Sustainable Artificial Intelligence: A Perspective on Society, Ethics, and the Environment

Chloe M. Barnes[0000−0002−6782−1773] and
Jordan J. Bird[0000−0002−9858−1231] and
Anikó Ekárt[0000−0001−6967−5397]

Abstract The rise of sociotechnical systems means that we and our systems are more interconnected than ever before. This is especially true in sociotechnical systems that incorporate Artificial Intelligence (AI), where systems can make decisions to achieve their goals that may unintentionally positively or negatively impact other entities if they are unable to consider the wider impact of their actions on other systems, people, and the environment. These unintended interactions, or unintended consequences of their actions, can stem from colocated systems pursuing conflicting or competing goals, a lack of awareness or consideration for the embodied societal or environmental cost of the system, or even through enabling AI systems to make potentially detrimental decisions or actions. Often, discussions on sustainability focus on the environment; however, given the prominence of sociotechnical systems, sustainability in the context of AI is now more than just this; although sustainable AI encompasses its effects on the environment that manifest over time, we argue that the sustainable AI systems of the future have an immediate effect on our social fabric that must be considered in their design and operation.

Key words: Sustainability, AI and Society, Sociotechnical Systems, Technological Ethics

---

Chloe M. Barnes
Department of AI & Robotics, ACAIRA, Aston University, Birmingham, UK
e-mail: `c.barnes1@aston.ac.uk`

Jordan J. Bird
Department of Computer Science, Nottingham Trent University, Nottingham, UK
e-mail: `jordan.bird@ntu.ac.uk`

Anikó Ekárt
Department of AI & Robotics, ACAIRA, Aston University, Birmingham, UK
e-mail: `a.ekart@aston.ac.uk`

1 The Rise of AI in Sociotechnical Systems

The term 'Artificial Intelligence' (AI) has transcended the laboratory into the public domain in recent years, with AI becoming increasingly integrated into the systems we use to support us on a daily basis – for example, voice assistants powered by AI such as Alexa or Siri [64]. The defining feature that distinguishes sociotechnical systems from technical systems is that they coexist with and are integrated into our society, therefore creating a two-way interaction between humans and systems; AI systems can therefore be defined as a subset of sociotechnical systems, with additional characteristics that incorporate artificial agents (in the case of AI, 'agent' refers to an artificial entity capable of interactivity, autonomy, and adaptability [33]), as well as adherence to social and technical rules [63]. As AI encroaches on every aspect of our human existence, the very fabric of our society is changing rapidly; this now brings the issue of how AI not only impacts society in the sociotechnical sense, but how AI systems – and our interaction with AI systems – impact the very environment in which we coexist.

This transcendence of the term AI has now grown beyond being a tool for exclusive use by academics and researchers, and has evolved into being used in everyday systems enabled by AI by the public that affect our activities both as individuals and as a society as a whole. At the societal level, AI systems have been introduced that have impact on the activities of the public in general, for example: using wearable sensors in healthcare applications, such as to support the care of the elderly [51]; assisting in the development and delivery of educational materials that can be individually tailored [61]; modelling and predicting urban traffic to inform city planning [37]; automating and supporting activities in the agri-food chain to promote sustainable practices and reduce waste [48, 55]; reducing dangers to life or injury in dangerous environments, such as through training simulations [15] and disaster management [74]; reducing repetitiveness and improving the efficiency of tasks in areas such as manufacturing and industry [19]; and generating or supporting the production of media to be consumed by the public, such as news and journalism [47, 62], or music, photos, and videos [43, 82]. Beyond the societal impact of AI where individuals may benefit from the institutional adoption of AI through education or healthcare, for example, individuals now have personal access to AI-enabled technologies to assist in their own daily activities and entertainment, such as: real-time access to information using natural language, due to the introduction and integration of Large Language Models (LLMs) such as OpenAI's GPT, Meta's LLaMA, and Google's Bard into personal devices and the internet [12, 66]; AI-enabled smartphones with features such as photo enhancement and Virtual Personal Assistants like Siri, Bixby or Google Assistant [18]; personalised, individual experiences in the form of recommended media such as films, TV shows, and music [45]; and targeted advertisements on online platforms and retail [49]. These innovations have been made possible through the method of *attention* in machine learning [80], which forms the basis of most modern Generative AI techniques. Classically, temporal techniques were used in natural language processing, which suffered from the open issue of more recent words being considered more important regardless of whether that is the case in reality. Attention mechanisms, on the other hand, are used to determine relative importance of tokens within the whole sequence (e.g. each word in a sentence). More recently, the concept of the Transformer network, the type of neural network behind the Large Language Model, has become a prominent technique due to its capability and efficiency compared with sequential approaches [84].

