use and improved productivity. For instance, ML technology used in precision planting with automated all-terrain vehicles (ATVs) has improved farming methods by allowing for accurate and flexible planting techniques[202]. ML algorithms can optimise planting depths, spacing, and seed kinds for maximum crop output by analysing real-time data from sensors on automated ATVs[203].

[196] Petersen, R., Weisse, M., Lewandowski, E., Swartz, T., and Wang, L., (2018). *Artificial Intelligence Helps Distinguish the Forest From the Trees: Part 1*. [online] Available at: https://www.globalforestwatch.org/blog/data/artificial-intelligence-helps-distinguish-the-forest-from-the-trees-part-1/?utm_campaign=gfw&utm_source=gfwblog&utm_medium=hyperlink&utm_term=orbitalinsights_11_2018 [Accessed on 03 November 2024].

[197] Rainforest Connection, (n/d). *Guardian Platform*. [online] Available at: https://rfcx.org/guardian [Accessed on 03 November 2024]

[198] WildMe, (n/d). [online] Available at: https://www.wildme.org/ [Accessed on 03 November 2024]

[199] Siemens, (n/d). *Leveraging Artificial Intelligence in Sustainable Digital Enterprise*. [online] Available at: https://www.siemens.com/global/en/company/insights/leveraging-artificial-intelligence-in-a-sustainable-digital-enterprise.html [Accessed on 03 November 2024]

[200] Rush, C., (2021). *Smart Building Solutions for Manufacturing Facilities.* [online] Available at: https://knowhow.distrelec.com/energy-and-power/smart-building-solutions-for-manufacturing-facilities/ [Accessed on 03 November 2024]

[201] Wood., J., (2023). *This AI helps buildings cool themselves and cut emissions.* [online] Available at: https://www.weforum.org/stories/2023/10/ai-buildings-heating-cooling-carbon-brainbox/ [Accessed on 03 November 2024]

[202] Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]

[203] Zhao, W., Li, T., Qi, B., Nie, Q. and Runge, T., (2021). *Terrain Analytics for Precision Agriculture with Automated Vehicle Sensors and Data Fusion*. *Sustainability*, 13(5), p.2905. [online] Available at: https://www.mdpi.com/2071-1050/13/5/2905 [Accessed on 03 November 2024]

The integration of AI and ML in agriculture, nonetheless, introduces challenges, particularly related to data privacy and security, as well as scalability and accessibility. With the reliance of AI on vast agricultural datasets, safeguarding data privacy becomes critical. Agricultural systems are increasingly vulnerable to cyber threats and data breaches, highlighting the need to protect data ownership, sharing, and access rights[204]. Additionally, the scalability and accessibility of AI-powered solutions remain constrained, especially for smaller farms and varying geographic regions[205]. Cost, infrastructure needs, patents, and usability are major barriers to widespread adoption. To make these technologies viable across different farm sizes and contexts, it is essential to address these challenges, ensuring that AI-driven agricultural tools are both affordable and adaptable to diverse agricultural environments.

Urban Sustainability and Resilient City Planning

With urban areas accounting for a large portion of global carbon emissions[206], AI is becoming essential for creating sustainable and resilient cities. AI-driven urban planning applications optimise infrastructure design, reducing resource consumption and promoting sustainable urban growth. For instance, ML models can simulate urban heat islands, informing the design of green spaces and cooling systems to mitigate heat stress in densely populated areas. Platforms like UrbanistAI[207] exemplify this approach by enabling human-AI collaboration to reimagine urban futures, facilitating participatory planning and co-design to create more sustainable and inclusive urban environments.

AI also supports resilient city planning by optimising transportation systems to lower greenhouse gas emissions. Intelligent transportation systems, which use AI to analyse traffic patterns and optimise routes, reduce fuel consumption and promote the use of low-emission vehicles[208]. Additionally, as discussed above, AI aids in energy-efficient building design and retrofitting, analysing architectural data to propose design modifications that improve thermal insulation and reduce energy use. In water and waste management, AI applications monitor and predict usage patterns, ensuring sustainable resource use within urban environments. An example of AI-driven urban sustainability is Amsterdam's use of the Object Detection Kit (ODK), an open-source framework developed to collect and analyse real-time street-level imagery through computer vision[209]. The ODK allows city officials to detect and manage waste by identifying and cataloguing objects on city streets. Municipal vehicles equipped with the ODK app scan streets for misplaced garbage, reporting locations to appropriate services, which then dispatch collection teams along optimised routes. This system enables Amsterdam to efficiently collect and sort waste, advancing circularity goals by separating materials like wood, furniture, and cardboard for specialised recycling.

While AI applications extend into a variety of other climate action areas—such as optimising carbon capture processes—this section has focused on four key domains to provide a concise overview. By honing in on Climate Modeling and Prediction, Monitoring and Conservation, Optimization in Energy and Resource Management, and Urban Sustainability, we present a broad view of AI's potentially impactful roles within climate action. These selected areas illustrate the primary ways AI can drive meaningful advancements toward SDG 13. In the following section, we delve into a specific case study: Global Forest Watch (GFW), a leading example of AI in environmental monitoring and conservation, examining its methodologies, achievements, and contributions to climate action.

3.3.2 | *Case Study: Climate TRACE*

As the urgency to combat climate change intensifies, accurate monitoring of greenhouse gas emissions has become paramount. Traditional methods of emissions reporting, often based on self-reported data by countries and industries, have been criticised for their lack of transparency and timeliness[210]. In response to this challenge, Climate TRACE[211] (Tracking Real-time Atmospheric Carbon Emissions) emerged as a groundbreaking initiative aiming to revolutionise how the world tracks emissions. Launched in 2020, Climate TRACE represents a coalition of organisations leveraging advanced technologies, particularly AI, to provide independent, near-real-time data on global greenhouse gas emissions.

The concept of Climate TRACE was born from the recognition that timely and accurate emissions data are essential for effective climate action. Historically, emissions data have been reported with significant delays—sometimes up to two years—hindering the ability of policymakers, researchers, and activists to make informed decisions. The founders of Climate TRACE envisioned a system that could offer transparent and up-to-date emissions monitoring, thereby holding emitters accountable and accelerating progress towards international climate goals like those outlined in the Paris Agreement.

204 Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]
205 Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]
206 Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]
207 Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]
208 Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]

209 Sukel, M., (2021). *An introduction to Object Detection Kit*. [online] Available at: https://amsterdamintelligence.com/posts/an-introduction-to-object-detection-kit [Accessed on 03 November 2024]
210 National Academies of Sciences, Engineering, and Medicine (2022: 53) *Advancing Climate Action: Insights from AI Technology*. Washington, D.C.: National Academies Press, Chapter 5. Available at: https://nap.nationalacademies.org/read/26641/chapter/5 (Accessed: 05 November 2024).
211 Climate TRACE, (n.d). [online] Available at: https://climatetrace.org/ [Accessed on 05 November 2024]

Climate TRACE operates by aggregating a vast array of data sources, including satellite imagery, remote sensing technologies, and open-source information. In December 2023, Climate TRACE released its second facility-level inventory, covering more than 352 million emissions sources, with detailed metadata providing unprecedented insight into the facilities driving the climate crisis and how to help fix them[212]. For instance, satellites equipped with spectrometers can detect specific wavelengths of light absorbed by greenhouse gases like carbon dioxide and methane. This data is then processed using AI algorithms that can identify emission sources, quantify emission levels, and even detect unreported or underreported emissions. As the coalition notes, "in 2024, Climate TRACE will begin releasing updated inventories multiple times per year, including increasingly up-to-date estimated emissions; even closer to 100% global coverage of all assets; and additional detailed information such as crosswalks with other datasets." (see https://climatetrace.org/about).

The use of AI is central to Climate TRACE's methodology. Machine learning models are trained to analyse patterns in the data, differentiating between various emission sources such as power plants, industrial facilities, and transportation networks. For example, AI algorithms can process thermal infrared imagery to detect heat signatures from power plants, correlating them with emission estimates based on known operational efficiencies and fuel types. This level of analysis allows Climate TRACE to provide detailed emissions inventories without relying on self-reported data, which can be incomplete or inaccurate.

An illustrative example of Climate TRACE's impact can be seen in the monitoring of methane emissions from both oil and gas operations and intensive farming practices. Methane is a potent greenhouse gas with a global warming potential significantly higher than carbon dioxide over a 20-year period. Traditional reporting methods often overlook fugitive methane emissions due to their diffuse nature and the difficulty of detection. In the agricultural sector, emissions from feedlots have been particularly hard to measure due to insufficient data sources. Climate TRACE addresses this challenge by utilising satellite imagery and AI to provide a clearer picture of emissions from feedlots. For instance, in the province of Jalisco, Mexico, satellite data revealed that feedlots are primarily concentrated in the northeastern region. Climate TRACE identified a total of 36 feedlots—24 beef, five dairy, and seven that may be a mix of both—housing approximately 200,000 head of cattle. Considering that a single cow typically produces about 99 kilograms of methane annually through enteric fermentation and manure, these feedlots collectively emit roughly 19.8 million kilograms of methane per year. This significant amount underscores the importance of accurately monitoring agricultural methane emissions[213].

Box 3.2 Leveraging Climate TRACE for Regional Emissions Insights in Abruzzo (Italy)

With agriculture and industry as its main economic sectors, Abruzzo has committed to significant greenhouse gas reduction targets in alignment with the European Union's goals. As a member of the Under2 Coalition and a signatory of the Global Covenant of Mayors for Climate and Energy, the region aims to reduce emissions by 40% by 2030 and achieve net-zero emissions by 2050[214].

Before integrating Climate TRACE data, Abruzzo faced difficulties in obtaining consistent and comprehensive emissions information. The regional authorities relied on outdated inventories—the latest being from 2012—and struggled with inconsistent data collected from various local sources. This lack of up-to-date and reliable data hindered the region's ability to formulate effective climate strategies and policies. By joining the STARRS (SpatioTemporal Asset Resolution for Sustainability) project, Abruzzo sought to overcome these challenges. Climate TRACE provided the region with recent and detailed emissions data, enabling a more accurate understanding of their current emissions landscape. Climate TRACE's methodology addressed the region's prior difficulties by offering verified and comparable data, providing consistent datasets that are easy to interpret and compare over time. This comprehensive coverage includes sectors and emission sources previously hard to measure, such as detailed transport emissions and land-use changes, thus overcoming significant data collection challenges.

The 2021 data[215] provided by Climate TRACE revealed that Abruzzo's emissions were estimated at 2.4 million metric tonnes of CO_2 equivalent, excluding sectors not currently tracked by Climate TRACE and land-use changes. Major emission sources included cement manufacturing, road transportation, and electricity generation. Cement manufacturing emerged as one of the largest contributors, with emissions fluctuating due to industry activity levels. A significant drop was observed in 2020 because of the COVID-19 pandemic, followed by a return to pre-pandemic levels in 2021. Road transportation was another major source of emissions, particularly in urban areas. Notably, the Greater Metropolitan Area of Pescara alone accounted for nearly 40% of urban road transport emissions in the region. Electricity generation also contributed significantly to Abruzzo's emissions profile.

Between 2015 and 2021, Abruzzo experienced a net gain in live forest and grassland biomass carbon. This increase resulted in the sequestration of approximately 300,000 tonnes of carbon, contributing to a reduction in overall emissions. Most of the biomass regrowth occurred in the central and western parts of the region, highlighting areas where conservation efforts may be particularly effective.

The detailed data provided by Climate TRACE has equipped Abruzzo's government to develop targeted policies focusing on high-emitting sectors and areas, such as implementing zero-emission zones in cities with the highest transportation emissions. Using recent data, the region can update climate strategies by refining regional programming tools to ensure they are based on current emissions realities. Additionally, sharing consistent and understandable data with municipalities and other stakeholders fosters collaboration in emissions reduction efforts, engaging stakeholders more effectively.

With access to up-to-date emissions data, Abruzzo is better positioned to meet its climate commitments. The region can now implement more effective medium- and long-term programming at both local and regional levels, contributing to Italy's national climate goals and the broader objectives of SDG 13 (Climate Action).

One key feature of Climate TRACE's work is its ongoing adaptation and refinement of methodologies, reflecting the diverse and dynamic nature of global emissions. The initiative continuously incorporates new data sources and enhances its AI models to improve accuracy and coverage. For example, integrating

[212] Climate TRACE, (n/d). Climate Trace - About the Coalition. [online] Available at: https://climatetrace.org/about [Accessed on 05 November 2024].
[213] Climate TRACE, (n/d: 40). A view from space: tracking emissions state by state. [pdf] Available at: https://climatetrace.org/api/download?file=starrs-2023-03-en [Accessed on 05 November 2024]

[214] Regione Abruzzo, (2020). *Initiatives and regional planning.* [online] Available at: https://www.regione.abruzzo.it/content/initiatives-and-regional-planning [Accessed on 05 November 2024]
[215] Climate TRACE, (n/d:. 15-23). A view from space: tracking emissions state by state. [pdf] Available at: https://climatetrace.org/api/download?file=starrs-2023-03-en [Accessed on 05 November 2024]

data from the European Space Agency's Sentinel satellites has expanded the spatial resolution of emissions monitoring, enabling more precise detection of emission hotspots. Additionally, Climate TRACE collaborates with local organisations to validate its findings and incorporate ground-based measurements, ensuring that its models remain robust across different contexts. Ultimately, rather than relying solely on AI or placing it at the centre of its approach, Climate TRACE utilises AI as one of several complementary tools within a diverse set of methodologies. This balanced integration of technology addresses the critical need for accurate and timely data, which is essential for informed decision-making and effective climate action.

Despite its innovative approach, Climate TRACE faces several challenges and criticisms[216]. It is important to note that *no independent peer-reviewed assessment has been made* of the recently released a database of global powerplant CO2 emissions at the facility-scale, which uses both AI and non-AI estimation approaches (as of October 2024, Gurney et al., 2024). This lack of external validation raises considerations about the accuracy and reliability of the data provided by Climate TRACE, highlighting the need for independent verification to strengthen confidence in its findings. While AI models can process vast amounts of data, they are not infallible. Factors such as cloud cover, satellite sensor limitations, and the complexity of disentangling overlapping emission sources can introduce uncertainties. Climate TRACE addresses these issues by employing cross-validation techniques, using multiple data sources and models to corroborate findings. However, the reliance on remote sensing data means that some small-scale or indoor emissions may still be overlooked. Another challenge is the potential for geopolitical tensions arising from independent emissions reporting. Countries or industries identified as significant emitters may dispute the findings, leading to conflicts over data sovereignty and the legitimacy of external monitoring. Climate TRACE advocates for transparency and collaboration, emphasising that its goal is to support global climate action rather than assign blame. By making its methodologies and data publicly available, Climate TRACE seeks to foster trust and encourage collective efforts to reduce emissions. Furthermore, there are technical and logistical hurdles associated with processing and storing the immense volumes of data required for global emissions monitoring. Climate TRACE leverages cloud computing and partnerships with technology companies to manage these demands, but scalability remains a critical concern. Ensuring that the system can handle increasing data inputs while maintaining rapid processing times is an ongoing area of development.

In terms of its contribution to SDG 13l 13 (Climate Action), Climate TRACE plays a pivotal role by enhancing the transparency and accountability of emissions reporting. Accurate data empowers policymakers to set realistic targets, track progress, and implement effective mitigation strategies. For example, a country may use Climate TRACE data to identify high-emitting sectors that require regulatory attention or investment in cleaner technologies. International bodies can also leverage this information to assess compliance with climate agreements and facilitate support for countries struggling to meet their commitments.

[216] Whitmee, S., Anton, B., & Haines, A., (2023). *Accountability for carbon emissions and health equity*. *The Lancet Planetary Health*, 7(1), pp. e20–e26. [pdf[Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9874368/pdf/BLT.22.289452.pdf [Accessed on 05 November 2024].

Moreover, Climate TRACE's approach has implications for scalability and replicability. By utilising open-source data and sharing its methodologies, the initiative encourages other organisations and countries to adopt similar practices. This collaborative model promotes the widespread adoption of advanced monitoring techniques, amplifying the global capacity to address climate change. The use of AI and remote sensing technologies in environmental monitoring is not limited to emissions; it can be extended to deforestation tracking, biodiversity assessments, and disaster response, showcasing the versatility of these tools in advancing sustainable development.

Despite the promise of Climate TRACE, it is essential to acknowledge the limitations and risks associated with its reliance on AI and technology. Algorithmic and estimation biases, data privacy concerns, and the digital divide between countries with varying technological capacities can affect the equitable distribution of benefits. There is a risk that nations or communities lacking access to advanced technologies may be left behind or unfairly scrutinised due to data gaps. Moreover, as Gurney et al. (2024: 7) argue "while the approach combining ML and high-resolution satellite imagery remains a promising technique, it was applied to a small minority of the individual powerplant facilities, [and] this was due to understandable limitations in the remote sensing information needed to actuate the AI method."

Although Climate TRACE represents a significant advancement in the global effort to combat climate change by providing an independent system for monitoring greenhouse gas emissions, it must navigate these ethical considerations by promoting inclusivity, capacity-building, and transparency in its operations. While challenges exist, the initiative represents a valuable asset in efforts to meet SDG 13 (Climate Action) targets and fostering transnational cooperation for a sustainable future.

CHAPTER 4
Critiquing AI

This chapter provides a critical examination of AI across five key areas of societal impact: corporate influence on digital development; bias, discrimination, and AI opacity; surveillance and privacy; environmental sustainability; and the new forms of digital colonialism. While some of these critiques have been addressed transversely across previous chapters, here we aim to provide a more concise overview of these issues. By dissecting these central themes, this chapter seeks to highlight how AI technologies are raising pressing ethical, environmental, and human rights concerns. Through an exploration of prominent debates, this chapter aims to provide a brief overview of some of the main critiques, challenges, and ethical concerns inherent in AI's rapid advancement.

SECTION 4.1
AI and Corporate Capture of Digital Development

While previous chapters have touched on these critiques (see, in particular, Chapter 2, Section 2.2: 'AI for Good—An Empty Signifier?'), this section explores briefly how corporate interests are capturing digital development on behalf of 'AI for Good' discourse.