This widespread integration of AI into society now raises questions about trust, morals, and values [35, 63, 64] in the systems we use. Beyond the individual impact of AI, the societal implications of AI in the form of decision-making [32, 54], employment [49], and management of natural resources [1, 28] are evermore prevalent. Further questions arise regarding the negative environmental impacts versus the benefits of using such systems [50, 81, 85], since the advancement and adoption of AI technologies affect our sociotechnical environment both explicitly and implicitly. In response to the rapid invention and global adoption of AI systems, there is now a growing call to consider explainability, responsibility, ethics, transparency, and sustainability [4, 32, 64, 85] in their design and operation to ensure their trustworthiness and acceptance for the future.

This chapter explores the use of AI and the impact that these systems can have on society, other systems, and the environment in the context of sustainable design, practice, and usage. We argue that the sustainable AI systems of the future need to not only consider their environmental impact in their design and operation, but that it is now also increasingly important to consider the complex, and often unintended, consequences of their inter-/actions with both their local and global environments, and the individual and collective social impacts as well, due to the ubiquitous and interwoven nature of modern AI-driven sociotechnical systems. We therefore posit that sustainability in the context of AI encompasses the following attributes, which will be explored throughout this chapter:

Sustainable Operation

The increase in technical, sociotechnical, and AI systems means that more natural resources and energy are required to build, maintain, and operate these systems. Sustainable operation reflects the environmental costs and effects of using AI on a day-to-day basis, including both the embodied and operational costs of designing the system through to daily operation, as well as considering the effects on the environment itself through natural resource consumption.

Sustainable Societal Use

One of the benefits of using AI is that tasks or decisions can be completed faster compared to human efforts. As the use of AI and other sociotechnical systems prevails, we are increasingly interconnected as a society both through our data and actions – with AI facilitating this on an accelerating scale. Sustainable social use of AI reflects and acknowledges these new and often unknowable connections, ensuring that human experience, employment opportunities, and autonomy are preserved.

Sustainable Situatedness

Our sociotechnical environment is increasingly enriched with diversity in human individuals, technical systems, AI systems, and the environment. AI systems do not exist in isolation – they are interwoven into the complex network of connections between actors and entities within our sociotechnical environment. Sustainable situatedness acknowledges the intricacies of both the local operating environment and operation at a global social and environmental scale, with designers taking explicit action to ensure AI systems act responsibly considering their impacts on both society, individuals, and the environment.

Sustainable Acceptance

There is no doubt that AI has the potential to be harnessed for the good of society and the environment if designed considerately and responsibly, with an understanding of the system's place in the sociotechnical landscape. Sustainable acceptance relates to the public's perception, trust, and acceptance of these systems, through education and governance to empower individuals to make informed decisions based on transparency and explainability in the system. AI systems may have remarkable capabilities for enhancing our daily lives and the environment in which we co-exist, but awareness of the consequences of their actions at an individual level would provide the public with information to understand and accept these systems as a sustainable option to continue using in the future.

The remainder of this chapter explores these concepts and the use and impact of AI on our society and environment, and ethical concerns raised about its widespread adoption. Section 2 looks at the interconnectivity that we face through the use of AI, and the unintended consequences of its use in terms of *sustainable situatedness*. Section 3 explores the unintended impacts of AI in social, ethical and environmental contexts in the context of *sustainable societal use* and *sustainable operation*. Section 4 then discusses the societal implications of AI in terms of it being a viable and accepted tool to integrate into our society through the lens of *sustainable acceptance*, with conclusions drawn in Section 5.

2 Increasing Interconnectivity between AI and Society

Modern systems are not just interconnected with themselves; the sociotechnical systems of today have complex interactions with other 'intelligent' systems, the environment, society and individuals, and systems that are not 'intelligent' or 'sociotechnical' in the sense described in Section 1. The nature of our current and future sociotechnical environment means that actors situated within this network may affect each other, often in ways that are difficult to predict or understand. Sustainable situatedness in AI systems here refers to systems that acknowledge the complexities of the sociotechnical environment in which they operate, and their effects on both humans, other actors in the system, and the environment, with consideration for how AI systems influence both individuals and society as a whole.

2.1 Towards Knowing the Unknown

As a society, we are both influenced by the increasing use of AI, and connected with individuals across the globe more than ever before. The individual actions of humans or the systems we use can have vast effects across our society, and these are often hard to anticipate due to the scale and complex nature of the current sociotechnical landscape. We use ‘society’ in this sense to refer to our global community, as even those who do not directly use, or do not have access to AI systems are affected by the environmental implications of their usage worldwide [44], thus connecting individuals around the world. The institutional or commercial use of AI systems also impact individuals – even those who do not explicitly or intentionally engage with this technology personally. For example, it has been noted that AI systems employed by law enforcement to curb violent or antisocial behaviour can lead to a ‘self-fulfilling prophecy’ where the very behaviour the AI system intended to reduce in fact increased due to the discriminatory nature of the data used to create the system [4]. The actions of individuals, paired with biased and discriminatory data, thus demonstrates the impact that AI can have on our society when used to unintentionally connect individuals through data and potentially influence human behaviour as a result. The role that bias in AI plays in our society is discussed in more detail in Section 3.1.

The presence of AI does not only affect us through facilitating connections between individuals in our society – we are also affected by the interconnectivity of the systems that we use. Pringle et al. [65] noted that a DDoS attack in 2016 had a significant impact on US internet access, due to the increasing number of interconnected smart devices; this interconnectedness enabled the attack to propagate further and at a larger scale than previously seen, which would reasonably be expected to increase in future attacks in line with the growing number of smart devices and their interconnectedness. While Section 4.1 discusses emerging AI governance in the UK and EU, regulations are being introduced around the world to respond to the rapidly developing advances in AI technologies [39]. Huang et al. [39] list coordination between nations as imperative for AI regulations to be successful at the global level, as otherwise there may be vulnerabilities that create opportunities for further attacks that exploit the connectedness of our modern world. Pringle et al. [65] also consider the concept of ‘unknown unknowns’, in that the sociotechnical systems of today are increasingly interconnected; thus, the consequences of these systems interacting can be difficult to predict, since the greater the number of systems with the potential to interact, the greater the number of unintended and unpredictable interactions arise. There are a number of terms for these unintended interactions, for example ‘mutual influences’ [76], but we take the broader term ‘interference’ from the author’s previous work [7] to encompass the positive and negative, intended and unintended interactions that may arise in these scenarios.

van de Poel [63] notes three reasons that unintended consequences arise in the context of AI: a lack of care or anticipation of the unintended consequences, which also encompasses a lack of awareness of how the system may interfere with other entities within its situated environment; a lack of knowledge, differing from the first point in that it more closely resembles the concept of 'unknown unknowns' discussed above however stemming from ignorance [65]; and indeterminacy, where the unintended consequences are largely affected by external factors to the system, closely aligned with the concept of interference described above [7]. Coeckelbergh [20] suggested in 2021 that humanity is in a state of hyper-agency with regard to our influence and control over our society and environment. This state, dubbed the 'Anthropocene' [20, 24], is increasingly evident as a result of our recent scientific and technological advances. These advances have led to substantial increases in the human population, which in turn requires greater amounts of food, water, energy, and materials to sustain [24], thus highlighting the complex nature of our growth as a species and the environment in which we live. Although the term 'Anthropocene' was formally rejected as a unit of geological time in 2024 [41], this discussion still highlights the importance of considering the impact of our actions and directing this hyper-agency towards sustainable existence, and ensuring that the development and adoption of AI systems will not 'exacerbate the problem' [20] of harming our environment in the pursuit of technological advancement.