Large technology firms, often called 'Big Tech,' dominate AI research and deployment due to their substantial resources, which allow them to attract top talent and acquire innovative startups[217]. As Iazzolino and Stremlau (2023) argue, these firms co-opt narratives of 'AI for social good' to justify this

[217] Zuboff, S., 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

dominance, using it as a tool to expand their influence in global development, often at the expense of local needs and contexts. Mazzucato and Collington (2023) expand this critique, highlighting how the over-reliance on private expertise weakens public institutions and warps economic priorities, with governments often dependent on corporate technologies and expertise. In the context of AI, this dependence can hinder effective regulation and shift policies to favour corporate profits over public welfare[218].

The corporate capture of AI can stifle innovation by prioritising proprietary technologies over open-source alternatives, limiting collaboration and knowledge sharing[219]. This environment marginalises smaller entities, including startups and academic institutions, hindering diverse contributions to AI development. The focus on profit-driven models often overlooks the needs of underserved communities, exacerbating digital divides[220]. This issue is particularly problematic in the Global South, where AI solutions from the Global North are often ill-suited to local contexts, leading to ineffective or harmful outcomes[221]. Heeks (2002) previously described this recurring issue as a 'design-reality gap,'[222] where imported technologies fail to align with local needs, often exacerbating digital divides rather than bridging them. In the context of AI, we see a similar dynamic: profit-driven, "one-size-fits-all" AI solutions are developed in the Global North and then 'transported' in the Global South without sufficient adaptation to the local context. This approach risks perpetuating digital divides and replicating previous ICT4D failures, as technologies designed for Northern markets may not only fail to address the unique challenges of Southern communities but also undermine local capacities and autonomy.

IBM's 'Project Lucy,' launched in 2014 through its Nairobi research lab, exemplifies this problematic dynamic. Aimed at adapting IBM Watson's AI for African contexts, the project faced immediate challenges that reflected the 'design-reality gap' described by Heeks. With its ambitious goals, Project Lucy quickly ran into challenges, particularly in reconciling IBM's corporate social responsibility (CSR) intentions with its commercial objectives in Kenya and South Africa. For instance, Iazzolino and Stremlau (2023) recount how IBM researchers tried to address school attendance data gaps by developing facial recognition technology to track student attendance. However, even before the pilot was launched, the project faced significant resistance. School leadership raised concerns due to the impact of biometric scanning on funding, which was linked to enrollment numbers, and students viewed the biometric approach as

[218] Frost (2024) expands on these concerns, arguing that algorithmic decision-making processes often lack sufficient public transparency and accountability, which can undermine the public's ability to oversee or challenge corporate influence in decision-making. See Frost, N., (2024). *The Impoverished Publicness of Algorithmic Decision Making*, *Oxford Journal of Legal Studies*. [doi] Available at: https://doi.org/10.1093/ojls/gqae027 [Accessed on 08 November 2024]
[219] Morozov, E., (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
[220] Couldry, N. & Mejias, U.A., (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.
[221] Iazzolino, G., & Stremlau, N. (2024). AI for social good and the corporate capture of global development. *Information Technology for Development*, *30*(4), 626–643. [doi] Available at: https://doi.org/10.1080/02681102.2023.2299351 [Accessed on 08 November 2024]
[222] Heeks, R., (2002). *Information systems in developing countries: Failure, success and local improvisations. The Information Society*, 18(2), 101-112. [doi] Available at: https://doi.org/10.1080/01972240290075039 [Accessed on 08 November 2024]

invasive surveillance. Technical issues further undermined the technology's efficacy, and the project ultimately fell short of its goals. Despite its early demise, Project Lucy exemplifies what Iazzolino and Stremlau describe as a growing trend in which tech corporations use data-driven projects not only to pursue CSR goals but also to build connections with policymakers, showcase AI's potential for African business sectors like business process outsourcing (BPO), agribusiness, and finance, and test predictive models for future commercial use.

In the end, the corporate capture of development under the banner of "AI for Good" risks replicating familiar development pitfalls: imposing solutions that serve corporate interests while sidelining local realities. As AI continues to expand globally, meaningful progress requires not just technological advancement but a commitment to aligning innovations with the unique needs and agency of the communities they aim to serve, in line with the *"Leave No One Behind"* agenda of the SDGs. Without this shift, the promises of "AI for good" will remain largely unfulfilled, leaving the Global South as a testing ground rather than a true partner in digital development.

SECTION 4.2

AI Bias, Discrimination and Opacity

In short, AI can be considered statistical pattern matching systems. As such AI systems learn patterns from data; therefore, they are susceptible to inheriting biases present in the datasets used for training. Biases can occur due to unrepresentative data samples, historical prejudices embedded in the data, or the way data is collected and labelled[223]. For instance, if an AI system is trained on a dataset that underrepresents women or certain ethnic groups, it may perform poorly for those populations, perpetuating inequality[224]. Additionally, a study[225] examining algorithmic ad delivery found that job ads promoting careers in Science, Technology, Engineering, and Math (STEM) fields, intended to be gender-neutral, were shown to fewer women than men. This occurred because younger women, a highly sought demographic, are more costly to reach with ads, resulting in apparent gender-based discrimination when the algorithm prioritised cost-efficiency.

In the context of three SDGs discussed in this chapter—SDGs 4, 10, and 13—AI biases and discrimination may have a devastating impact on people's lives. For example, in the context of SDG 4 (Quality Education),

[223] For an in-depth exploration of these issues, consult: Prabhu, V. U. & Birhane, A., (2020). *Large image datasets: A pyrrhic win for computer vision?* [pdf] Available at: https://arxiv.org/abs/2006.16923 [Accessed on 08 November 2024]; Birhane, A., Prabhu, V. U. & Kahembwe, E., (2021). *Multimodal datasets: misogyny, pornography, and malignant stereotypes.* [pdf] Available at: https://arxiv.org/abs/2110.01963 [Accessed on 08 November 2024]; Birhane, A., Prabhu, V., Han, S. & Boddeti, V. N., (2023). *On Hate Scaling Laws For Data-Swamps*. [pdf] Available at: https://arxiv.org/abs/2306.13141[Accessed on 08 November 2024]; Birhane, A., Prabhu, V., Han, S., Boddeti, V. N. & Luccioni, A. S., (2023). *Into the LAIONs Den: Investigating Hate in Multimodal Datasets*. [pdf] Available at: https://arxiv.org/abs/2311.03449; [Accessed on 08 November 2024]; Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A., (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. [pdf] Available at: https://arxiv.org/abs/1607.06520 [Accessed on 08 November 2024].
[224] Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. *Proceedings of Machine Learning Research*, 81, 1-15. [pdf] Available at: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf [Accessed on 08 November 2024].
[225] Lambrecht, A., & Tucker, C., (2019). *Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads*. *Management Science*, 65(7), pp. 2966-2981. [pdf] Available at: https://doi.org/10.1287/mnsc.2018.3093 [Accessed on 08 November 2024].

contrary to the view that technology is neutral, research increasingly reveals that AI systems inherit biases from their human developers and the data used in training. For example, a University of Washington study[226] found that AI models often perpetuate gender stereotypes in language translation tasks, misrepresenting gender-neutral pronouns from low-resource languages such as Bengali and Farsi. These biases not only reflect but also reinforce societal inequalities, particularly in contexts where AI tools influence decisions on hiring, media representation, and access to resources. Regarding SDG 10 (Reduced Inequality), AI-driven financial services may discriminate against minority groups if credit scoring algorithms are trained on biassed financial data[227]. And, for SDG 13 (Climate Action), biassed data in environmental monitoring can lead to inequitable resource allocation, affecting vulnerable communities disproportionately[228].

Biassed AI systems can have profound negative impacts on marginalised communities by amplifying existing social and economic inequalities. For example, predictive policing algorithms have been criticised for targeting minority neighbourhoods, leading to over-policing and the criminalization of certain communities. Such practices not only violate human rights but also hinder progress towards achieving equitable development outcomes[229][230].

Many AI models, especially deep learning algorithms, function as "black boxes" where the decision-making processes are not transparent or understandable to users or those affected by the decisions[231][232]. This opacity poses significant challenges for accountability and trust, particularly in international development contexts where AI decisions can affect large populations. The lack of transparency makes it difficult to identify and correct biases, undermining efforts to ensure fair and equitable outcomes[233]. In international development, the opacity of AI systems can hinder stakeholders from understanding how resources are allocated or how decisions are made, potentially leading to mistrust and resistance to AI

[226] David, J., (2023). *Students investigate how artificial intelligence perpetuates biases*. Information School, University of Washington. [online] Available at: https://ischool.uw.edu/news/2023/10/students-investigate-how-artificial-intelligence-perpetuates-biases [Accessed on 08 November 2024].

[227] Hurley, M., & Adebayo, J. (2016). *Credit scoring in the era of big data. Yale Journal of Law and Technology*, 18(1), 148-216. [pdf] Available at: https://openyls.law.yale.edu/bitstream/handle/20.500.13051/7808/Hurley_Mikella.pdf [Accessed on 08 November 2024].

[228] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233. [online] Available at: https://www.nature.com/articles/s41467-019-14108-y [Accessed on 08 November 2024].

[229] Angwin, J., Larson, J., Mattu, S., and Kirchner, L., (2016). *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*. [online] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed on 08 November 2024].

[230] Koepke, L., (2020). *Predictive policing algorithms are racist. They need to be dismantled*. [online] Available at: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ [Accessed on 08 November 2024].

[231] Lipton, Z. C. (2018). *The Mythos of Model Interpretability*. *Communications of the ACM*, 61(10), 36-43. [pdf] Available at: https://dl.acm.org/doi/pdf/10.1145/3233231 [Accessed on 08 November 2024].

[232] Yampolskiy, R. V., (2019). *Unexplainability and Incomprehensibility of Artificial Intelligence*. Computer Engineering and Computer Science, University of Louisville. [pdf] Available at: https://arxiv.org/abs/1907.03869 [Accessed on 08 November 2024].

[233] Pasquale, F., 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

interventions. In humanitarian contexts, this lack of transparency and explainability is particularly concerning, as it can obscure the causal link between triggering events and the delivery of critical services to vulnerable populations. This phenomenon, often referred to as the "black box effect," describes how the complexity and technical opacity of AI algorithms can hide their inner workings, making it challenging to hold these systems accountable. As Coppi, Jimenez, and Kyriazi (2021) argue, without clear, explicable processes in place, AI-driven decision-making can inadvertently harm those it aims to assist, undermining key humanitarian principles such as neutrality, impartiality, and the commitment to "do no harm."[234].

While addressing AI bias, discrimination, and opacity remains a significant challenge, doing so is essential not only for technical accuracy but also for upholding human rights and equity in AI applications, particularly as AI becomes increasingly embedded in global development and humanitarian initiatives.

SECTION 4.3

AI Surveillance, Privacy and Security

AI has significantly advanced surveillance technologies, enabling unprecedented capabilities in monitoring and data collection. AI-driven surveillance systems can process vast amounts of data in real-time, including facial recognition, behavioural analysis, and predictive analytics[235]. While these technologies promise to enhance security, they pose serious risks to privacy rights and civil liberties.

The integration of AI into surveillance has led to concerns about a pervasive "surveillance society," where individuals are constantly monitored without their explicit consent (Zuboff, 2019). This constant monitoring can infringe upon several fundamental rights, most notably the right to privacy and the rights to freedom of assembly and association—key human rights recognized in the Universal Declaration of Human Rights. Additionally, the right to freedom of expression may be compromised, as individuals may self-censor their speech or activities if they believe they are under surveillance. The right to non-discrimination is also at risk, as biases in AI systems often result in disproportionate surveillance of marginalised groups, potentially leading to unequal treatment. Furthermore, surveillance can restrict the right to freedom of movement by monitoring or flagging individuals in certain areas, thereby limiting their freedom to move without oversight. The right to a fair trial can be impacted if biassed surveillance data is used in legal proceedings, jeopardising due process. Lastly, AI's capacity—particularly in multimodal AI systems—to analyse and interpret vast amounts of personal data intensifies concerns over the erosion of personal autonomy and human dignity, as individuals feel less able to make decisions free from monitoring or scrutiny. These rights are crucial for personal freedom and are increasingly threatened by the intrusive capabilities of AI-enhanced surveillance.

A recent case in Trento, Italy, illustrates the challenges associated with multimodal AI surveillance systems and their potential infringements on rights. The Trento municipality participated in the EU-funded Marvel and Protector projects, which aimed to enhance urban security through advanced AI surveillance. These

[234] Coppi, G., Jimenez, R. M., & Kyriazi, S., 2021. *Explicability of Humanitarian AI: A Matter of Principles*. Journal of International Humanitarian Action, 6(1), p. 11. [pdf] Available at: https://jhumanitarianaction.springeropen.com/articles/10.1186/s41018-021-00096-6 [Accessed on 08 November 2024].

[235] van Brakel, R. (2017). *Big Data Surveillance: The Case of Policing. Surveillance & Society*, 14(1), 1–15. [pdf] Available at: https://www.asanet.org/wp-content/uploads/attach/journals/oct17asrfeature.pdf [Accessed on 08 November 2024].

projects used AI to analyse audio and video data from public surveillance cameras to detect potential security threats. However, Italy's data protection authority found that the anonymization techniques used were insufficient, as citizens' identities could still potentially be inferred from the data collected. The projects were subsequently fined, with regulators citing significant risks to privacy and the potential to alter public behaviour, highlighting the profound ethical implications of using multimodal AI for public surveillance[236].

Governments and corporations are increasingly deploying AI-powered surveillance tools for various purposes, ranging from national security to targeted marketing. Governments utilise AI for mass surveillance to monitor populations, often justified by the need to prevent crime or terrorism[237]. For example, in a trial by South Wales Police, facial recognition technology was found to be only 19% accurate, misidentifying innocent people at a high rate[238]. Additionally, studies showed that the system disproportionately misidentified people from Black and other minority ethnic backgrounds, raising concerns about racial bias and the profiling of individuals with protected characteristics. Whereas corporations employ AI to collect and analyse consumer data to tailor advertisements and influence purchasing behaviour[239]. This data collection often occurs without informed consent.

AI systems, while powerful, are vulnerable to hacking and misuse. Security flaws can be exploited by malicious actors to gain unauthorised access to sensitive data or to manipulate AI behaviour. For instance, "prompt injection" attacks can manipulate AI language models into disclosing confidential information or performing unintended actions. Recent research has demonstrated that prompt injection attacks can compromise real-world LLM-integrated applications, leading to data theft and unauthorised actions[240][241]. Additionally, a recent example highlights other types of vulnerabilities in AI-integrated devices: in October 2024, two Harvard students, demonstrated how Meta's Ray-Ban smart glasses could be exploited to access individuals' personal information without their consent[242]. The glasses, equipped with AI features like voice assistants and real-time data processing, were intended to enhance user experience. However, the students discovered that by manipulating certain features, they could hijack the device to record conversations and collect personal data surreptitiously. This incident underscores the potential for AI devices to be repurposed for unauthorised surveillance and the importance of robust security measures. Equipped with AI capabilities, these glasses were intended to enhance user experience through features like voice assistance and real-time data processing. However, the students discovered they could use the glasses, combined with additional software, to identify people on campus and extract sensitive details such as names, home addresses, phone numbers, and even relatives' names. Their process involved livestreaming to Instagram, where they used facial recognition software to detect faces in the stream and connected these to a reverse image search engine, PimEyes. This then surfaced links with the identified person's images, and a large language model was used to extract personal data from these sources. This project, dubbed "I-XRAY," was created as a public service announcement to demonstrate the privacy risks associated with large language models and AI-enhanced devices, which can scrape and analyse data on a large scale.

The widespread use of AI surveillance has significant implications for SDGs, particularly SDG 4 (Quality Education) and SDG 10 (Reduced Inequalities). In educational settings, increased surveillance can impact students' privacy and freedom of expression, ultimately affecting the quality of education and the openness of the learning environment[243][244]. AI-driven monitoring tools in schools can lead to self-censorship and discomfort among students and staff, creating environments that may hinder educational engagement and critical discussion. Further, AI surveillance technologies often disproportionately target marginalised communities, thus exacerbating social inequalities—relevant to SDG 10 (Reduced Inequalities). Biassed AI algorithms can lead to increased surveillance of specific ethnic or socio-economic groups, reinforcing discriminatory practices and amplifying existing societal biases (see Benjamin, 2019).

As AI-driven surveillance becomes more integrated into public and private spaces, its implications for human rights, democracy, and security demand even closer and more rigorous examination, and especially when deployed in high-vulnerable, low-rights contexts.

SECTION 4.4

AI and Environmental Impact

The global distribution of AI infrastructure is heavily influenced by political and economic factors, resulting in uneven geographical development. Data centres—the backbone of AI computations—are often situated in regions with favourable tax laws, inexpensive electricity, and lenient environmental

[236] Redazione Trento, (2024). *Progetti Marvel e Protector, il comune condannato a pagare 25 mila euro*. La Voce del Trentino. [online] Available at: https://www.lavocedeltrentino.it/2024/01/24/progetti-marvel-e-protector-il-comune-condannato-a-pagare-25-mila-euro/ [Accessed on 08 November 2024].

[237] Harwell, D., & Timberg, C. (2019). *Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Casts Doubt on Their Expanding Use*. [online] Available at: https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/ [Accessed on 08 November 2024]

[238] Liberty, (2020). *Legal challenge: Ed Bridges v South Wales Police*. Liberty. [online] Available at: https://www.libertyhumanrights.org.uk/issue/legal-challenge-ed-bridges-v-south-wales-police/ [Accessed on 08 November 2024].

[239] Cheney-Lippold, J. (2011). *A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control*. *Theory, Culture & Society*, 28(6), 164–181. [doi] Available at: https://doi.org/10.1177/0263276411424420 [Accessed on 08 November 2024]

[240] Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M., (2023). *Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection*. arXiv. [pdf] Available at: https://arxiv.org/abs/2302.12173 [Accessed on 08 November 2024]

[241] Lee, D. & Tiwari, M., 2023. *Prompt Infection: LLM-to-LLM Prompt Injection within Multi-Agent Systems*. [pdf] Available at: https://arxiv.org/html/2410.07283 [Accessed 12 Nov. 2024].

[242] Choo, L. (2024). *Meta's Ray-Ban Smart Glasses Exposed to Privacy Breach by Students*. Forbes. [online] Available at: https://www.forbes.com/sites/lindseychoo/2024/10/04/meta-ray-bans-ai-privacy-surveillance/ [Accessed on 08 November 2024]

[243] Williamson, B., & Hogan, A. (2020). *Commercialisation and Privatisation in/of Education in the Context of COVID-19. Education International*. [pdf] Available at: https://www.ei-ie.org/file/129 [Accessed on 08 November 2024]

[244] Lindh, M., & Nolin, J. (2016). *Information We Collect: Surveillance and Privacy in the Implementation of Google Apps for Education. European Educational Research Journal,* 15, 644-663. [doi] Available at: https://doi.org/10.1177/1474904116654917 [Accessed on 08 November 2024]

[245] Crawford, K., 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

The Cartography of Generative AI, a project by the Spanish collective Estampa, provides a comprehensive (visual) mapping of the global AI landscape, illustrating the concentration of AI resources in certain regions and the implications for global equity[246]. Estampa's work critically examines the imaginaries surrounding AI that lead to alienation and mystification, particularly as AI becomes more popularised and integrated into everyday life. They argue that as AI technologies consolidate, it becomes pertinent to map their connections with human activities and more-than-human territories to understand the extractions, agencies, and resources that enable the operation of AI tools.

The continuous demand for AI hardware accelerates resource depletion and has significant environmental and social impacts. Mining activities not only deplete finite resources but also cause significant ecological disruptions, including deforestation, loss of biodiversity, and contamination of water sources, affecting both ecosystems and human communities[247]. The production of AI hardware requires significant amounts of finite minerals and metals, such as cobalt, lithium, and rare earth elements. The extraction of these materials often involves environmentally damaging mining practices, leading to habitat destruction, soil erosion, and water pollution[248][249]. For instance, copper mining in Chile and Peru has led to conflicts over mining revenues and health problems due to water pollution (Estampa, 2023). Similarly, lithium extraction in Chile's Salar de Atacama threatens local communities' access to water and biodiversity.

AI models, especially deep learning algorithms, require vast computational resources. Training large-scale models consumes significant amounts of electricity, contributing to increased greenhouse gas emissions[250]. For example, training a single deep learning model can emit as much carbon dioxide as five cars during their entire lifespans. This high energy demand underscores the importance of sourcing electricity from renewable resources to mitigate environmental impacts. The surge in investment and applications in recent years has multiplied the power requirements of servers in data centres. A single data centre can consume the equivalent energy of 50,000 homes[251]. The growth of AI has significantly increased energy dependency, with dedicated AI servers consuming substantially more power than traditional servers. Monserrate (2022) argues that the electricity used by data centres is estimated to account for 0.3% of total carbon emissions, and when personal connected devices are included, the total rises to 2% of global carbon emissions.

[246] Estampa, (2024). *Cartography of Generative AI.* [online] Available at: https://cartography-of-generative-ai.net/ [Accessed on 08 November 2024]
[247] Mazzucato, M. (2024). *The ugly truth behind ChatGPT: AI is guzzling resources at planet-eating rates. [online]* Available at: https://www.theguardian.com/commentisfree/article/2024/may/30/ugly-truth-ai-chatgpt-guzzling-resources-environment [Accessed on 08 November 2024].
[248] Gabrys, J. (2013). *Digital rubbish: A Natural History of Electronics.* Michigan, University of Michigan Press.
[249] Ali, S.H., Giurco, D., Arndt, N., Nickless, E., Brown, G., Demetriades, A., & Yakovleva, N. (2017). *Mineral supply for sustainable development requires resource governance*. *Nature*, 543(7645), 367–372. [doi] Available at: 10.1038/nature21359 [Accessed on 08 November 2024]
[250] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics. [pdf] Available at: https://aclanthology.org/P19-1355.pdf [Accessed on 08 November 2024]
[251] Gonzalez Monserrate, S. (2022). *The Cloud is Material: On the Environmental Impacts on Computation and Data Storage* [online] Available at: https://mit-serc.pubpub.org/pub/the-cloud-is-material/release/2 [Accessed on 08 November 2024].

Additionally, data centres consume large quantities of water for cooling purposes to prevent servers from overheating[252]. This water usage can strain local water supplies, particularly in areas prone to drought or with limited water resources. The expansion of data centres exacerbates water scarcity issues, prompting the emergence of activist movements that challenge the environmental footprint of these infrastructures and advocate for more sustainable practices[253]. For example, conflicts between local communities and technology companies are arising, such as protests by Montevideo residents against plans to build a

The environmental impact of AI technologies presents significant challenges to achieving SDG 13 (Climate Action), as AI infrastructure's energy intensity, resource demands, and waste byproducts contribute to a growing ecological footprint. Addressing these issues within the framework of SDG 13 requires urgent, multi-sectoral commitments to mitigate the environmental costs of AI through investment in renewable energy, sustainable resource management, and equitable distribution of resources. Yet, the very idea that organisations, governments, and private companies aim to achieve SDG 13 through AI is paradoxical, given AI's own substantial environmental footprint. While AI promises to aid in climate monitoring and sustainable practices, its heavy reliance on energy-intensive infrastructure and extractive industries raises a fundamental contradiction, challenging its alignment with genuine climate action goals. It's time for all stakeholders to move beyond a narrow, technology-determinist mindset and recognize the environment not as a limitless resource for innovation, but as a finite, common good—a shared foundation upon which all sustainable progress depends.

SECTION 4.5

AI, Neocolonialism, and Invisible Labour

AI colonialism refers to the practices by which data and resources are extracted from less powerful regions, communities, or groups to benefit more powerful entities, often without adequate compensation or consent.[254] This paradigm mirrors historical colonialism, where colonial powers exploited colonies for resources and labour. In the digital age, data—along with natural resources and labour—has become a valuable commodity, and its extraction raises ethical concerns regarding consent, ownership, and equitable benefit-sharing. A recent example is Worldcoin, a startup co-founded by Sam Altman, CEO of OpenAI, which promises a cryptocurrency-based universal basic income but instead has gathered biometric data from vulnerable populations in economically disadvantaged regions. Through limited disclosure and monetary incentives, the company collected iris scans and other personal data from individuals across countries such as Indonesia, Kenya, and Sudan.[255] Worldcoin's activities raise ethical

[252] Pasek, A., Vaughan, H., & Starosielski, N. (2023). *The world wide web of carbon: Toward a relational footprinting of information and communications technology's climate impacts*. Big Data & Society, 10(1). [doi] Available at: https://doi.org/10.1177/20539517231158994 [Accessed on 08 November 2024]
[253] Lehuedé, S. (2022). *Big tech's new headache: Data centre activism flourishes across the world. [online]* Available at: https://blogs.lse.ac.uk/medialse/2022/11/02/big-techs-new-headache-data-centre-activism-flourishes-across-the-world/ [Accessed 08 November 2024].
[254] Mohamed, S., Png, M.-T., & Isaac, W. (2020). *Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology*, 33(4), 659–684. [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s13347-020-00405-8.pdf [Accessed on 08 November 2024]
[255] Guo, E. & Renaldi, A., (2022). *Deception, exploited workers, and cash handouts: How Worldcoin recruited its first half a million test users.* [online] Available at: https://www.technologyreview.com [Accessed 09 November 2024].

concerns over informed consent, privacy, and exploitation, echoing historical colonial practices of extracting resources from the Global South without equitable compensation or transparency. This case reflects how, in the digital era, AI-driven enterprises continue to reinforce global power asymmetries, and establish new forms of colonial practices, by commodifying data and labour from less powerful communities.

Data from developing countries is often harvested to train AI systems without proper consent or compensation to the data subjects.[256] The asymmetrical power relations between data collectors and brokers (often corporations from developed countries) and data subjects (often individuals from developing countries, though not exclusively) can lead to a form of digital exploitation. [257] AI systems heavily rely on human labour for tasks like data annotation, content moderation, and algorithm training. This labour is often outsourced to workers in lower-income countries who work under precarious conditions for minimal wages[258]. These workers perform essential functions that enable AI technologies to operate effectively, yet their contributions remain largely invisible and undervalued in the global AI economy. For instance, one of the first large-scale deep learning datasets, ImageNet, consists of over 14 million labelled images, spanning more than 20,000 categories. This vast dataset was made possible by the efforts of thousands of anonymous workers recruited through Amazon's Mechanical Turk platform, which introduced 'crowdwork'—a practice in which large volumes of tasks are broken down and distributed among millions of people globally. These crowdworkers, who made ImageNet possible, received minimal payment, sometimes as little as a few cents per completed task[259]. Similarly, OpenAI, the company that developed ChatGPT, outsourced labour to Kenyan workers who earned less than $2 per hour to help reduce the toxicity of its AI chatbots[260]. This work entailed exposing these labourers to graphic and disturbing content to label harmful materials, often at a significant mental health cost.

AI technologies reinforce existing social hierarchies by perpetuating biases present in training data. For instance, facial recognition systems have demonstrated higher error rates for women and people of colour due to biassed datasets (see subchapter 4.2 above). Such disparities can lead to discriminatory practices in law enforcement, hiring, and access to services. Moreover, algorithmic decision-making in areas like credit scoring, employment, and healthcare can disproportionately disadvantage marginalised groups.[261][262] In a 2023 study[263] on large language models (LLMs), four popular models—Bard, ChatGPT, Claude, and GPT-4—were tested with questions about medical advice. The researchers posed nine questions to each model five times, yielding a total of 45 responses per model. The findings revealed that all models, including those from OpenAI, Anthropic, and Google, perpetuated race-based medicine stereotypes in their responses. For instance, GPT-4 falsely asserted that the "normal" lung function value for Black people is 10-15% lower than that of white people, a misconception rooted in outdated race-based medical practices. Additionally, the models showed inconsistency in their responses, indicating a lack of reliability and potential reinforcement of harmful stereotypes. Such examples underscore the potential for AI to reinforce outdated and inaccurate stereotypes, particularly in sensitive areas like healthcare, where biassed information can have serious consequences for marginalised groups.

Technology in general, and AI specifically, does not come into existence without an ideology; it is embedded within the layered apparatus of technology, institutions, and ideologies that influence AI at large[264]. AI has often been harnessed to advance neoliberal visions of society, endorse projects of dispossession and resource accumulation, and legitimise practices of mass incarceration and surveillance.[265][266] Historically, tools like 'IQ testing' served as early "psychopolitical dispositifs" (or mechanisms of social control) that helped naturalise racial and economic inequalities by framing them as scientifically grounded, inherent traits rather than socially constructed outcomes. Neoliberal thought—espoused by figures ranging from radical proponents like Murray Rothbard, Richard J. Herrnstein, and Charles Murray, to mainstream figures like Milton Friedman—emphasises this competitive, hierarchical view of society. Within this framework, IQ served as a way to "justify" inequalities as natural differences, supporting the neoliberal ideals of free markets, minimal state intervention, and the privatisation of public goods[267]. This linkage between IQ and social hierarchy has extended into the realm of AI, which now operates as a more technologically sophisticated and obscured extension of these early biases. Algorithms in AI systems are often designed to predict and classify people's capabilities, reproducing a social stratification that reinforces existing power structures. In this way, AI has become an even more opaque psychopolitical dispositif, lending a seemingly objective veneer to biases deeply rooted in data and social contexts.

[256] Birhane, A. (2021). *Algorithmic injustice: A relational ethics approach. Patterns*, 2(2), 100205. [doi] Available at: https://doi.org/10.1016/j.patter.2021.100205 [Accessed on 09 November 2024].
[257] Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
[258] Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
[259] Conger, K. & Gebeloff, R., (2019). *The Humans Working Behind the AI Curtain*. [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed 09 November 2024].
[260] Perrigo, B., (2023). *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. *[online]* Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed 09 November 2024].
[261] Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
[262] O'Neil, C., (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
[263] Omiye, J.A., Lester, J.C., Spichak, S., Rotemberg, V., & Daneshjou, R., 2023. *Large language models propagate race-based medicine*. [pdf] https://www.nature.com/articles/s41746-023-00939-z.pdf [Accessed on 08 November 2024].
[264] McQuillan, D., (2022). *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*. Bristol: Bristol University Press.
[265] Katz, Y., (2020). *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.
[266] As Pasquinelli (2023) explains, Friedrich Hayek's ideas on spontaneous order and market self-organization influenced the early development of connectionist AI, notably in the work of Frank Rosenblatt. Hayek's theories appropriated concepts of pattern recognition, reinterpreting them as neoliberal principles of market regulation, where order is seen to emerge naturally without central control.
[267] Winston, A. S. (2018). *Neoliberalism and IQ: Naturalizing economic and racial inequality.* Theory & Psychology, 28(5), 600-618. [doi] Available at: https://doi.org/10.1177/0959354318798160 [Accessed on 08 November 2024]

The impact of eugenics on AI is significant too. Key statistical concepts developed within the eugenics movement continue to influence AI system design and methodology. For instance, foundational statistical tools such as Francis Galton's concept of 'regression to the mean', Karl Pearson's 'correlation coefficient', and Ronald Fisher's 'maximum likelihood estimation' were initially developed to support eugenic research and have since become integral to statistical modelling in AI. Additionally, Charles Spearman's work on 'IQ[268] and General Intelligence (g-factor)' and the controversial arguments on intelligence and social hierarchy presented by Richard J. Herrnstein and Charles Murray in *The Bell Curve* have influenced how AI systems approach classification, prediction, and the ranking of human traits. These concepts are applied in various degrees in the algorithms, data analysis, and predictive models underlying AI systems today, often without full recognition of their origins and the biases they may perpetuate. Regression, correlation and the notion of general intelligence mark historical entanglements with eugenics that the field of AI has never fully faced up (McQuillan, 22: 91). In fact, eugenicist ideologies manifest still in contemporary AI applications that draw on reductive metrics and data-driven categorization, implicitly valuing certain traits over others and entrenching social inequalities. Gebru and Torres (2024) argue that many of the discriminatory attitudes that animated eugenicist thinking—such as racism, xenophobia, classism, ableism, and sexism—persist within the drive to create Artificial General Intelligence (AGI). These biases are often masked by language that emphasises "safety" and the purported goal of "benefiting humanity," which can serve to evade accountability while reinforcing systems that disproportionately harm marginalised communities and consolidate power among a few actors. This eugenics-informed approach within AI reflects a continuation of harmful ideologies that emphasise hierarchical classifications and exclusionary practices[269].

To conclude, AI neocolonialism interlinks with broader critiques of AI's role in reinforcing global inequities. The corporate capture of digital development (subchapter 4.1) fuels neocolonial practices, as powerful corporations from developed countries dominate AI deployment in the Global South, often at the expense of local communities. This dominance extends to perpetuating biases and discriminations (subchapter 4.2) embedded in AI systems that frequently disadvantage marginalised groups and people with protected characteristics worldwide, mirroring and reinforcing historic power imbalances. AI-driven surveillance (subchapter 4.3) also disproportionately affects vulnerable populations, infringing not only on privacy and autonomy but also on fundamental freedoms and the rights, much like colonial governance structures once did. Furthermore, the environmental impact of AI (subchapter 4.4) disproportionately strains regions that are already burdened by climate vulnerability, adding layers of ecological exploitation to the digital extraction economy. Together, these critiques reveal a pattern where AI technologies, under the guise of progress, risk consolidating global inequalities, highlighting an urgent need for an ethical recalibration that centres equity, accountability, solidarity, and shared benefit.

[268] It should be noted, however, that IQ has been developed originally by French psychologist Alfred Binet, in order to identify school children who needed additional support, and rejected the idea that it represented a fixed and innate attribute of the individual (McQuillan, 2022: 86).
[269] Gebru, T. & Torres, É.P., 2023. *The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence*. *First Monday*, 29(4). [doi] Available at: https://doi.org/10.5210/fm.v29i4.13636 [Accessed 09 November 2024].

CHAPTER 5
Recommendations

SECTION 5.1
Towards the Democratisation of Expertise and Policymaking in Times of AI

In the current climate of increasing scepticism towards democratic institutions and scientific expertise[270]—manifested in the denial of critical issues like the climate crisis—the democratisation of expertise and policymaking becomes even more crucial. This growing denialism can lead to misinformed policies, undermine public trust, and widen the gap between technological advancements and societal well-being[271]. The erosion of trust in independent science and democratic processes hampers collective efforts to address global challenges and may result in the misuse or unethical deployment of AI technologies. By fostering inclusive and participatory policymaking, we can bridge this divide, ensuring that AI governance is both ethically sound and aligned with societal values.

The democratisation of expertise involves expanding participation in policymaking processes to include a diverse array of stakeholders—ranging from scientists and technologists to civil society organisations, educators, and the general public. This inclusive approach aligns with the principles of deliberative democracy, fostering plural forms of knowledge creation and ensuring that policies reflect a broad spectrum of interests and perspectives. Alvin M. Weinberg (1972) introduced the concept of ***'trans-science'*** to address issues that, while framed in scientific terms, extend beyond empirical verification and enter the realm of societal values and ethics. He emphasised the importance of public participation in such matters:

> We scientists value our republic of science with its rigorous peer group review. The uninformed public is excluded from participation in the affairs of the republic of science rather as a matter of course. But when what we do transcends science and impinges on the public, we have no choice but to welcome public participation. Such participation by the uninitiated in matters that have both scientific and trans-scientific elements that may pose some threat to the integrity of the republic of science. To my mind, however, this is a lesser threat than is the threat to our democratic process that would be posed by excluding the public in trans-scientific debate.[272]

AI transcends the contours of pure science due to its profound impact on society. AI technologies influence various aspects of daily life, thereby embedding themselves deeply within social, ethical, and

[270] For a more nuanced view about different types of science scepticism, please see: Većkalov, B., van Stekelenburg, A., van Harreveld, F., & Rutjens, B. T. (2023). *Who Is Skeptical About Scientific Innovation? Examining Worldview Predictors of Artificial Intelligence, Nanotechnology, and Human Gene Editing Attitudes*. *Science Communication*, 45(3), 337-366. [doi] Available at: https://doi.org/10.1177/10755470231184203 [Accessed on 09 November 2024]; Rutjens, B. T., Sengupta, N., der Lee, R. van, van Koningsbruggen, G. M., Martens, J. P., Rabelo, A., & Sutton, R. M. (2022). *Science Skepticism Across 24 Countries*. *Social Psychological and Personality Science*, 13(1), 102-117. [doi] Available at: https://doi.org/10.1177/19485506211001329 [Accessed on 09 November 2024]
[271] Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). *Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era*. Journal of Applied Research in Memory and Cognition, 6(4), 353–369. [doi] Available at: https://doi.org/10.1016/j.jarmac.2017.07.008 [Accessed on 09 November 2024]
[272] Weinberg, M.A., 1972. *Science and Trans-Science*. Volume 177, Number 4045. American Association for the Advancement of Science.

political spheres. In Weinberg's terms, AI presents questions that cannot be answered solely through scientific methods because they inherently involve value judgments, ethical considerations, and societal norms.