As the number and variety of AI systems integrated into our society increases, the occurrence of unintended interactions due to the interference of their operation on other systems and other people will increase – as well as both the environments the systems are situated in and the wider environment we inhabit together. The result of these unintended interactions of AI systems in the context of humanity's increasing hyper-agency therefore means that it is more important than ever before to understand the context in which AI systems interact in their local and global environment, with what systems or entities they interact with, and how they might interact. Do et al. [26] note that researchers tend to acknowledge and understand the importance of considering the societal and environmental impacts of their work arising from the unintended consequences of their operation; however, this acknowledgement rarely influences the technological design, and instead is often considered in retrospect. The findings of Bingley et al. [13] also support this line of thinking, in that the societal impacts of AI systems are generally not prioritised by designers, even though this is beneficial to user experiences. The concept of the Anthropocene age enforces the notion that the actions of humans, and consequently the systems that we use and create, are closely tied with our environment; thus, we have a responsibility to anticipate the 'unknown unknowns' and unintended consequences of using these systems that we increasingly rely on in every aspect of our lives. Although this may seem paradoxical, combatting the 'unknown unknowns' and mitigating the unintended consequences of the interactions of AI systems in the future should come from a greater understanding of the system's place in the social and environmental landscape it operates in, by anticipating how the system will operate when situated in the complex network of interconnected sociotechnical and environmental entities. The increasing interconnectivity of AI systems with both society and the environments they are situated in – both their local operating environment including other systems, and the broader sense of environment – means that this anticipation and understanding will be crucial moving forward if the use of AI is to be a sustainable option to integrate into our sociotechnical environment for the benefit of humanity.

2.2 AI and the Human Experience

Aside from the interconnectedness of modern AI systems, the use of AI affects individuals around the world both directly and indirectly, for example, through personal use, institutional or commercial use, or the unintended consequences of use in the sociotechnical environment. Relating to Coeckelbergh's notion of humanity's state of hyper-agency [20] as discussed in Section 2.1, in terms of an increasing need for control over our sociotechnical environment, Usmani et al. [79] highlight that *user* agency is an important quality when designing human-centred AI systems. User agency in this context is described as empowering users with both control and autonomy over their use of AI systems; this also includes the transparency and explainability of the system's behaviour to support trust [79]. This bidirectional relationship – e.g. a user's ability to communicate needs through control, and the system's ability to communicate outcomes through explainability to aid user's understanding – is also discussed by Bingley et al. [13], however, they find that users value when their needs are understood and satisfied when using AI more, compared to understanding how the AI system achieved the outcome. Whether through an individual user's desire for control over an AI system, or through the societal desire to influence the environment, the use of AI affects the human experience and how we experience the sociotechnical environment we inhabit.

AI is now also being used to find trends and connections between individuals in our society through their data, thus affecting individuals indirectly through institutional and commercial use of AI. For example, Duhigg [27] discusses how AI can be used in retail settings to predict and recommend vouchers for products that customers may be interested in, based on their purchase history, thus driving profits for the company and potentially inspiring customers to purchase products they might not necessarily have bought if left to their own free will. This may seem harmless, but Duhigg [27] describes an instance where a retailer sent vouchers to a teenager for pregnancy-related products, based on her spending habits, before it was known that she was pregnant. This example is also discussed by Corrigan et al. [23], outlining how data mining can be used to find relationships in vast quantities of data – in this case, finding patterns and relationships from the spending trends of large customer bases, such that life events can be predicted. From a societal perspective, the sustainable use of AI in consumer-facing industries moving forward requires ethical consideration – not just for customers, but for employees, too. Makridakis [54] highlights that the use of AI in retail and other establishments requiring skilled workers could negatively impact both employment and wealth distribution if companies replace or reduce the number of workers with AI systems; an example of this shift can be seen in the semi-automated Amazon Go chain of shops, where the 'convenience ideology' has eliminated the need for cashiers since AI systems both detect what items customers have taken while walking around the shop, and automatically bill customers when they leave [40]. The consequences of using AI as part of a data-driven or AI-enabled business strategy can be vast, affecting not just the customer-base, but the livelihoods of the workforce who may be replaced by these systems. If customer data is used to connect and make inferences from the data of individual customers and their spending habits, unintended consequences may arise that impact individuals if life events may be predicted based on the activity abstracted across a collective. From a business standpoint, AI-based recommendations or automation could lead to increased sales, however, the impact of this at an individual level can be morally or ethically grey. For workers, this could mean facing unemployment due to new AI systems being introduced without viable opportunities for upskilling [77]; for individuals, this could arise from reducing the human experience to purchase history and manipulating customer data to influence spending and predict life experiences.

Bellman et al. [10] discuss a similar scenario, where a nursery introduced a fee to deter parents from picking up their children late; the nursery saw that this actually exacerbated the problem and had an adverse effect, since the 'lateness' behaviour then had a monetary value attached to it that overruled the moral value. By quantifying human morals and values as numerical or financial values, this can result in reinforcement of the behaviour that the measure intended to reduce. While this example does not involve the use of AI, it highlights the importance of preserving the human experience and the values we hold when designing AI systems, coined as *socially-sensitive* system design [10]. For AI to be fully integrated into our society sustainably, designers should consider how the AI system will operate in and affect the complex interactions of both humans and other machines in accordance with the values we hold as a society. This line of thinking relates to Feldman's [31] idea of 'comparing the incomparable' when designing utility functions of AI systems, where systems must consider trade-offs for how they may achieve their goals; in line with the arguments of Bellman et al. [10], we argue that the consideration of trade-offs to achieve goals should be extended to consider also the situatedness of the system, and how achieving these goals impact the direct and indirect sociotechnical environment. If AI systems do not consider their situatedness in terms of their sociotechnical environment, then they will not reasonably be a viable option for continued use due to the unintended consequences that may arise due to their effects on society.

The word 'society' in this chapter, as described at the beginning of Section 2.1, is used here in the context of AI to encompass all individuals across the world. It is important to acknowledge that while our entire society is influenced by the use of AI systems due to the interconnected nature of our sociotechnical environment, the implications this has on our human experience can differ greatly between individuals. There exists a 'digital divide' both internationally and within nations, affected by access to digital infrastructure, individual and national average income, knowledge, access to education, and even language [44]. These attributes, especially access to high-speed internet, all affect the personal benefit that individuals can glean from using AI – or prevent them from being able to access it directly at all. Free-to-use AI tools, such as ChatGPT, are helping to remove barriers related to digital infrastructure and income; individuals now only need an internet-enabled device and internet connection to use AI tools, rather than purchasing additional devices such as those for Alexa or Google assistants [44]. Regardless of whether an individual uses or has access to AI systems explicitly, our society as a whole remains affected by the use of these AI systems that stem from the way we are connected to and experience our sociotechnical environment [44, 50, 85]. Sustainable AI moving forward will not only be about sustainable operation for the benefit of the environment, but also to ensure that wealth, employment opportunities and human experience are preserved in terms of its sustainable societal use and situatedness. This also means that new employment opportunities will be created where AI replaces the need for humans to carry out a task, human morals and values are conserved, and the interconnected nature of our modern world is respected and considered, to enable everyone to benefit from AI both implicitly and explicitly. This is especially important considering that modern AI systems affect individuals and communities even without direct interaction. Socially, this could be through the use of companies or institutions in ways that affect or influence our behaviour as individuals. Moreover, the passive effect of using AI systems globally is that the environment we inhabit is being affected to an increasing magnitude. Since AI has the ability to affect our human experience and the ways we interact with our sociotechnical environment, AI has the potential to have a positive impact on the achievement of UN Sustainable Development Goals (SDGs) such as SDG 8 (Decent Work and Economic Growth), 9 (Industry, Innovation and Infrastructure), and 11 (Sustainable Cities and Communities) [81]. AI has the potential to be a significant enabler in supporting the sustainable development of our society – *if* designed with caution and consideration for its wider impact on, and the complex interconnectivity of, the sociotechnical environment in which it is situated and the humans it encompasses.

3 Unintended Social, Ethical and Environmental Impacts of AI

In 1927, Charles A. Beard adapted pre-existing social theories to the rapid development of technology [8], noting that *“Technology marches in seven-league boots from one ruthless, revolutionary conquest to another, tearing down old factories and industries, flinging up new processes with terrifying rapidity.”*. Beard’s work largely began the notion of *technological determinism*, the idea that technology has increasingly important effects on everyday life. This, along with the growing prevalence of sociotechnical systems, was discussed in part within *Risk Society* [9], where Beck argues that technological developments occur faster than society can react in terms of control and risk, thus leading to unintended consequences. In this section, we discuss these unintentional consequences with a focus on sustainable Artificial Intelligence on the perspective of social, ethical, and environmental impact.