In line with Weinberg's call for inclusion of the public in trans-scientific debate, Ivan Ilich (1973) proposed the concept of ***'counterfoil research'***, advocating for critical engagement with technology and public involvement in identifying and challenging potential harms. For Ilich, "counterfoil research [...] provides guidelines for detecting the incipient stages of murderous logic in a tool and involves the public by showing that the demands for freedom of any group or alliance can be identified with the implicit interest of all".[273] In the context of increasing denialism, Illich's ideas highlight the importance of collective scrutiny and the democratisation of scientific expertise—and technological design—to prevent technological developments from undermining societal well-being.

Perhaps even more relevant to the context of this research is the concept of Post-Normal Science (PNS), introduced by Funtowicz and Ravetz (1993), which is particularly significant in times "when facts are uncertain, values in dispute, stakes high and decisions urgent".[274] They call for the co-production of knowledge, dialogue-based policy processes and democratic participation. Central to this is what they call '***extended peer communities'***. As Funtowicz and Ravetz note:

> For [...] new problems, the maintenance of quality depends on open dialogue between all those affected. This we call an 'extended peer community', consisting not merely of persons with some form or other of institutional accreditation, but rather of all those with a desire to participate in the resolution of the issue. Since this context of science is one involving policy, we might see this extension of peer communities as analogous to earlier extensions of the franchise in other fields, such as women's suffrage and trade union rights. This is not merely a matter of extensions of liberty of individuals. With PNS we can guide the extension of the accountability of governments (the foundation of modern democratic society) to include the institutions involved in the governance of science and technology.[275]

An example of this approach to democratising policymaking in the field of emerging technologies is the recent initiative launched by the City of Turin, Italy. Under its "Casa delle Tecnologie Emergenti di Torino – CTE NEXT" initiative, Turin has established Italy's first *Board of Ethics for Emerging Technologies*. This board, acting as an advisory body to the City of Turin, aims to assess the impacts of emerging technologies, particularly AI, on municipal policies, ensuring alignment with both national and international directives, and address practical issues such as AI applications, algorithmic data aggregation, surveillance, profiling, privacy, and the digital rights of citizens.

In times of polycrisis marked by high uncertainty and distrust in the foundations of democratic society, such deliberative approaches—including those of feminist science (Haraway, 1988; Harding, 1998; Roy,

[273] Ilich, I. [1973], 2009 (p. 77-83). *Tools for Conviviality*. Marion Boyars Publishers Ltd
[274] Funtowicz, S.O., Ravetz, J.R., 1993. *Science for the Post-Normal Age.* Futures, 25[7], pp. 739-55. [doi] Available at: https://doi.org/10.1016/0016-3287(93)90022-L [Accessed on 08 November 2024]
[275] Funtowicz, S., Ravetz, J. 2003. *Post-Normal Science*. *International Society for Ecological Economics' Internet Encyclopaedia of Ecological Economics.*

2004)—in policymaking help to address imperfections in a fair and inclusive manner.[276] In other words, the policymaking and scientific communities should engage in continuous dialectical interaction where scientific expertise is democratised while simultaneously "expertising" democracy.[277]

SECTION 5.2

Recommendations

The aim of these recommendations is to support, guide, and facilitate policymakers and other stakeholders, particularly institutions such as the Ministry of Foreign Affairs of Italy (MAECI) and the Italian Agency for Development and Cooperation (AICS), in navigating the complexities of AI. The recommendations presented here emphasise the importance of ethical standards, participatory governance, and inter-epistemic collaboration to empower communities and address complex development challenges effectively.

1. **Promote, Develop and Adopt 'Ethics by Design'**
   - **Establish an Ethics by Design Framework**: Collaborate with international and national organisations, academia, and civil society to develop an AI ethical framework that integrates ethical considerations throughout the AI development lifecycle. This approach ensures that transparency, accountability, fairness, and respect for human rights are embedded from the inception of AI systems across various sectors. To make *Ethics by Design* an operational requirement, this framework should be adopted at all governmental levels, from ministries to public administrations, agencies, and local municipalities. Each level of government would implement this ethical framework in a structured, granular way, setting clear ethical standards and accountability measures within their respective jurisdictions. Ministries would establish high-level policies, while agencies and local municipalities would operationalize these guidelines in their daily practices, ensuring that ethical principles are upheld consistently across all AI applications. At each level, there should be an ***independent and multidisciplinary Board of Ethics*** tasked with overseeing the ethical application of AI technologies. These boards, comprising experts in fields such as law, technology, sociology, philosophy, public policy, among others, would provide objective, cross-disciplinary perspectives on ethical considerations, helping to prevent conflicts of interest and ensure accountability.
   - **Promote Responsible AI Practices**: Ensure all AI initiatives are guided by principles that prevent biases and discrimination, particularly in applications impacting vulnerable communities and people with protected characteristics, by embedding ethical considerations into the design and implementation phases. To support this, launch a nationwide awareness-raising campaign—similar to past tobacco awareness campaigns—to educate the public about the risks and ethical implications of AI. This

[276] European Commission: Joint Research Centre, Strand, R., Krieger, K. and Melchor, L., (2022). *Indicator dashboards in governance of evidence-informed policymaking – Thoughts on rationale and design criteria*. [pdf] Available at: https://data.europa.eu/doi/10.2760/328204 [Accessed on 09 November 2024]
[277] Funtowicz, S., & Liberatore, A. (2003). *'Democratising' expertise, 'expertising' democracy: What does this mean, and why bother?* Science and Public Policy, 30(3), 146–150. [pdf] Available at: https://academic.oup.com/spp/article-abstract/30/3/146/1628309?redirectedFrom=fulltext [Accessed on 10 November 2024]

campaign would inform citizens about issues such as algorithmic bias, privacy risks, and the potential for AI misuse, empowering them to engage critically with AI technologies and advocate for responsible practices.

2. **Support Contextual and Transdisciplinary AI Research and Development**
   - **Support Contextualized AI Research**: Direct funding towards independent research that examines evidence of AI's impact across context-specific environments, such as humanitarian aid, healthcare, agriculture, education, etc.. Research of this type would assess how AI technologies influence these fields in real-world settings, identifying both potential benefits and potential risks. The insights gained from such studies would inform more nuanced, effective, and responsible AI deployments, ensuring that AI applications are aligned with the unique needs and ethical considerations of each sector.
   - **Foster Transdisciplinary Tech Hubs:** Establish and support transdisciplinary technology hubs, where diverse fields—including but not limited to technology and innovation—work together. These hubs would bring together expertise from areas such as social sciences, ethics, arts, and public policy to collaborate on community-driven AI solutions that reflect local needs and values. By fostering local talent and promoting a diversity of perspectives, these hubs can help resist the monocultural tendencies of AI, creating space for culturally relevant, locally-rooted and inclusive AI applications that support sustainable development.

3. **Enhance Capacity Building and Critical AI Literacy**
   - **Implement Critical AI Literacy Programs**: Develop and support initiatives that not only boost digital literacy but also foster a critical understanding of AI among policymakers, practitioners, and the public. These programs should be implemented across all levels of governance—tailored to the specific needs of national, regional, and local contexts. They should cover the implications, limitations, and ethical considerations of AI, enabling informed decision-making and responsible use of AI at every societal level.

4. **Address AI Discrimination, Transparency, and Explainability**
   - **Establish Independent Oversight Bodies for Human Rights Compliance:** Following the recommendations of the Council of Europe's Human Rights Commissioner (2019)[278], establish independent and effective oversight bodies to ensure human rights compliance in the development, deployment, and use of AI systems by both public authorities and private entities. These bodies should operate independently from the entities they oversee, and they must be equipped with interdisciplinary expertise, competencies, and sufficient resources. Their mandate would include investigating and monitoring AI systems to ensure alignment with human rights standards, particularly regarding issues of bias, discrimination, transparency, and explainability.
   - **Implement Human-First Decision-Making and Bias Detection Mechanisms:** Ensure that, in cases when AI systems are deployed, humans remain the ultimate decision-makers,

[278] Council of Europe, Commissioner for Human Rights, 2019. *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*. Available at: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 [Accessed 09 November 2024].

especially within public sectors, by preventing fully automated decision-making until independent evidence confirms the reliability, fairness, and compliance of AI technologies. Invest in advanced research and tools to identify, measure, and address biases within AI algorithms. Additionally, develop mechanisms that ensure fully transparent and explainable AI systems, allowing users to understand decision-making processes, which is essential for building trust and ensuring equitable treatment across diverse populations (e.g., explainable AI models, audit trails and documentation, user-friendly interfaces, and regular bias and fairness audits).

5. **Mitigate Environmental Impact of AI**
   - **Monitor Environmental Impact and Promote Sustainable AI**: Advocate for energy-efficient AI models that minimise environmental footprints, particularly in large-scale ML applications with substantial energy demands. Require AI projects to assess and disclose their environmental impacts, including energy consumption and emissions, as part of their standard reporting to ensure accountability and promote sustainable practices.
   - **Invest in Decomputing as a Sustainable Alternative to AI:** Allocate funding towards research and development of *decomputing*—an emerging approach in AI that emphasises low-computation, low-energy methodologies. Decomputing emphasises a reduced dependency on large-scale computation, encouraging AI practices that prioritise minimal environmental impact, community participation, and a focus on real-world context rather than abstracted, large-scale data-driven predictions. This approach values human judgement and local expertise, countering the extractive, resource-intensive tendencies of conventional AI models.

6. **Foresight and Anticipatory Analysis in AI**
   - **Integrate Foresight and Speculative Design for Future-Ready AI Strategies**: Support and incorporate foresight analysis and speculative design as essential tools for exploring potential futures in the development and humanitarian sectors. Use these approaches to envision and prototype possible, probable, and preferable scenarios for AI deployment, examining emerging trends and anticipating ethical, social, and environmental implications. This anticipatory analysis will enable policymakers and practitioners to prepare for challenges and harness potential opportunities within the rapidly evolving landscape of AI for development.

SECTION 5.3

AI Hype-Reality Gap Model

To complement the previous discussions on democratising expertise and the provided recommendations, we introduce the AI Hype-Reality Gap Model[279]. This framework specifically addresses the challenges posed by the hype surrounding AI technologies in development and humanitarian sectors, which can often obscure the true reality of their capabilities, limitations, and impacts, leading to misinformed decisions

[279] AI Hype-Reality Gap Model is inspired and draws from Richard Heeks' Design-Reality Gap Model (2002), which examines the challenges in aligning ICT projects with real-world conditions in development contexts and e-governance.

and misplaced resources. The AI Hype-Reality Gap Model aims to help policymakers, practitioners, and other stakeholders critically assess the discrepancy between the expectations created by AI hype and the actual, empirical and independent evidence and outcomes achieved in real-world applications. By identifying this gap, the model aims to prevent misallocated resources, unmet development goals, and the exacerbation of existing inequalities that can arise when AI implementations fall short of their promised impact.

The model serves as acritical assessment high-level tool that enables stakeholders to:

- **Assess Realistic Expectations**: Determine whether the promises made about an AI technology are achievable in the given context.
- **Identify Overhyped Claims**: Highlight areas where AI solutions may be oversold[280] or misrepresented in terms of their capabilities and benefits.
- **Inform Decision-Making**: Provide evidence-based insights to guide policy and investment decisions, avoiding pitfalls associated with overhyped technologies.
- **Promote Accountability**: Encourage transparency and empirical validation of AI projects' claimed benefits, ensuring responsible innovation.

The model evaluates the gap across seven dimensions, focusing on factors that contribute to hype and the realities on the ground, and the gaps to be analysed by policymakers and other stakeholders

[280] For example, a 2019 audit of over 2,800 purported AI startups in Europe found that two-fifths of these firms were making no meaningful use of artificial intelligence in their products (see: Schulze, E., (2019), *40% of A.I. start-ups in Europe have almost nothing to do with A.I., research finds*. [online] Available at: www.cnbc.com/2019/03/06/40-percent-of-ai-start-ups-in-europe-not-related-to-ai-mmc-report.html [Accessed on 09 November 2024]). Similarly, investigations have uncovered cases where IT firms claimed to employ sophisticated AI capabilities but instead relied on low-tech manual procedures (see: Johnson, K., (2021), *Government audit of AI with ties to white supremacy finds no AI*. [online] Available at: https://venturebeat.com/2021/04/05/government-audit-of-ai-with-ties-to-white-supremacy-finds-no-ai/ [Accessed on 09 November 2024]; Morris, M., (2022), *AI' shopping startup exaggerated tech capabilities to potential investors*. [online] Available at: www.theinformation.com/articles/shaky-tech-and-cash-burning-giveaways-ai-shopping-startup-shows-excesses-of-funding-boom [Accessed on 09 November 2024]).

Table 5.1 AI Hype-Reality Gap Model

Dimensions	Hype	Reality	Gap Analysis
1. Promised Capabilities vs. Actual Functionality	High-performance claims, revolutionary features, transformative impacts	Actual performance in the given context; independent empirical validation of capabilities	Identify discrepancies between promised capabilities and functionality in real-world use
2. Claimed Evidence vs. Independent Evidence	Claims of efficacy and impact based on internal or anecdotal evidence provided by developers or vendors	Results from independent and peer-reviewed studies, third-party evaluations/audits, and on-the-ground observations of AI impacts	Assess reliability and validity of claimed evidence; identify any discrepancies between claimed and independently verified results
3. Expected Outcomes vs. Measured Impact	Projected benefits like improved efficiency, significant cost savings, developmental gains, or correct predictions	Tangible outcomes and any unintended consequences observed post-implementation	Assess how actual outcomes measure up to promised benefits, including any unintended impacts
4. Resource Requirements vs. Available Resources	Assumptions about the availability of necessary data, infrastructure, and expertise	Real availability and quality of resources such as data access, infrastructure robustness, and local expertise	Evaluate gaps in resource quality, availability, and accessibility
5. Scalability Claims vs. Practical Scalability	Claims that AI solutions can be easily scaled across various regions or sectors	Practical challenges in scaling, including contextual, logistical, legal, and cost-related limitations	Identify specific barriers to scaling (e.g., regional variability, cost constraints, legal compliances, etc.)
6. Ease of Integration vs. Integration Challenges	Promises of seamless integration with existing systems and processes	Actual difficulties in integrating with current infrastructure, compatibility issues, resistance to change	Document integration barriers and evaluate the feasibility of promised ease of integration
7. Ethical Considerations vs. Ethical Implementation	Claims that AI technology aligns with ethical standards and promotes social benefits	Observed ethical impacts, including unintended negative consequences or breaches in practice	Review ethical gaps and ensure alignment with community and organisational values

Policymakers and stakeholders can apply the model through the following approach:

- **Evaluate Claims Critically**: Begin by reviewing promotional materials, marketing claims, and proposals to identify the main capabilities and outcomes being promised by the AI technology.
- **Contextual Assessment**: Gather empirical independent evidence from pilot projects, case studies, and independent evaluations. Engage with local communities, practitioners, and experts to understand the ground realities and context-specific needs.
- **Conduct Gap Analysis**: Compare the hype with the actual outcomes, documenting discrepancies for each dimension where the technology falls short or exceeds expectations.
- **Evaluate Risks**: Assess the impact of identified gaps on project success, resource allocation, and stakeholder trust. Consider any unintended consequences that may arise from unmet expectations.
- **Decision-Making and Mitigation**: Adjust project goals to align with realistic assessments. Develop mitigation strategies, such as additional training or technology adjustments, to address specific gaps. Based on these findings, decide whether to proceed, modify the approach, or consider alternative solutions.

By introducing the AI Hype-Reality Gap Model, we aim to provide a tool for policymakers to critically assess AI initiatives and systems, ensuring that they are grounded in reality, supported by independent evidence, and capable of delivering tangible benefits. We hope that this tool serves as a starting point for fostering scrutiny, accountability, and evidence-based decision-making in the adoption of AI technologies, and that it can be adapted and appropriated to suit various contexts, including international development and humanitarian settings.

Bibliography

Aborujilah, A., et al., 2023. *IoT Integration in Agriculture: Advantages, Challenges, and Future Perspectives: Short Survey*. *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Istanbul, Turkiye, pp. 1-7. [online] Available at: https://ieeexplore.ieee.org/document/10322958 [Accessed on 03 November 2024]

AccessNow, (2024). *Joint statement – A dangerous precedent: how the EU AI Act fails migrants and people on the move.* [online] Available at: https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/ [Accessed on 01 June 2024].