3.1 Social and Ethical Impacts on Society and Individuals

Artificially intelligent systems, due to their transition from the laboratory to the real world for impactful change, now exhibit effects on society and the individual. Given the use of these systems in daily life, we must now consider the impacts that such systems have on society as a whole (social issues) and society's idea of what is right and wrong (ethical issues).

From mid- to late- 2010s, members of the general public were unknowingly embroiled in a data-driven controversy that would later become known as the Cambridge Analytica scandal [11]. Following personality tests taken by an estimated 50 million people via the Facebook social media platform (and, in addition, their extended network of connections), data scientists discovered that politically useful intelligence could be derived. This was due to high-dimensional data and their underlying relationships, with five thousand data points collected that describe an individual later extended to several million by external data brokers [38]. The company was shown to be active in 68 countries [29], and through the use of highly targeted ads, the scandal was linked to electoral influence in the United Kingdom, the United States, Australia, India, Kenya, Malta, and Mexico [67]. These data-driven approaches where machine learning techniques can uncover underlying information exemplify how AI techniques can be used to influence society as a whole, through psychological manipulation towards an aim, such as voting behaviour. Societal implications of algorithmic automation are growing, and Lambrecht and Tucker [49] describe how Amazon's automated recruitment tool preferred men over women and Apple's credit scoring system offered men higher credit limits than women. The gap between men and women involved in the STEM field is a persistent issue, and researchers argue that structural barriers in high school education, psychological factors, influence, expectation, and perceptions all play a role [42]. Referring back to an experiment performed by Lambrecht and Tucker [49], a Facebook advertising campaign on STEM careers was targeted at men and women aged 18 to 65 years of age in 191 countries with no demographic goals. They found that, although no specific demographics were targeted, 20% more men than women were displayed the ad on their newsfeed, and women aged 25-34 were 40% less likely to see the ad than their male counterparts. For those who did not see the ad, engagement with it was therefore impossible. It is clear that AI has a significant impact on individuals in our society, however, for AI to be used sustainably to support our society, issues regarding gender inequality and bias need to be carefully considered and robust measures taken to promote fairness and equality, especially in terms of UN SDG 5 (Gender Equality).

Generative AI is a current paradigm shift in the field, significantly impacting several academic fields through the method of generating synthetic media [43], including, but not limited to, text, images, videos, and audio. Since the dawn of humanity, our species has exhibited creativity in the form of art long before the formation of civilisation [83]. As noted by Belfiore and Bennett, the societal impacts of creativity include transformation, positive individual impact, and the argument of *l'art pour l'art* (art for art's sake) in that, for something to be truly creative, it is beyond all social values and utilitarian function [69], and that these ideas have been commonplace since the days of Plato and Aristotle. Although technology has had a considerable impact on society's idea of *creativity*, these concepts have remained for more than 2000 years. According to Epstein et al. [30], the rise and developments within the field of Generative AI are likely to transform the nature of creative work and related employment. However, they note that generative AI is less likely to have a negative impact on these non-routine and high-skill tasks. This is supported by Polanyi's paradox "*We can know more than we can tell*" [5], which describes cognitive phenomena that we humans intuitively understand but cannot adequately explain, such as the creative process that leads to tangible outputs. For example, in the music industry, automation techniques have led to more musicians, not fewer [46], where AI can be used to augment an individual's abilities rather than replace their expertise [44, 79]. Celise, Claudio, and Chow [16] argue in favour of a relational-materialistic definition of creativity in the context of these systems, arguing against both universalism (a singular definition of the concept of creativity) and reductionism (the idea that only humans can be *truly* creative). This labour-focused approach moves away from the idea that AI technology is augmenting creativity but examines its impact on practices and social structures.