Addy A., 2023. *Artificial Intelligence in the Supply Chain Management for Vaccine Distribution in the West African Healthcare Sector with a focus on Ghana.* [online] Available at: https://www.researchgate.net/publication/375603718_Artificial_Intelligence_in_the_Supply_Chain_Management_for_Vaccine_Distribution_in_the_West_African_Healthcare_Sector_with_a_focus_on_Ghana [Accessed on 24 May 2024]

AdmitYogi, (n.d.). *AdmitYogi: Your Admissions Helper*. [online] Available at: https://admityogi.com [Accessed on 04 August 2024]

African Union Development Agency, (2023). *Regulation and Responsible Adoption of AI for Africa Towards Achievement of AU Agenda 2063*. [pdf] Available at: https://onedrive.live.com/?authkey=%21AKJcwcnXeRGANKQ&id=14DDAD979C3656DF%2145404&cid=14DDAD979C3656DF [Accessed on 01 June 2024]

AI for SDGs Academy (n.d.). [online] Available at: https://ai-for-sdgs.academy/ [Accessed on 29 March 2024]

AI for SDGs Academy (n.d.). [online] Available at: https://ai-for-sdgs.academy/ [Accessed on 15 May 2024]

AI White Paper and Roadmap Review, (n.d). [online] Available at: https://onedrive.live.com/?authkey=%21AKJcwcnXeRGANKQ&id=14DDAD979C3656DF%2145404&cid=14DDAD979C3656DF [Accessed on 01 June 2024]

AI4Good (n.d.). *About Us*. [online] Available at: https://ai4good.org/about-us/ [Accessed on 29 March 2024]

Aiken, E.L., Bedoya, G., Coville, A., & Blumenstock, J.E. (2020). Targeting Development Aid with Machine Learning and Mobile Phone Data. COMPASS '20: Proceedings of the 3rd ACM SIG- CAS Conference on Computing and Sustainable Societies, 310–311. [doi] Available at: https://dl.acm.org/doi/10.1145/3378393.3402274 [Accessed on 14 October 2024]

Ajay, Agrawal., Joshua, S., Gans., Avi, Goldfarb. (2023). *Do we want less automation?*. Science, 381, 155-158. [online] Available at: https://www.science.org/doi/abs/10.1126/science.adh9429 [Accessed on 14 October 2024]

Al Jazeera ,(2023). *UN's Volker Turk: A Quarter of Humanity is Caught in 55 Global Conflicts*. Talk to Al Jazeera, 22 December 2023. [online] Available at: https://www.aljazeera.com/program/talk-to-al-jazeera/2023/12/22/uns-volker-turk-a-quarter-of-humanity-is-caught-in-55-global-conflicts [Accessed on 15 May 2024]

ALEKS, (n.d.). *ALEKS: Assessment and Learning in Knowledge Spaces*. [online] Available at: https://www.aleks.com/?_s=2894576279172929 [Accessed on 04 August 2024]

Algeria Press Service (APS), (n.d.). *National Artificial Intelligence Strategy 2020–2030 Presented*. [online] Available at: https://www.aps.dz/en/health-science-technology/37591-higher-education-national-artificial-intelligence-strategy-2020-2030-presented [Accessed on 01 June 2024]

Algorithmic Sabotage Research Group, 2024. Theorizing "Algorithmic Sabotage". [pdf] Available at: https://cryptpad.fr/file/# [Accessed on 01 June 2024].

Ali, S.H., Giurco, D., Arndt, N., Nickless, E., Brown, G., Demetriades, A., & Yakovleva, N. (2017). *Mineral supply for sustainable development requires resource governance*. *Nature*, 543(7645), 367–372. [doi] Available at: 10.1038/nature21359 [Accessed on 08 November 2024]

American Institutes for Research (AIR) (n.d.). *Methods for Extremely Rapid Observation of Nutritional Status (MERON)*. [online] Available at: https://www.air.org/project/methods-extremely-rapid-observation-nutritional-status-meron [Accessed on 24 May 2024]

Amnesty International, (2023). *Serbia: World Bank-funded digital welfare system exacerbating poverty, especially for Roma and people with disability.* [online] Available at: https://www.amnesty.org/en/latest/news/2023/12/serbia-world-bank-funded-digital-welfare-system-exacerbating-poverty-especially-for-roma-and-people-with-disabilities/#:~:text=Launched%20in%202022%20and%20aimed,determining%20eligibility%20for%20social%20assistance. [Accessed on 30 October 2024].

Amnesty International, (2023). *Trapped by Automation: Poverty discrimination in Serbia's welfare state.* [online] Available at: https://www.amnesty.org/en/latest/research/2023/12/trapped-by-automation-poverty-and-discrimination-in-serbias-welfare-state/ [Accessed on 30 October 2024]

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016) *Machine Bias*. [online] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed on 14 October 2024]

Aula, V. & Bowles, J. (2023). *Stepping back from Data and AI for Good – current trends and ways forward*. Big Data & Society, 10(1). [doi] Available at: https://doi.org/10.1177/20539517231173901 [Accessed on 29 March 2024]

BakerHostetler, (2022). *Proxy problems: solving for discrimination in algorithms*. [online] Available at: https://www.bhfs.com/insights/alerts-articles/2022/proxy-problems-solving-for-discrimination-in-algorithms [Accessed 21 October 2024].

Bankhwal M., et. al., 2024. *AI for social good: Improving lives and protecting the planet.* McKinsey Global Institute. [pdf] Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/ai-for-social-good#/download/%2F~%2Fmedia%2Fmckinsey%2Fbusiness%20functions%2Fquantumblack%2Four%20insights%2Fai%20for%20social%20good%2F2024%2Fai-for-social-good-improving-lives-and-protecting-the-planet-v2.pdf%3FshouldIndex%3Dfalse [Accessed on 20 May 2024]

Barua, P.D., Vicnesh, J., Gururajan, R., Oh, S.L., Palmer, E., Azizan, M.M., Kadri, N.A. & Acharya, U.R., (2022). *Artificial Intelligence Enabled Personalised Assistive Tools to Enhance Education of Children with Neurodevelopmental Disorders: A Review*. International Journal of Environmental Research and Public Health, 19(3), p. 1192. [doi] Available at: https://doi.org/10.3390/ijerph19031192 [Accessed on 04 August 2024]

Bashford, A. & Levine, P. (eds.), (2010). *The Oxford Handbook of the History of Eugenics*. Oxford University Press, Oxford.
Beijing Academy of Artificial Intelligence, (2019). *Beijing AI Principles*. [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s11623-019-1183-6.pdf [Accessed on 01 June 2024]
Benin Ministry of Digitalisation, (2024). *Benin National Artificial Intelligence and Big Data Strategy*. [pdf] Available at: https://www.d4daccess.eu/sites/default/files/knowledge_base_products_files/national-artificial-intelligence-and-big-data-strategy-1682673348%20%281%29.pdf [Accessed on 01 June 2024]
Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, Cambridge.
Berendt, B. (2019) *AI for the Common Good?! Pitfalls, challenges, and ethics pen-testing.* Paladyn, Journal of Behavioral Robotics, Vol. 10 (Issue 1), pp. 44-65. [doi] Available at: https://doi.org/10.1515/pjbr-2019-0004 [Accessed on 29 March 2024]
Biesta, G. J. J. (2011). *Good Education in an age of measurement: Ethics, politics, democracy.* Paradigm Publishers.
Birhane, A. (2021). *Algorithmic injustice: A relational ethics approach. Patterns*, 2(2), 100205. [doi] Available at: https://doi.org/10.1016/j.patter.2021.100205 [Accessed on 09 November 2024].
Birhane, A., Prabhu, V. U. & Kahembwe, E., (2021). *Multimodal datasets: misogyny, pornography, and malignant stereotypes.* [pdf] Available at: https://arxiv.org/abs/2110.01963 [Accessed on 08 November 2024]
Birhane, A., Prabhu, V., Han, S. & Boddeti, V. N., (2023). *On Hate Scaling Laws For Data-Swamps*. [pdf] Available at: https://arxiv.org/abs/2306.13141[Accessed on 08 November 2024]
Birhane, A., Prabhu, V., Han, S., Boddeti, V. N. & Luccioni, A. S., (2023). *Into the LAIONs Den: Investigating Hate in Multimodal Datasets*. [pdf] Available at: https://arxiv.org/abs/2311.03449; [Accessed on 08 November 2024]
Black E., 2001. *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation*. NYC: Three Rivers Press
BlueDot (n.d.). *Home*. [online] Available at: https://bluedot.global/ [Accessed: 24 May 2024]
Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A., (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. [pdf] Available at: https://arxiv.org/abs/1607.06520 [Accessed on 08 November 2024].
Bonan, G.B. & Doney, S.C., (2018). *Climate, Ecosystems, and Planetary Futures: The Challenge to Predict Life in Earth System Models*. Science, 359, eaam8328. Available at: https://doi.org/10.1126/science.aam8328 [Accessed on 01 November 2024]
Bridges, K.M. (2017). *The Poverty of Privacy Rights*. Stanford University Press, Stanford, CA.
Brownie J., 2019. A Gentle Introduction to Generative Adversarial Networks. Available at: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/ [Accessed on 17 March 2024]
Buolamwini, J. & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (FAT* 2018). [pdf] Available at: https://proceedings.mlr.press/v81/buolamwini18a.html [Accessed on 04 August 2024]
Carboni, K., (2023). *Nel 2023 la crisi del clima ha scatenato un estremo evento al giorno in Italia.* [online] Available at: https://www.wired.it/article/clima-crisi-2023-italia-evento-estremo-numeri/ [Accessed on 01 November 2024].
Central Information Board (CIB), (2024). *Mauritius AI Strategy*. [pdf] Available at: https://cib.govmu.org/Documents/Strategies/Mauritius%20AI%20Strategy%20(7).pdf [Accessed on 01 June 2024]
Centre for Humanitarian Data (n.d.). *Catalogue for Predictive Models in the Humanitarian Sector*. [online] Available at: https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/ [Accessed on 29 March 2024]
Chen, Z., Zhu, Z., Jiang, H., and Sun, S., (2020). *Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods*. *Journal of Hydrology*, 591, p.125286. [online] Available at: https://www.sciencedirect.com/science/article/abs/pii/S0022169420307460 [Accessed on 01 November 2024]
Cheney-Lippold, J. (2011). *A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control*. *Theory, Culture & Society*, 28(6), 164–181. [doi] Available at: https://doi.org/10.1177/0263276411424420 [Accessed on 08 November 2024]
Choo, L. (2024). *Meta's Ray-Ban Smart Glasses Exposed to Privacy Breach by Students*. Forbes. [online] Available at: https://www.forbes.com/sites/lindseychoo/2024/10/04/meta-ray-bans-ai-privacy-surveillance/ [Accessed on 08 November 2024]
Christiaens, T. (2024). *Nationalize AI!*. AI & Society. [doi] Available at: https://doi.org/10.1007/s00146-024-01897-0 [Accessed on 01 June 2024]
Chui M., et. al. 2018. *Notes from the Frontier: Applying AI for Social Good.* McKinsey Global Institute. [pdf] Available at: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf [Accessed on 13 May 2024]
Civitas Learning, (n.d.). *Civitas Learning: Transforming Higher Education*. [online] Available at: https://www.civitaslearning.com/ [Accessed on 04 August 2024]
Climate TRACE, (n.d). [online] Available at: https://climatetrace.org/ [Accessed on 05 November 2024]
Climate TRACE, (n/d). Climate Trace - About the Coalition. [online] Available at: https://climatetrace.org/about [Accessed on 05 November 2024].
Climate TRACE, (n/d: 40). A view from space: tracking emissions state by state. [pdf] Available at: https://climatetrace.org/api/download?file=starrs-2023-03-en [Accessed on 05 November 2024]
Climate TRACE, (n/d:. 15-23). A view from space: tracking emissions state by state. [pdf] Available at: https://climatetrace.org/api/download?file=starrs-2023-03-en [Accessed on 05 November 2024]
Conger, K. & Gebeloff, R., (2019). *The Humans Working Behind the AI Curtain*. [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed 09 November 2024].

Copeland J., 2000. Alan Turing and the origins of AI. [online] Available at: https://www.alanturing.net/turing_archive/pages/reference%20articles/what_is_ai/What%20is%20AI03.html [Accessed on 20 March 2024]
Copernicus, (2022). *Collapse of the Marmolada glacier, Italy.* [online] Available at: https://www.copernicus.eu/it/node/11693 [Accessed on 01 November 2024]
Coppi, G., Moreno Jimenez, R. & Kyriazi, S. (2021). *Explicability of Humanitarian AI: A Matter of Principles*. International Journal of Humanitarian Action, 6, p. 19. [doi] Available at: https://doi.org/10.1186/s41018-021-00096-6 [Accessed on 01 June 2024]
Couldry, N. & Mejias, U.A., (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.
Council of Europe, (as per 18 December 2023). *Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*. [pdf] Available at: https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043 [Accessed on 15 March 2024]
Council of Europe, (n.d.). *The Framework Convention on Artificial Intelligence*. [online] Available at: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence [Accessed: 01 June 2024]
Council of Europe, Commissioner for Human Rights, 2019. *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*. Available at: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 [Accessed 09 November 2024].
Cowls, J., Tsamados, A., Taddeo, M. & Floridi, L. (2021). *A definition, benchmark and database of AI for social good initiatives*. Nature Communications, 12, Article 419. [pdf] Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3893654_code2644503.pdf?abstractid=3826465&mirid=1 [Accessed on 29 March 2024]
Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
Crawford, K., 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
Crisis Group (n.d.). *CrisisWatch*. [online] Available at: https://www.crisisgroup.org/crisiswatch [Accessed: 24 May 2024]
Cuong, Nguyen., Dat, Tho, Tran. (2018). *4. Proxy Means Tests to Identify the Income Poor: Application for the Case of Vietnam: Journal of Asian and African Studies.* [doi] Available at: https://doi.org/10.1177/00219096177094 [Accessed on 21 Oct. 2024].
Das Sh., 2017. *CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more…* [online] Available at: https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5 [Accessed on 21 March 2024]
David, J., (2023). *Students investigate how artificial intelligence perpetuates biases*. Information School, University of Washington. [online] Available at: https://ischool.uw.edu/news/2023/10/students-investigate-how-artificial-intelligence-perpetuates-biases [Accessed on 08 November 2024].
Denton E., Hanna A., Amironesei R., 2021. *On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet, Big data & Society,* 8(2). [doi] Available at: https://doi.org/10.1177/20539517211035955 [Accessed on 21 March 2024]

Department of Industry, Science and Resources, (n.d.). *Australia's AI Ethics Principles*. [online] Available at: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles [Accessed on 01 June 2024]
Development Pathways blog, 2018. *Targeting humanitarian aid: something to be left to opaque algorithms?* [online] https://www.developmentpathways.co.uk/blog/targeting-humanitarian-aid-something-to-be-left-to-opaque-alogorithms/ [Accessed on 01 June 2024]
Drahl C., 2023. *AI was asked to create images of Black African docs treating white kids. How'd it go?* [online] Available at: https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it- [Accessed on 01 June 2024]
Düben, P., (2020). *AI and machine learning at ECMWF.* [online] Available at: https://www.ecmwf.int/en/newsletter/163/news/ai-and-machine-learning-ecmwf [Accessed on 01 November 2024]
Dwivedi, Y.K., Hughes, D.L., Baabdullah, A.M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M.M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C.M.K. & Wade, M. (2022). *Metaverse Beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice, and Policy*. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S0268401222000767 [Accessed on 04 August 2024]
Earl C. C., 2021. *Towards an Abolitionist AI: the role of Historically Black Colleges and Universities*. [pdf] Available at: https://arxiv.org/abs/2101.02011 [Accessed on 01 June 2024]
EE Times Asia, (2021). *Using AI to Keep City Clean Makes Amsterdam GO SMART Award Winner.* [online] Available at: https://www.eetasia.com/using-ai-to-keep-city-clean-makes-amsterdam-2021-go-smart-award-winner/ [Accessed on 14 October 2024]
Elhra, 2015. *Text messages used to deliver humanitarian aid*. [online] Available at: https://www.elrha.org/news-and-blogs/text-messages-used-deliver-humanitarian-aid/ [Accessed on 14 May 2024]
Emergency Telecommunications Cluster (ETC) (n.d.). *Introducing the New ETC Chatbot: AI-powered Buddy in Times of Crisis*. [online] Available at: https://www.etcluster.org/blog/introducing-new-etc-chatbot-ai-powered-buddy-times-crisis#:~:text=The%20ETC%20Chatbot%20has%20the,actions%2C%20enhancing%20resilience%20and%20empowerment [Accessed on 24 May 2024]
Emerson P, 1995. *Building IBM: Shaping an Industry and its Technology*. Massachusetts: MIT Press
Estampa, (2024). *Cartography of Generative AI.* [online] Available at: https://cartography-of-generative-ai.net/ [Accessed on 08 November 2024]
Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, New York.
European Commission: Joint Research Centre, Strand, R., Krieger, K. and Melchor, L., (2022). *Indicator dashboards in governance of evidence-informed policymaking – Thoughts on rationale and design criteria*. [pdf] Available at: https://data.europa.eu/doi/10.2760/328204 [Accessed on 09 November 2024]

European Union, (n.d.). *Artificial Intelligence Act*. [online] Available at: https://artificialintelligenceact.eu/ [Accessed on 01 June 2024]
Facebook (n.d.). *AI for India*. [online] Available at: https://fbaiforindia.splashthat.com [Accessed on 29 March 2024]
Fakiya V., 2023. *Nigeria to create a National AI Strategy.* [online] Available at: https://techpoint.africa/2023/08/29/techpoint-digest-658/ [Accessed on 01 June 2024]
Fakiya V., 2023. *Nigeria to create a National AI Strategy.* [online] Available at: https://techpoint.africa/2023/08/29/techpoint-digest-658/ [Accessed on 01 June 2024]
Fallon K., 2020. *UN warns of impact of smart borders on refugees: 'Data collection isn't apolitical'*. [online] Available at: https://www.theguardian.com/global-development/2020/nov/11/un-warns-of-impact-of-smart-borders-on-refugees-data-collection-isnt-apolitical [Accessed on 01 June 2024]
Fan, Y., Huang, T., Meng, Y. & Cheng, S. (2023). *The current opportunities and challenges of Web 3.0*. [pdf]Available at: https://doi.org/10.48550/arXiv.2306.03351 [Accessed on: 23 March 2024]
Feathers S.. (2024). *Revealed: A Californian city is training AI to spot homeless encampments.* [online] Available at: https://www.theguardian.com/technology/2024/mar/25/san-jose-homelessness-ai-detection [Accessed on 14 October 2024]
Ferili S., et. al. 2021. *Artificial Intelligence for Sustainable Development.* Available only in Italian language: https://www.cnr.it/sites/default/files/public/media/attivita/editoria/VOLUME%20FULL%2014%20digital%20LIGHT.pdf [Accessed on 13 May 2024]
FEWS NET (n.d.). *Home*. [online] Available at: https://fews.net/ [Accessed on 24 May 2024]
Floridi, L. & Cowls, J. (2019). *A Unified Framework of Five Principles for AI in Society*. [doi] Available at: http://dx.doi.org/10.2139/ssrn.3831321 [Accessed on 01 June 2024]
Flynn Sh., 2020. *The Difference Between Symbolic AI and Connectionist AI*. [online] Available at: https://blog.re-work.co/the-difference-between-symbolic-ai-and-connectionist-ai/ [Accessed on 21 March 2024]
Friedrich K., *The Dehomag D11 Tabulator - A Milestone in the History of Data Processing.*
Friel K., 2023. *A Look Back on the Dartmouth Summer Research Project on Artificial Intelligence.* [online] Available at: https://www.thedartmouth.com/article/2023/05/a-look-back-on-the-dartmouth-summer-research-project-on-artificial-intelligence [Accessed on 21 March 2024]
Frost, N., (2024). *The Impoverished Publicness of Algorithmic Decision Making*, *Oxford Journal of Legal Studies*. [doi] Available at: https://doi.org/10.1093/ojls/gqae027 [Accessed on 08 November 2024]
Funtowicz, S., & Liberatore, A. (2003). *'Democratising' expertise, 'expertising' democracy: What does this mean, and why bother?* Science and Public Policy, 30(3), 146–150. [pdf] Available at: https://academic.oup.com/spp/article-abstract/30/3/146/1628309?redirectedFrom=fulltext [Accessed on 10 November 2024]
Funtowicz, S., Ravetz, J. 2003. *Post-Normal Science*. *International Society for Ecological Economics' Internet Encyclopaedia of Ecological Economics.*
Funtowicz, S.O., Ravetz, J.R., 1993. *Science for the Post-Normal Age.* Futures, 25[7], pp. 739-55. [doi] Available at: https://doi.org/10.1016/0016-3287(93)90022-L [Accessed on 08 November 2024]

G20, (2019). *AI Principles*. [pdf] Available at: https://www.mofa.go.jp/mofaj/files/000486596.pdf [Accessed on 04 June 2024]
Gabrys, J. (2013). *Digital rubbish: A Natural History of Electronics.* Michigan, University of Michigan Press.
Gebru, T. & Torres, É.P., 2023. *The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence*. *First Monday*, 29(4). [doi] Available at: https://doi.org/10.5210/fm.v29i4.13636 [Accessed 09 November 2024].
Gelb A., Clark J., 2013. *Identification for Development: The Biometric Revolution*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2226594 [Accessed on 24 May 2024]
Gershgorn D., 2017. The data that transformed AI research—and possibly the world. [ionline] Available at: https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world [Accessed on 21 March 2024]
Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
Global Forest Watch (n.d.). *Home*. [online] Available at: https://www.globalforestwatch.org/ [Accessed on 24 May 2024]
Global Parenting Initiative (n.d.). *ParentText*. [online] Available at: https://globalparenting.org/parenttext [Accessed on 24 May 2024]
Goel, A. K., & Polepeddi, L. (2016). *Jill Watson: A virtual teaching assistant for online education*. Georgia Institute of Technology.
Gonzalez Monserrate, S. (2022). *The Cloud is Material: On the Environmental Impacts on Computation and Data Storage* [online] Available at: https://mit-serc.pubpub.org/pub/the-cloud-is-material/release/2 [Accessed on 08 November 2024].
Google AI (n.d.). *Why AI*. [online] Available at: https://ai.google/why-ai/ [Accessed on 29 March 2024]
Gradescope, (n.d.). *Gradescope: Simplifying Grading for Educators*. [online] Available at: https://www.gradescope.com/ [Accessed on 04 August 2024].
Graesser, A.C., Chipman, P., Olney, A. & Haynes, B.C. (2005). *AutoTutor: An Intelligent Tutoring System With Mixed-Initiative Dialogue*. IEEE Transactions on Education, 48(4), pp. 612–618. [doi] Available at: https://doi.org/10.1109/TE.2005.856149 [Accessed on 04 August 2024]
Green, B. (2019). *"Good" isn't good enough*. [pdf] Available at: https://www.semanticscholar.org/paper/%E2%80%9CGood%E2%80%9D-isn%E2%80%99t-good-enough-Green/dc2fed36474b1d1dd497b8f08e06183bb65cf48f [Accessed on 29 March 2024]
Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M., (2023). *Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection*. arXiv. [pdf] Available at: https://arxiv.org/abs/2302.12173 [Accessed on 08 November 2024]
Grosh M.E. and Baker J.L., (1995). *Proxy means tests for targeting social programs: simulations and speculation. Living Standards Measurement Study (LSMS) working paper, no. LSM 118. Washington, D.C.:* [online] Available at: http://documents.worldbank.org/curated/en/750401468776352539/Proxy-means-tests-for-targeting-social-programs-simulations-and-speculation [Accessedon 21 Oct. 2024].

Guo, E. & Renaldi, A., (2022). *Deception, exploited workers, and cash handouts: How Worldcoin recruited its first half a million test users.* [online] Available at: https://www.technologyreview.com [Accessed 09 November 2024].
Gurney, K. R., Aslam, B., Dass, P., Gawuc, L., Hocking, T., Barber, J. J., & Kato, A. (2024). Assessment of the Climate TRACE global powerplant CO_2 emissions. *Environmental Research Letters*. [online] Available at: https://iopscience.iop.org/article/10.1088/1748-9326/ad8364 [Accessed on 05 November 2024]
GVU Center, Georgia Tech, (n.d.). *Virtual Teaching Assistant: Jill Watson*. [online] Available at: https://gvu.gatech.edu/research/projects/virtual-teaching-assistant-jill-watson [Accessed on 04 August 2024].
Harwell, D., & Timberg, C. (2019). *Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Casts Doubt on Their Expanding Use*. [online] Available at: https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/ [Accessed on 08 November 2024]
Heeks, R. (2020). *ICT4D 3.0? Part 1—The components of an emerging "digital-for-development" paradigm*. The Electronic Journal of Information Systems in Developing Countries, 86(3), pp. e12124. [doi] Available at: https://doi.org/10.1002/isd2.12124 [Accessed on: 23 March 2024]
Heeks, R. (2020). *ICT4D 3.0? Part 2—The patterns of an emerging "digital-for-development" paradigm*. The Electronic Journal of Information Systems in Developing Countries, 86(3), pp. e12123. [doi] Available at: https://doi.org/10.1002/isd2.12123 [Accessed on: 23 March 2024]
Heeks, R., (2002). *Information systems in developing countries: Failure, success and local improvisations. The Information Society*, 18(2), 101-112. [doi] Available at: https://doi.org/10.1080/01972240290075039 [Accessed on 08 November 2024]
Herzog, D. (2000). *Poisoning the Minds of the Lower Orders*. Paperback edition. Harvard University Press.
Holmes, W. & Porayska-Pomsta, K. (eds.) (2022). *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*. Routledge, Abingdon.
Holmes, W. & Tuomi, I., (2022). *State of the Art and Practice in AI in Education*. European Journal of Education, 57(4), pp. 542–560. [doi] Available at: https://doi.org/10.1111/ejed.12533 [Accessed on 30 July 2024]
Holmes, W., (2023). *The Unintended Consequences of Artificial Intelligence and Education* [pdf] Available at: https://www.ei-ie.org/file/740 [Accessed on 30 July 2024]
Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V., (2018). *The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. In Proceedings of the 8th international conference on learning analytics and knowledge - LAK '18 (pp. 79–88).* Association for Computing Machinery. [doi] Available at: https://doi.org/10.1145/3170358.317037 [Accessed on 04 August 2024]
Holzmeyer, C. (2021). *Beyond 'AI for Social Good' (AI4SG): Social transformations—not tech-fixes—for health equity*. Interdisciplinary Science Reviews, 46(1–2), pp. 94–125. Available at: https://doi.org/10.1080/03080188.2020.1840221 [Accessed on 29 March 2024]
Houssou I. N., (2013). *Operational Poverty Targeting by Proxy Means Tests: Models and Policy Simulations for Malawi.* 1st ed. Frankfurt: Peter Lang GmbH.
Hsiang, S., Kopp, R.E., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D.J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K. and Houser, T., (2017). *Estimating economic damage from climate change in the United States*. *Science*, 356(6345), pp.1362-1369. [doi] Available at: https://www.science.org/doi/10.1126/science.aal4369 [Accessed on 01 November 2024] https://doi.org/10.1177/1474904116654917 [Accessed on 08 November 2024]
Human Rights Watch, (2023). *Automated Neglect: How the World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights*. [pdf] Available at: https://www.hrw.org/sites/default/files/media_2023/11/thr_jordan0623%20web.pdf [Accessed on 30 October 2024]
Humanitarian OpenStreetMap Team (HOTOSM) (n.d.). *FAIR-DEV: Data for Equitable and Inclusive Development*. [online] Available at: https://fair-dev.hotosm.org/
Hurley, M., & Adebayo, J. (2016). *Credit scoring in the era of big data. Yale Journal of Law and Technology*, 18(1), 148-216. [pdf] Available at: https://openyls.law.yale.edu/bitstream/handle/20.500.13051/7808/Hurley_Mikella.pdf [Accessed on 08 November 2024].
Iazzolino, G., & Stremlau, N., (2024). AI for social good and the corporate capture of global development. *Information Technology for Development*, *30*(4), 626–643. [doi] Available at: https://doi.org/10.1080/02681102.2023.2299351 [Accessed on 08 November 2024]
IBM (n.d.). *AI for Social Good: Advantage Reports*. [online] Available at: https://www.ibm.com/watson/advantage-reports/ai-social-good.html [Accessed on 29 March 2024]
IBM, (n.d.). *Watson Education Classroom*. [online] Available at: https://www.ibm.com/mysupport/s/topic/0TO50000000Qei8GAC/watson-education-classroom?language=en_US [Accessed on 04 August 2024]
ICRC, 2022. *Cyber attack on ICRC: What we know*. [online] Available at: https://www.icrc.org/en/document/cyber-attack-icrc-what-we-know%E2%80%8B [Accessed on 01 June 2024]
Ilich, I. [1973], 2009 (p. 77-83). *Tools for Conviviality*. Marion Boyars Publishers Ltd
Imagenet. [online] Available at: https://image-net.org/challenges/LSVRC/ [Accessed on 21 March 2024]
Inasafe (n.d.). *Inasafe Home*. [online Available at: https://inasafe.org/home/index.html [Accessed: 24 May 2024]
Intel (n.d.). *AI for Social Good*. [online] Available at: https://www.intel.ai/ai4socialgood/ [Accessed on 29 March 2024]
Internal Displacement Monitoring Centre (IDMC) (n.d.). *Monitoring Tools: Monitoring Platform*. [online] Available at: https://www.internal-displacement.org/monitoring-tools/monitoring-platform/ [Accessed on 24 May 2024]
International Federation of Red Cross and Red Crescent Societies, 2023. *Chatbots in Humanitarian Contexts*. [pdf] Available at: https://communityengagementhub.org/wp-content/uploads/sites/2/2023/06/20230623_CEA_Chatbots.pdf [Accessed on 24 May 2024]
International Labour Organisation, (2022). *The Economic Assistance Programme in Albania: Challenges and Reform Trends.* [pdf] Available at: https://albania.un.org/sq/download/136372/237264 [Accessed on 27 October 2024].

International Monetary Fund, (2024). *Unemployment Rate - World Economic Outlook (WEO) Data Mapper - IMF Data Mapper*. [online] Available at: https://www.imf.org/external/datamapper/LUR@WEO/VNM/THA/SGP/PHL/MYS/IDN [Accessed on 17 October 2024].

International Rescue Committee (IRC), (2024). *OpenAi x International Rescue Committee: Leveraging AI to Scale Ed-Tech in Crisis Affected Settings.* [online] https://www.rescue.org/press-release/openai-x-international-rescue-committee-leveraging-ai-scale-ed-tech-crisis-affected [Accessed on 04 August 2024]

International Telecommunication Union (ITU) (n.d.). *About AI for Good*. [online] Available at: https://aiforgood.itu.int/about-ai-for-good/#:~:text=The%20goal%20of%20AI%20for,United%20Nations%20platform%20on%20AI [Accessed on 29 March 2024]

Introna, L. D., (2016). *Algorithms, Governance, and Governmentality: On Governing Academic Writing*. Science, Technology, & Human Values, 41(1), 17-49. [doi] Available at: https://doi.org/10.1177/0162243915587360 [Accessed on 04 August 2024]

Jean N., et al., 2016. *Combining satellite imagery and machine learning to predict poverty.* [online] Available at: https://www.science.org/doi/10.1126/science.aaf7894 [Accessed on 17 March 2024]

Johnson, K., (2021), *Government audit of AI with ties to white supremacy finds no AI*. [online] Available at: https://venturebeat.com/2021/04/05/government-audit-of-ai-with-ties-to-white-supremacy-finds-no-ai/ [Accessed on 09 November 2024]

Jones, N., (2017). *How machine learning could help to improve climate forecasting.* [pdf] Available at: https://doi.org/10.1038/548379a [Accessed on 01 November 2024]

Kaplan J., Morgan S. (2018). *Predicting Displacement: Using predictive analytics to build a better future for displaced children.* [pdf] Available at: https://resourcentre.savethechildren.net/pdf/predicting_displacement_report_-_save_the_children_mdi.pdf/ [Accessed on 14 October 2024]

Katz, B. (2020). *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. Columbia University Press, New York.

Katz, M.B. (1996). *In the Shadow of the Poorhouse: A Social History of Welfare in America*. 10th Anniversary edn. Basic Books, New York.

Khazanchi, R., & Khazanchi, P., (2021). *Artificial intelligence in education: A closer look into intelligent tutoring systems.* In A. Singh, C. J. Yeh, S. Blanchard, & L. Anunciação (Eds.), *Handbook of research on critical issues in special education for school rehabilitation practices* (pp. 256–277). Information Science Reference/IGI Global. [doi] Available at: https://doi.org/10.4018/978-1-7998-7630-4.ch014 [Accessed on 04 August 2024]

Kidd, S. and Wylde, E., (2011: 4). *Targeting the Poorest: An assessment of the proxy means test methodology. Canberra: Australian Agency for International Development (AusAID).* [pdf] Available at: https://www.dfat.gov.au/sites/default/files/targeting-poorest.pdf [Accessed 21 October 2024].

Kidd, S., Gelders, B. & Bailey-Athias, D., (2017). *Exclusion by Design: An Assessment of the Effectiveness of the Proxy Means Test Poverty Targeting Mechanism*. ESS Working Paper, No. 56. International Labour Organization (ILO). [pdf] Available at: https://researchrepository.ilo.org/view/pdfCoverPage?instCode=41ILO_INST&filePid=13131727070002676&download=true [Accessed 21 October 2024].

Kira Talent, (n.d.). *Kira Talent: Holistic Admissions Solution*. [online] Available at: https://www.kiratalent.com/ [Accessed on 04 August 2024]

Knox, J. (2020). *Artificial Intelligence and Education in China. Learning, Media and Technology, 45(3)*, pp. 298–311. Available at: https://eric.ed.gov/?id=EJ1265613 [Accessed on 04 August 2024]

Koepke, L., (2020). *Predictive policing algorithms are racist. They need to be dismantled*. [online] Available at: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ [Accessed on 08 November 2024].

Komljenovic, J., Sellar, S., Birch, K. & Hansen, M., (2024). *Assetization of Higher Education's Digital Disruption*. In: World Yearbook of Education 2024: Digitalisation of Education in the Era of Algorithms, Automation and Artificial Intelligence, pp. 122–139. Routledge Taylor & Francis Group. [pdf] Available at: https://www.researchgate.net/profile/Janja-Komljenovic/publication/376025761_Assetization_of_higher_education's_digital_disruption/links/65673337b1398a779dc6c477/Assetization-of-higher-educations-digital-disruption.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19 [Accessed on 30 July 2024]

Komljenovic, J., Williamson, B., Eynon, R. & Davies, H.C., (2023). *When Public Policy 'Fails' and Venture Capital 'Saves' Education: Edtech Investors as Economic and Political Actors*. Globalisation, Societies and Education. [doi] Available at: https://doi.org/10.1080/14767724.2023.2272134 [Accessed on 30 July 2024]

Krizhevsky A., Sutskever I., Hinton G, 2012. *ImageNet Classification with Deep Convolutional Neural Networks*. [pdf] Available at: https://proceedings.neurips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf [Accessed on 21 March 2024]

Lacuna Fund (n.d.). *Language Datasets*. [online] Available at: https://lacunafund.org/datasets/language/ [Accessed on 24 May 2024]

Lambert GA iGEM Team (2018). *CALM Model*. [online] Available at: https://2018.igem.org/Team:Lambert_GA/CALM_MODEL#target6 [Accessed: 24 May 2024]

Lambrecht, A. & Tucker, C.E. (2018). *Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads*. [doi] Available at: http://dx.doi.org/10.2139/ssrn.2852260 (Accessed on 24 May 2024]

Lawal S., (2024). *South Africa: 30 years after apartheid, what has changed?* [online] Available at: https://www.aljazeera.com/news/2024/4/27/south-africa-30-years-after-apartheid-what-has-changed [Accessed: on 17 October 2024].

Lee, D. & Tiwari, M., 2023. *Prompt Infection: LLM-to-LLM Prompt Injection within Multi-Agent Systems*. [pdf] Available at: https://arxiv.org/html/2410.07283 [Accessed 12 Nov. 2024].

Leetaru K., 2018. *Does AI Truly Learn And Why We Need to Stop Overhyping Deep Learning*. [online] Available at: https://www.forbes.com/sites/kalevleetaru/2018/12/15/does-ai-truly-learn-and-why-we-need-to-stop-overhyping-deep-learning/ [Accessed on 17 March 2024].