The transition from universalism and reductionism towards relational-materialistic points of view epitomise the paradigm shift that the concept of Generative AI has had on modern society, leading us to question the very definition of *creativity*. Given the rapid developments of Generative AI that lead to media that cannot be adequately recognised by humans in visual Turing tests [22, 60], this leads to ethical issues regarding our understanding of authenticity and truth itself. For example, a study from Silva et al. [73] showed that a sample of 50 participants could only detect AI-generated artwork 58% of the time. Ethical issues in synthetic media revolve around the ontological, on the nature of being, and epistemological, on theory and scope. Ontologically, given that humans have an increasingly waning ability to recognise what is and is not real, there is therefore a growing epistemological reality that there are concerns surrounding the reliability of human knowledge. In 2021, an AI-generated image was entered in the Colorado State Fair's annual art competition, winning an award in the category of emerging digital artists and causing a backlash from the community [68]. Here, the judges in the art competition provide an example of ethical dilemmas both ontologically and epistemologically. The judges were faced with the issue of the authenticity of the artworks, effectively tasked with defining what art was, while Polanyi's paradox suggests that the cognitive phenomena that lead to art are unexplainable. Epistemologically, the incident raises questions about the reliability of the knowledge and judgement of the panel.

In terms of concerns regarding reliability of academic knowledge, a scientific article originally published in 2023 was later retracted due to ChatGPT's *"Regenerate* *response"* string being erroneously included in the published manuscript [21]. Later, an investigation found that the use of Generative AI violated the ethical policies of the journal. A similar incident occurred in 2024, with a published manuscript beginning with the phrase *"Certainly, here is a possible introduction for your topic:"* [87]. The phrase was attributed to the use of a Large Language Model, and the article was subsequently retracted by the journal due to both the use of Generative AI and the plagiarism of text and image data. Beyond these clear examples due to unintended inclusion of text from the Generative AI interface, recent studies in Natural Language Processing by Liang et al. [52] argue the revelation of distinctive increases in the use of Generative AI in scientific publishing, estimating a distribution of 17.5% of sentences in Computer Science papers being either generated or modified by Large Language Models. Referring back to the aforementioned theories, this leads to a significant ethical issue in the reliability of our knowledge, suggesting that considerable portions of recent scientific publications contributing to human knowledge were not written by a human being. According to the UK Government's *Ethics, Transparency and Accountability Framework for Automated Decision-Making* [78], "organisations and individuals should be held accountable to ensure the proper functioning of artificial intelligence". The unintended consequences of unethical behaviour can be seen here, as it is researchers who produce retracted articles that suffer reputational damage and a 10% decrease in citation rate on average [6], not the algorithm that was used for this unethical activity. Large Language Models specifically tailored for academic writing that are marketed for *enhancing*, rather than *replacing* academic writing, such as AcademicGPT, jenni.ai and Paperpal are increasingly more accepted and used by academics; after all, if our academic peers use these tools and are consequently perceived to be more effective and successful by doing so (e.g. they publish more academic papers and/or faster, obtain more grants), then should we ourselves also be using these tools if we are not to be left behind? The line between what is ethical and what is not becomes more blurred.

Biases often arise from unexplained AI model predictions, due to, for example, their *black box* nature [2]. That is, a complex system where the translation from input (data) to output (prediction or decision) is not easily understood, and thus may lead to unintentional misrepresentation of said data to achieve a given goal. Take, for example, a binary image classification problem with a model that scores highly in metrics to discern between the two. In this example, the difference in preprocessing techniques between the two classes of images could have been different, e.g. if a centre cropping technique was used to resize class A to preferred dimensions, while a stretch technique was used to resize images belonging to class B, then it is not beyond doubt that such an image classifier is learning to simply detect an image resizing technique and thus score well with respect to the chosen metrics. A more practical example can be seen in observations made by Murphy [59], where a reinforcement learning algorithm tasked with playing the classical Tetris game reacts to a potential game-over state by pausing the game, effectively preventing the state transition indefinitely. A solution to these potential issues is Explainable AI (xAI), in which model predictions are reverse-engineered by various methods to provide an explanation of why a prediction was made given the data provided and the inner workings of the algorithm. Evidence of where explainable AI is useful is most easily understood in visual problems, such as image recognition, where techniques such as Grad-CAM [70] are used to overlay a heatmap on the input data to provide context of why an image processing-based prediction was made. Examples of this are provided by Fuhrman et al. [34], where visual overlays are presented when detecting COVID-19 from x-ray scans, ultimately leading a human expert's attention to areas of the image.