Legambiente, (2024). *L'Italia e l'Europa rischiano di rimanere senza ghiacciai.* [online] Available at: https://www.legambiente.it/comunicati-stampa/litalia-e-leuropa-rischiano-di-rimanere-senza-ghiacciai/ [Accessed on 01 November 2024].
Lehuedé, S. (2022). *Big tech's new headache: Data centre activism flourishes across the world. [online]* Available at: https://blogs.lse.ac.uk/medialse/2022/11/02/big-techs-new-headache-data-centre-activism-flourishes-across-the-world/ [Accessed 08 November 2024].
Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). *Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era*. Journal of Applied Research in Memory and Cognition, 6(4), 353–369. [doi] Available at: https://doi.org/10.1016/j.jarmac.2017.07.008 [Accessed on 09 November 2024]
Lewis, H. (2023). *Toward AI Realism*. [online] Available at: https://spectrejournal.com/toward-ai-realism/ [Accessed 18 November 2024]
Lexplore, (n.d.). *Lexplore Assessment*. [online] Available at: https://lexplore.com/lexplore-assessment [Accessed on 04 August 2024]
Liberty, (2020). *Legal challenge: Ed Bridges v South Wales Police*. Liberty. [online] Available at: https://www.libertyhumanrights.org.uk/issue/legal-challenge-ed-bridges-v-south-wales-police/ [Accessed on 08 November 2024].
Lindh, M., & Nolin, J. (2016). *Information We Collect: Surveillance and Privacy in the Implementation of Google Apps for Education. European Educational Research Journal,* 15, 644-663. [doi] Available at:
Lipton, Z. C. (2018). *The Mythos of Model Interpretability*. *Communications of the ACM*, 61(10), 36-43. [pdf] Available at: https://dl.acm.org/doi/pdf/10.1145/3233231 [Accessed on 08 November 2024].
Luqman, M., Rayner, P.J. and Gurney, K.R., 2023. *On the impact of urbanisation on CO_2 emissions*. *npj Urban Sustainability*, 3, p.6. [pdf] Available at: https://www.nature.com/articles/s42949-023-00084-2.pdf [Accessed on 03 November 2024]
M. Gray, S. Suri, 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston, New York: Houghton Mifflin Harcourt
Mazzucato, M. (2024). *The ugly truth behind ChatGPT: AI is guzzling resources at planet-eating rates. [online]* Available at: https://www.theguardian.com/commentisfree/article/2024/may/30/ugly-truth-ai-chatgpt-guzzling-resources-environment [Accessed on 08 November 2024].
Mazzucato, M. and Collington, H., (2023). *The Big Con: How the Consulting Industry Weakens our Businesses, Infantilizes our Governments and Warps our Economies*. London: Allen Lane.
McCarthy J., et. al, 1955. *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* [online] Available at: https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html [Accessed on 20 March 2024].
McQuillan D., 2019. *Non-Fascist AI.* [online] Available at: https://osf.io/preprints/socarxiv/b64sw/download [Accessed on 17 March 2024]
McQuillan, D. (2022). *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*. Bristol University Press, Bristol.
Merchant, B. (2023). *Blood in the Machine: The Origins of the Rebellion Against Big Tech*. Little, Brown and Company, New York.
Miao F., Holmes W., Huang R., Zhang H., 2021. *AI and education: Guidance for policy-makers*. [online] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000376709 [Accessed on 18 March 2024]
Microsoft (n.d.). *AI for Good*. [online] Available at: https://www.microsoft.com/en-us/ai/ai-for-good [Accessed on 29 March 2024]
Ministère de la Communication, des Télécommunications et de l'Économie Numérique (2023). *Stratégie Nationale et Feuille de Route du Sénégal sur l'Intelligence Artificielle (Version Résumée)*. [pdf] Available at: https://drive.google.com/file/d/1kXmzNmOayAHo4ESE2rGEh2AaUTZMA-u2/view [Accessed on 01 June 2024]
Ministero dell'Ambiente e della Sicurezza Energetica, (2022). *Strategia Nazionale per lo Sviluppo Sostenibile 2023: Strategia e allegati.* [pdf] Available at: https://www.mase.gov.it/sites/default/files/archivio/allegati/sviluppo_sostenibile/ALL1_SNSvS_2023_Strategia_e_allegati.pdf [Accessed on 18 November 2024]
Ministry of Economy, Trade and Industry (METI), (2021). *AI Implementation in Society*. [pdf] Available at: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf [Accessed on 01 June 2024]
Ministry of Finance, Ghana, (2024). *Ghana Charts Collaborative Path to Catalyse AI for Africa's Development*. [online] Available at: https://mofep.gov.gh/news-and-events/2024-03-19/ghana-charts-collaborative-path-to-catalyse-ai-for-africas-development [Accessed on 01 June 2024]
Ministry of Foreign Affairs of Japan, (2024). Hiroshima Process: International Guiding Principles for Organizations Developing Advanced AI Systems. [pdf] Available at: https://www.mofa.go.jp/files/100573471.pdf [Accessed on 04 June 2024]
MIT Media Lab, (2024). *Discrimination by Proxy. AI Blindspot. [online]* Available at: https://aiblindspot.media.mit.edu/discrimination_by_proxy.html [Accessed 21 October 2024].
Mohamed, S., Png, M.-T., & Isaac, W. (2020). *Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology*, 33(4), 659–684. [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s13347-020-00405-8.pdf [Accessed on 08 November 2024]
Moore, J. (2019). *AI for Not Bad*. Frontiers in Big Data, 2. [doi] Available at: https://doi.org/10.3389/fdata.2019.00032 [Accessed on 29 March 2024]
Morozov E., 2023. *The problem with artificial intelligence? It's neither artificial nor intelligent*. [online] Available at: https://www.theguardian.com/commentisfree/2023/mar/30/artificial-intelligence-chatgpt-human-mind
Morozov, E., (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
Morris, M., (2022), *AI shopping startup exaggerated tech capabilities to potential investors*. [online] Available at: www.theinformation.com/articles/shaky-tech-and-cash-burning-giveaways-ai-shopping-startup-shows-excesses-of-funding-boom [Accessed on 09 November 2024]
Motalebi N., Verity A., 2023. *Generative AI for Humanitarians.* [pdf] Available at: https://reliefweb.int/attachments/fe1baf61-33a6-4f55-955d-

b977e8a76937/Generative%20AI%20for%20Humanitarians%20-%20September%202023.pdf
[Accessed on 24 May 2024]
Narayanan A., 2019. *How to recognize AI snake oil.* [pdf] Available at: https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf [Accessed on 17 march 2024]
National Academies of Sciences, Engineering, and Medicine (2022: 53) *Advancing Climate Action: Insights from AI Technology*. Washington, D.C.: National Academies Press, Chapter 5. Available at: https://nap.nationalacademies.org/read/26641/chapter/5 (Accessed: 05 November 2024).
National Centre for Social Research, (2022). *Evaluation of Third Space Learning's Affordable Maths Tuition*. [pdf] Available at: https://whatworks-csc.org.uk/research-report/pilot-evaluation-of-affordable-maths-tuition/ Accessed on 04 August 2024]
Nederveen Pieterse, J. (2010). *Development Theory: Deconstructions/Reconstructions*. 2nd edn. Sage Publications, London. (Originally published in 2001)
Newman A., 2019. *I Found Work on an Amazon Website. I Made 97 Cents an Hour.* [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed on 21 March 2024].
Newman A., 2019. *I Found Work on an Amazon Website. I Made 97 Cents an Hour.* [online] Available at: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html [Accessed on 01 June 2024]
Nguyen D., et. al., 2024. *Patient Characteristics Impact Performance of AI Algorithm in Interpreting Negative Screening Digital Breast Tomosynthesis Studies*. [online] Available at: https://pubs.rsna.org/doi/10.1148/radiol.232286 [Accessed on 01 June 2024]
Nichols, S. (2022). *Metaverse Rollout Brings New Security Risks, Challenges*. [online] Available at: https://www.techtarget.com/searchsecurity/news/252513072/Metaverse-rollout-brings-new-security-risks-challenges?utm_campaign=20220209_Metaverse+brings+new+security+challenges+to+businesses%3B+Plus%2C+manual+vs.+automated+pen+testing&utm_medium=EM&utm_source=NLN&track=NL- [Accessed on 04 August 2024]
NITI Aayog (2023). *National Strategy for Artificial Intelligence*. [pdf] Available at: https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf [Accessed on 01 June 2024]
NITI Aayog, (2021). *Responsible AI*. [pdf] Available at: https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf [Accessed on 01 June 2024]
Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, New York.
OECD, (n.d.). *AI National Plan of Argentina*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26935 [Accessed on 01 June 2024]
OECD, (n.d.). *AI National Policy, Chile*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24840 [Accessed on 01 June 2024]
OECD, (n.d.). *Australia's AI Action Plan*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26948 [Accessed on 01 June 2024]
OECD, (n.d.). *Brazilian AI National Strategy*. [online] Available at: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-27104 [Accessed on 01 June 2024]
OECD, (n.d.). *OECD AI Principles*. [online] Available at: https://oecd.ai/en/ai-principles [Accessed on 04 June 2024]
OECD, 2019. *Recommendation of the Council on Artificial Intelligence*. [online] Available at: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 [Accessed 15 March 2024]
Office of the High Commissioner for Human Rights (OHCHR) (n.d.). *Economic, Social and Cultural Rights*. [online] Available at: https://www.ohchr.org/en/human-rights/economic-social-cultural-rights [Accessed on 29 March 2024]
Omiye I. J., et. al., 2023. *Large language models propagate race based medicine*. [online] Available at: https://www.nature.com/articles/s41746-023-00939-z [Accessed on 01 June 2024]
Osseiran, N., Asher-Schapiro, A. and Farouk, M., (2024). *In Middle East, poor excluded from welfare by 'faulty' algorithms.* [online] Available at: https://www.japantimes.co.jp/news/2023/10/05/world/society/middle-east-poor-algorithms-aid/ [Accessed 21 October 2024].
Othot, (n.d.) *Enrollment Management*. [onl;ine] Available at: https://www.othot.com/products/enrollment-management [Accessed on 04 August 2024]
Petersen, R., Weisse, M., Lewandowski, E., Swartz, T., and Wang, L., (2018). *Artificial Intelligence Helps Distinguish the Forest From the Trees: Part 1*. [online] Available at: https://www.globalforestwatch.org/blog/data/artificial-intelligence-helps-distinguish-the-forest-from-the-trees-part-1/?utm_campaign=gfw&utm_source=gfwblog&utm_medium=hyperlink&utm_term=orbitalinsights_11_2018 [Accessed on 03 November 2024].
Padhiary, M., Saha, D., Kumar, R., Sethi, L.N., and Kumar, A., (2024). *Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicles for farm automation*. *Advanced Technology*, p.100483. [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S2772375524000881/pdfft?md5=526fc481832db8786364db930ddb071e&pid=1-s2.0-S2772375524000881-main.pdf [Accessed on 03 November 2024]
Pangburn, D. J., (2019). *Schools are using AI to help pick students. What could go wrong?* [online] Available at: https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong [Accessed on 04 August 2024]
Pasek, A., Vaughan, H., & Starosielski, N. (2023). *The world wide web of carbon: Toward a relational footprinting of information and communications technology's climate impacts*. Big Data & Society, 10(1). [doi] Available at: https://doi.org/10.1177/20539517231158994 [Accessed on 08 November 2024]

Pasquale, F., 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
Pasquinelli, M. (2023). *The Eye of the Master*. *A Social History of Artificial Intelligence.* Verso.
Perrigo B., 2023. *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. [online] Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed on 01 June 2024]
Perrigo, B., (2023). *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. *[online]* Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/ [Accessed 09 November 2024].
Prabhu, V. U. & Birhane, A., (2020). *Large image datasets: A pyrrhic win for computer vision?* [pdf] Available at: https://arxiv.org/abs/2006.16923 [Accessed on 08 November 2024]
Rainforest Connection, (n/d). *Guardian Platform*. [online] Available at: https://rfcx.org/guardian [Accessed on 03 November 2024]
Rechtbank Den Haag, (2020). 'ECLI:NL:RBDHA:2020:865'. *District Court of The Hague*. [online] Available at: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865 [Accessed on 14 October 2024]
Redazione Trento, (2024). *Progetti Marvel e Protector, il comune condannato a pagare 25 mila euro*. La Voce del Trentino. [online] Available at: https://www.lavocedeltrentino.it/2024/01/24/progetti-marvel-e-protector-il-comune-condannato-a-pagare-25-mila-euro/ [Accessed on 08 November 2024].
Regione Abruzzo, (2020). *Initiatives and regional planning.* [online] Available at: https://www.regione.abruzzo.it/content/initiatives-and-regional-planning [Accessed on 05 November 2024]
Regione Emilia Romagna, (2024). *Alluvione, un anno dopo: il primo pensiero alle vittime. Il punto su quanto fatto dalla Regione*. [online] Available at: https://www.regione.emilia-romagna.it/notizie/2024/maggio/alluvione-un-anno-dopo [Accessed on 01 November 2024]
Riddell, R., et al., (2024). *Inequality Inc. How corporate power divides our world and the need for a new era of public action.* [pdf] Available at: https://oi-files-d8-prod.s3.eu-west-2.amazonaws.com/s3fs-public/2024-01/Davos%202024%20Report-%20English.pdf [Accessed on 14 October 2024]
Roio, D. (2024). *Algorithmic Sovereignty*. PhD dissertation. Plymouth University. [pdf] Available at: https://pearl.plymouth.ac.uk/foahb-theses-other/79/ [Accessed on 01 June 2024]
Rolnick, D., et al. (2019). *Tackling climate change with machine learning. Nature Climate Change*, 9(7), 498-504. [pdf] Available at: https://arxiv.org/pdf/1906.05433 [Accessed on 01 November 20234]
Rozite, V., Miller, J., Oh S., (2023). *Why AI and energy are the new power couple.* [pdf] Available at: https://www.digital-energy.ru/wp-content/uploads/2024/07/IEA_Why_AI.pdf [Accessed on 01 November 2024]
Rush, C., (2021). *Smart Building Solutions for Manufacturing Facilities.* [online] Available at: https://knowhow.distrelec.com/energy-and-power/smart-building-solutions-for-manufacturing-facilities/ [Accessed on 03 November 2024]
Russell S., Perset K., Grobelnik M., 2023. *Updates to the OECD's definition of an AI system explained.* [online] Available at: https://oecd.ai/en/wonk/ai-system-definition-update [Accessed on 15 March 2024]
Rutjens, B. T., Sengupta, N., der Lee, R. van, van Koningsbruggen, G. M., Martens, J. P., Rabelo, A., & Sutton, R. M. (2022). *Science Skepticism Across 24 Countries*. *Social Psychological and Personality Science*, 13(1), 102-117. [doi] Available at: https://doi.org/10.1177/19485506211001329 [Accessed on 09 November 2024]
Salesforce, (n.d.). *Education Recruitment and Admissions Software*. [online] Available at: https://www.salesforce.com/education/recruitment-admissions-software/ [Accessed on 04 August 2024]
Salganik, M. J., Lundberg, I., Kindel, A. T., & McLanahan, S., (2019). *Introduction to the Special Collection on the Fragile Families Challenge.* Socius, 5. [doi] Available at: https://doi.org/10.1177/2378023119871580 [Accessed on 04 August 2024]
Sanchez, T. W., Shumway, H., Gordner, T., & Lim, T. (2022). The prospects of artificial intelligence in urban planning. *International Journal of Urban Sciences*, *27*(2), 179–194. [online] Available at: https://doi.org/10.1080/12265934.2022.2102538 [Accessed on 03 November 2024]
Sardar, Z. (2009). *Welcome to Postnormal Times*. *Futures*, 42(5), pp. 435–444. [doi] Available at: https://doi.org/10.1016/j.futures.2009.11.028 [Accessed on 18 November 2024]
Save the Children, (2023). *2023 in review: climate disasters claimed 12,000 lives globally in 2023.* [online] Available at: https://www.savethechildren.net/news/2023-review-climate-disasters-claimed-12000-lives-globally-2023 [Accessed on 30 October 2024]
Scheuerman M., et. al., 29019. *How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis and Image Labeling Services*. [online] Available at: https://dl.acm.org/doi/10.1145/3359246 [Accessed on 01 June 2024)
Schulze, E., (2019), *40% of A.I. start-ups in Europe have almost nothing to do with A.I., research finds*. [online] Available at: www.cnbc.com/2019/03/06/40-percent-of-ai-start-ups-in-europe-not-related-to-ai-mmc-report.html [Accessed on 09 November 2024]
Sclater, N. (2016). *Developing a Code of Practice for Learning Analytics*. Journal of Learning Analytics, 3(1), pp. 16–42. [online] Available at: https://www.researchgate.net/publication/307842715_Developing_a_Code_of_Practice_for_Learning_Analytics [Accessed on 04 August 2024]
Scott-Smith T., 2016. *Humanitarian neophilia: the 'innovation turn' and its implications. Third World Quarterly*, *37*(12), 2229–2251. [doi] Available at: https://doi.org/10.1080/01436597.2016.1176856 [Accessed on 14 May 2024]
Sefala R., Gebru T., Mfupe L., Moorosi N., Klein R., (2021). *Constructing a Visual Dataset to Study the Effects of Spatial Apartheid in South Africa. Constructing a Visual Dataset to Study the Effects of Spatial Apartheid in South Africa', Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS).* [pdf] Available at: https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/07e1cd7dca89a1678042477183b7ac3f-Paper-round2.pdf [Accessed on 17 October 2024]
Selwyn, N 2016, *Is Technology Good for Education?* Polity Press, Cambridge UK. Available at: http://au.wiley.com/WileyCDA/WileyTitle/productCd-0745696465.html [Accessed on 04 August 2024]

Selwyn, N. (2022). *The Future of AI and Education: Some Cautionary Notes*. European Journal of Education, 57(4), pp. 620–631. [online] Available at: https://eric.ed.gov/?id=EJ1355024 [Accessed on 30 July 2024]
Sen, A., (1995). *The Political Economy of Targeting*. In D. van de Walle & K. Nead (Eds.), *Public Spending and the Poor: Theory and Evidence* (pp. 11–24). [pdf] Available at: https://adatbank.ro/html/cim_pdf384.pdf [Accessed 21 October 2024].
Shi, Z.R., Wang, C. & Fang, F. (2020). *Artificial Intelligence for Social Good: A Survey*. [pdf]. Available at: https://doi.org/10.48550/arXiv.2001.01818 [Accessed on 29 March 2024]
Siemens, (n/d). *Leveraging Artificial Intelligence in Sustainable Digital Enterprise*. [online] Available at: https://www.siemens.com/global/en/company/insights/leveraging-artificial-intelligence-in-a-sustainable-digital-enterprise.html [Accessed on 03 November 2024]
Simonite T., 2021. *This Researcher says AI is Neither Intelligent nor Artificial.* [online] Available at: https://www.wired.com/story/researcher-says-ai-not-artificial-intelligent/ [Accessed on 17 March 2024]
Singh, A., Karayev, S., Gutowski, K. & Abbeel, P. (2017). *Gradescope: A Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work*. In: Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale. [online] Available at: https://www.researchgate.net/publication/317003983_Gradescope_A_Fast_Flexible_and_Fair_System_for_Scalable_Assessment_of_Handwritten_Work [Accessed on 04 August 2024]
Solow-Niederman A., 2024. *Can AI Standards Have Politics?* [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4714812 [Accessed on 01 June 2024]
Stanford University (2017). *Full Translation: China's New Generation Artificial Intelligence Development Plan (2017)*. [online] Available at: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ [Accessed on 01 June 2024]
Stanford University's Human-Centered AI, (2024). *AI Index Report 2024*. [pdf] Available at: https://aiindex.stanford.edu/report/ [Accessed on 13 May 2024]
Statistics South Africa, 2023. *Quarterly Labour Force Survey: Quarter 1, 2023*. Available at: https://www.statssa.gov.za/publications/P0211/P02111stQuarter2023.pdf [Accessed on 17 October 2024].
Stefan Feuerriegel S., HartmannJ., Janiesch C, Zschech P., 2023. *Generative AI.* [pdf] Available at: https://link.springer.com/content/pdf/10.1007/s12599-023-00834-7.pdf [Accessed on 17 March 2024]
STOP LAPDS Spying Coalition, 2020. *The Algorithmic Ecology: An Abolitionist Tool for Organizing Against Algorithms*. [online] Available at: https://freerads.org/2020/03/02/the-algorithmic-ecology-an-abolitionist-tool-for-organizing-against-algorithms/ [Accessed on 01 June 2024]
Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics. [pdf] Available at: https://aclanthology.org/P19-1355.pdf [Accessed on 08 November 2024]
Suel, E., Polak, J.W., Bennett, J.E., Ezzati, M., (2019). Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports*, 9, p.6229. [online] Available at: https://doi.org/10.1038/s41598-019-42036-w [Accessed on 14 October 2024]
Sukel, M., (2021). *An introduction to Object Detection Kit*. [online] Available at: https://amsterdamintelligence.com/posts/an-introduction-to-object-detection-kit [Accessed on 03 November 2024]
The Alan Turing Institute. *Data Science and AI Glossary.* [online] Available at: https://www.turing.ac.uk/news/data-science-and-ai-glossary [Accessed on 15 March 2024]
The United Nations Office for Disasters Risk Reduction, (2015). *The Human Cost of Weather Related Disasters.* [pdf] Available at: https://www.preventionweb.net/files/46796_cop21weatherdisastersreport2015.pdf [Accessed on 30 October 2024]
Third Space Learning (n.d.). *How It Works*. [online] Available at: https://thirdspacelearning.com/how-it-works/ [Accessed on 04 August 2024]
Third Space Learning, (n/d). *Data Protection & Privacy Policy.* [pdf] Available at: https://thirdspacelearning.com/data-protection-privacy-policy/ [Accessed on 04 August 2024]
This Person Does Not Exist, (n.d). [online] Available at: https://thispersondoesnotexist.com/ [Accessed on 17 March 2024]
Translators Without Borders (n.d.). *Chatbot Release in Northeast Nigeria*. [online] Available at: https://translatorswithoutborders.org/chatbot-release-northeast-nigeria/ [Accessed: 24 May 2024]
Tsanni A., (2024). *How satellite images and AI could help fight spatial apartheid in South Africa* . [online] Available at: https://www.technologyreview.com/2024/01/19/1086837/satellite-images-ai-spatial-apartheid-south-africa/ [Accessed on 17 October 2024]
Tsanni A., 2023. *This company is building AI for African languages.* [online] Available at: https://www.technologyreview.com/2023/11/17/1083637/lelapa-ai-african-languages-vulavula/ [Accessed on 14 October 2024]
Turing A., 1947. *Lecture on the Automatic Computing Engine*, in *The Essential Turing,* ed. B. Jack Copeland, London: Clarendon Press, 2004.
Turnitin, (n.d.). *Turnitin: Integrity in Every Assignment*. [online] Available at: https://www.turnitin.com/ [Accessed on 04 August 2024]
UK Government, (2023). *A Pro-Innovation Approach to AI Regulation: Government Response*. [online] Available at: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response [Accessed on 01 June 2024]
UK Government, (2023). *AI Regulation: A Pro-Innovation Approach – White Paper*. [online] Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [Accessed on 01 June 2024]
UN Environmental Programme, (2023). *Broken Record: Temperatures hit new heights, yet world fails to cut emissions (again).* [pdf] Available at: https://wedocs.unep.org/bitstream/handle/20.500.11822/43922/EGR2023.pdf?sequence=3&isAllowed=y [Accessed on 30 October 2024]
UN OCHA, Centre for Humdata, 2024. *Briefing note on Artificial Intelligence and Humanitarian Sector*. [pdf] Available at: https://centre.humdata.org/note-briefing-note-on-artificial-intelligence-and-the-humanitarian-sector/ [Accessed on 24 May 20234]

UN OCHA, n.d. Predictive Analytics. [online] Available at: https://centre.humdata.org/glossary-2/predictive-analytics/ [Accessed on 24 May 2024]
UNESCO, (2019). *Beijing Consensus on Artificial Intelligence in Education*. [pdf] Available at: https://unesdoc.unesco.org/ark:/48223/pf0000368303 [Accessed on 04 June 2024]
UNESCO, (n.d.). *Digital Education and Artificial Intelligence*. [online] Available at: https://www.unesco.org/en/digital-education/artificial-intelligence [Accessed on 04 June 2024]
UNHCR (n.d.). *Biometric Identity Management System*. [online] Available at: https://www.unhcr.org/media/biometric-identity-management-system & https://www.unhcr.org/blogs/unhcrs-biometric-tools-in-2023/ [Accessed on 24 May 2024]
UNHCR (n.d.). *Jetson: Predictive Analytics Tool*. [online] Available at: https://jetson.unhcr.org/ [Accessed on 24 May 2024]
UNICEF (n.d.). *Blockchain*. [online] Available at: https://www.unicef.org/innovation/blockchain [Accessed on 29 March 2024]
UNICEF (n.d.). *Drones*. [online] Available at: https://www.unicef.org/innovation/drones [Accessed on 29 March 2024]
UNICEF, (2021). *Policy Guidance on AI for Children*. [pdf] Available at: https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf [Accessed on 04 June 2024]
UNICEF, 2021. *Policy guidance on AI for children.* [online] Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Accessed on 15 March 2024] https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf [Accessed on 15 March 2024]
United Nations (2022). *The Sustainable Development Goals Report 2022*. [pdf] Available at: https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf [Accessed on 15 May 2024]
United Nations (2023). *The Sustainable Development Goals Report 2023: Special Edition*. [pdf] Available at: https://desapublications.un.org/publications/sustainable-development-goals-report-2023-special-edition [Accessed on 15 May 2024]
United Nations Conference on Trade and Development (UNCTAD), (2023). *Technology and Innovation Report 2023.* [pdf] Available at: https://unctad.org/system/files/official-document/tir2023_en.pdf (Accessed on 29 March 2024)
United Nations Development Programme (UNDP) (n.d.). *Artificial Intelligence at UNDP*. [online] Available at: https://www.undp.org/digital/ai [Accessed on 04 June 2024]
United Nations Development Programme (UNDP) (n.d.). *Making AI Work for Us*. [online] Available at: https://feature.undp.org/making-ai-work-for-us/ [Accessed on 29 March 2024]
United Nations, (2022). *Principles for the Ethical Use of AI in the UN System*. [pdf] Available at: https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf [Accessed on 04 June 2024]
United Nations, (2024). *The Sustainable Development Goals Report 2024*. [pdf] Available at: https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf [Accessed on 14 October 2024]

United Nations, 2024. *The Sustainable Development Goals Report 2024*. [pdf] Available at: https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf [Accessed on 30 July 2024]
United States Congress, (2020). *Congressional Report on AI Regulation*. [pdf] Available at: https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210 [Accessed on 01 June 2024]
University of Memphis, (n.d.). *AutoTutor Project*. [online] Available at: https://www.memphis.edu/iis/projects/autotutor.php [Accessed on 04 August 2024]
Uppsala University (n.d.). *VIEWS: Violence Early-Warning System*. [online] Available at: https://www.uu.se/en/department/peace-and-conflict-research/research/views/ [Accessed: 24 May 2024]
UrbanistAI, (n/d). [online] Available at https://site.urbanistai.com/ [Accessed on 03 November 2024]
Ushahidi blog, 2010. *Crisis Mapping Haiti: Some Final Reflections.* [online] Available at: https://www.ushahidi.com/about/blog/crisis-mapping-haiti-some-final-reflections/ [Accessed on 14 May 2024]
Ushahidi blog, 2012. *Haiti and the Power of Crowdsourcing*. [online] Available at: https://www.ushahidi.com/about/blog/haiti-and-the-power-of-crowdsourcing/ [Accessed on 14 May 2024]
van Brakel, R. (2017). *Big Data Surveillance: The Case of Policing. Surveillance & Society*, 14(1), 1–15. [pdf] Available at: https://www.asanet.org/wp-content/uploads/attach/journals/oct17asrfeature.pdf [Accessed on 08 November 2024].
van Brakel, R., (2020). *Dutch Court Rules that SyRI Welfare Fraud Detection System Violates Human Rights.* [online] Available at: https://edri.org/our-work/dutch-court-rules-that-syri-welfare-fraud-detection-system-violates-human-rights/ [Accessed on 14 October 2024]
Većkalov, B., van Stekelenburg, A., van Harreveld, F., & Rutjens, B. T. (2023). *Who Is Skeptical About Scientific Innovation? Examining Worldview Predictors of Artificial Intelligence, Nanotechnology, and Human Gene Editing Attitudes*. *Science Communication*, 45(3), 337-366. [doi] Available at: https://doi.org/10.1177/10755470231184203 [Accessed on 09 November 2024]
Verity, A. & Wright, J. (2020). *Artificial Intelligence Principles for Vulnerable Populations in Humanitarian Contexts*. Digital Humanitarian Network. [pdf] Available at: https://www.academia.edu/41716578/Artificial_Intelligence_Principles_For_Vulnerable_Populations_in_Humanitarian_Contexts [Accessed on 01 June 2024].
Verma, J., Sandys, L., Matthews, A., & Goel, S., (2024). *Readiness of artificial intelligence technology for managing energy demands from renewable sources.* [pdf] Available at: https://www.sciencedirect.com/science/article/pii/S0952197624009898?via%3Dihub [Accessed on 01 November 2024]
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Nerini, F. F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233. [online] Available at: https://www.nature.com/articles/s41467-019-14108-y [Accessed on 08 November 2024].
Watson, J.B. (1998). *Behaviorism*. 1st edn. Routledge, New York.

Watters, A. (2021). *Teaching Machines: The History of Personalized Learning*. MIT Press, Cambridge, MA.
Wei, Cui., Zhen, Xue., Khanh-Phuong, Thai. (2018). Performance Comparison of an AI-Based Adaptive Learning System in China. [online] Available at: https://ieeexplore.ieee.org/document/8623327 [Accessed on 30 July 2024]
Weinberg, M.A., 1972. *Science and Trans-Science*. Volume 177, Number 4045. American Association for the Advancement of Science.
Welsh T., 2019. *Biometrics disagreement leads to food aid suspension in Yemen.* [onl;ine] Available at: https://www.devex.com/news/biometrics-disagreement-leads-to-food-aid-suspension-in-yemen-95164 [Accessed on 01 June 20234]
WeRobotics (n.d.). *Watch These Cargo Drones Bring Essential Health Services to Remote Communities*. [online] Available at: https://werobotics.org/blog/watch-these-cargo-drones-bring-essential-health-services-to-remote-communities/ [Accessed on 24 May 2024]
White House ,(n.d.). *AI Bill of Rights*. [online] Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [Accessed on 01 June 2024]
White House, (2023). *Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*. [online] Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [Accessed on 01 June 2024]
Whitmee, S., Anton, B., & Haines, A., (2023). *Accountability for carbon emissions and health equity*. *The Lancet Planetary Health*, 7(1), pp. e20–e26. [pdf[Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9874368/pdf/BLT.22.289452.pdf [Accessed on 05 November 2024]
WildMe, (n/d). [online] Available at: https://www.wildme.org/ [Accessed on 03 November 2024]
Wiley, (n.d.). *Alta*. [online] Available at: https://www.wiley.com/en-it/education/alta [Accessed on 04 August 2024]
William S., 1998. *Population Statistics, the Holocaust, and the Nuremberg Trials,* Population and Development Review
Williamson, B. & Hogan, A., (2020). *Commercialisation and Privatisation in/of Education in the Context of Covid-19*. Education International, Brussels, Belgium. Available at: https://eprints.qut.edu.au/216577/1/76301358.pdf [Accessed on 30 July 2024]
Williamson, B., (2017). *Big Data in Education: The Digital Future of Learning, Policy, and Practice*. Sage Publications.
Winner, L. (1977). *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA: MIT Press.
Wood., J., (2023). *This AI helps buildings cool themselves and cut emissions.* [online] Available at: https://www.weforum.org/stories/2023/10/ai-buildings-heating-cooling-carbon-brainbox/ [Accessed on 03 November 2024]
World Bank, (2022). *Inequality in Southern Africa: An Assessment of the Southern African Customs Union*. Washington, DC: World Bank. [pdf] Available at: https://documents1.worldbank.org/curated/en/099125003072240961/pdf/P1649270b73f1f0b5093fb0e644d33bc6f1.pdf [Accessed on 17 October 2024]

World Economic Forum, (n.d.). *AI Governance Alliance*. [online] Available at: https://initiatives.weforum.org/ai-governance-alliance/home [Accessed on 04 June 2024]
World Food Programme (WFP) (n.d.). *HungerMap*. [online] Available at: https://hungermap.wfp.org/ [Accessed on 24 May 2024]
World Food Programme (WFP) (n.d.). *SCOPE Cash Accounts User Manual*. [online] Available at: https://usermanual.scope.wfp.org/cash-accounts/content/common_topics/introduction/1_introduction.htm & https://scope.wfp.org/login/?next=/ [Accessed on 24 May 2024]
World Intellectual Property Organisation, n/d. *Frontier Technologies Fact-sheet.* [pdf] Available at: https://www.wipo.int/about-ip/en/frontier_technologies/pdf/frontier-tech-6th-factsheet.pdf (Accessed on 29 March 2024)
World Meteorological Organization (WMO), (2023). *Climate Change Indicators Reached Record Levels in 2023*. [online] Available at: https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-wmo#:~:text=2023%20was%20the%20warmest%20year%20in%20the%20174%2Dyear%20observational,above%20the%201850%E2%80%931900%20average [Accessed on 15 May 2024]
World Meteorological Organization (WMO), (2023). *Climate Change Indicators Reached Record Levels in 2023*. [online] Available at: https://wmo.int/news/media-centre/climate-change-indicators-reached-record-levels-2023-wmo#:~:text=The%20WMO%20report%20confirmed%20that,ten%2Dyear%20period%20on%20record [Accessed on 30 October 2024]
Yale Climate Connections, (2023). *Heat contributed to 47,000 deaths in Europe during summer 2023, study finds.* [online] Available at: https://yaleclimateconnections.org/2024/11/heat-contributed-to-47000-deaths-in-europe-during-summer-2023-study-finds/#:~:text=The%20summer%20of%202023%20was,47%2C000%20deaths%20across%20the%20continent. [Accessed on 30 October 2024]
Yampolskiy, R. V., (2019). *Unexplainability and Incomprehensibility of Artificial Intelligence*. Computer Engineering and Computer Science, University of Louisville. [pdf] Available at: https://arxiv.org/abs/1907.03869 [Accessed on 08 November 2024].
Yazdinejad, A., Zolfaghari, B., Azmoodeh, A., Dehghantanha, A., Karimipour, H., Fraser, E., Green, A.G., Russell, C., and Dunn, E., 2021. *A Review on Security of Smart Farming and Precision Agriculture: Security Aspects, Attacks, Threats and Countermeasures. Journal of Sensor and Actuator Networks*, 10(4), p.49. [online] Available at: https://www.mdpi.com/2076-3417/11/16/7518 [Accessed on 03 November 2024]
Zhao, W., Li, T., Qi, B., Nie, Q. and Runge, T., (2021). *Terrain Analytics for Precision Agriculture with Automated Vehicle Sensors and Data Fusion*. *Sustainability*, 13(5), p.2905. [online] Available at: https://www.mdpi.com/2071-1050/13/5/2905 [Accessed on 03 November 2024]
Zipline (n.d.). *Fly Zipline*. [online] Available at: https://www.flyzipline.com/ [Accessed on 24 May 2034]
Zuboff, S., (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

510 Global (2021). *Impact-Based Forecasting for Typhoons in the Philippines*. [online] Available at: https://510.global/2021/10/impact-based-forecasting-for-typhoons-in-the-philippines/ [Accessed: 24 May 2024]

A Project of

impactskills

UNIVERSITÀ DI TORINO

Dipartimento di Culture, Politica e Società

MANCHESTER 1824
The University of Manchester

With the Support of

Ministero degli Affari Esteri e della Cooperazione Internazionale