In visual problems, techniques for explainability are often themselves, also visual in nature. Interpretability in this case arises from our inherent ability to process visual information. However, explainability in the case of non-visual data, such as when processing numerical or categorical attributes, often comes in the form of inspecting model parameters globally or tracing back a model's single prediction given input values [57]. In this case, the predictions are explainable but not necessarily interpretable and may require verification by a human domain expert. As AI research makes the transition from the laboratory to the real world, effective explainable and interpretable techniques are required for validity.

The integration of AI into our society poses ethical and social questions about how these systems can be used sustainably for the good of society. Care in the design process that acknowledges the power that AI systems hold over behaviour and experiences in both individuals and society as a whole is necessary moving forward to ensure that the effects of AI are for the benefit of society and the individuals within it, rather than unintentionally causing harm or influencing behaviour that leads to a shift in societal changes or perspectives.

3.2 Environmental Impacts of AI

AI is becoming more ubiquitous, and as such, consideration is needed for how these systems not only impact society, but the environment as well. AI has the potential to be of great benefit to our society for providing more efficient ways to manage our environmental resources, optimise and automate processes to reduce energy and resource consumption, and find innovative solutions to environmental issues such as climate change. However, to reap this benefit, AI systems must be constructed from materials sourced from the environment by members of our society, powered by energy, cooled with water, and occupy physical space in the environment in order to operate. For sustainable operation of AI systems, the design and operation of these systems should consider the wider embodied and operational costs of their use, in order to offset any negative environmental consequences raised in comparison to the generated environmental benefits.

3.2.1 AI for the Environment

As our society becomes more reliant on technology, the environmental impacts of producing and using these technologies grows in tandem. AI has enabled the development of new technologies to manage and support our environment and natural resources. For example, in the field of smart agriculture, the introduction of AI into agricultural practices has led to improvements of both crop and soil quality, while reducing agricultural waste and water usage [55]; this translates to improvements for the agricultural environment and availability of food, whilst working towards tackling water, air and ground pollution. A further benefit noted by Mana et al. [55] is the optimisation of irrigation in smart agriculture, thus reducing water wastage. Additional effects of AI on water resources are emerging from further studies, such as managing and modelling water quality in reservoirs [17], or utilising ChatGPT for water management and optimising usage [28]. AI has also been harnessed for the monitoring of air pollution to improve environmental and human health [3], thus creating a way to automate the process and reduce the costs associated with it. In addition to the management of air pollution, water, and agriculture, AI has been used to optimise energy resources, including renewable energy such as solar and wind [1]. The role of AI in developing sustainable practices for managing marine ecosystems has also been explored, with researchers noting the interplay between the preservation of marine environments and climate change [36]. AI is already being harnessed for environmental benefits, managing our usage of natural resources and reducing environmental harm caused by human activities; Vinuesa et al. [81] note that using AI in this way could enable the achievement of 25 of the 27 targets in UN SDGs 13 (Climate Action), 14 (Life Below Water), and 15 (Life on Land). Consequently, there is a responsibility on the researchers and developers of these AI systems to ensure that environmental and societal benefits are balanced throughout all stages of the AI systems' operation, in order to achieve true sustainability in AI.

3.2.2 Detriments of AI in the Environment

While usage of AI may be well-intentioned as in the above applications, there are often hidden or unintended consequences of using such systems that transpire; these consequences can stem from a lack of education in how to properly use the system in a sustainable manner, lack of domain understanding coupled with eagerness to accept the AI-generated solution, lack of transparency in how the system was built or powered, or even flaws in the design or data used to create the AI system. There is also a trade-off to consider, since AI systems designed to tackle issues of climate change, air pollution, clean water or energy amongst others can unintentionally contribute to those problems if sustainability is not considered during their development [12, 20, 44, 85]. Vinuesa et al. [81] note the merits of AI when addressing UN SDGs, but also highlight that eight of the 25 targets in UN SDGs 13 (Climate Action), 14 (Life Below Water), and 15 (Life on Land) could be inhibited by the use or application of AI, either through the resources required to operate these systems, or through misuse or exploitation of natural resources through more accurate information being gained. For example, Wu et al. [85] note that both the *operational* and *embodied* carbon footprints are of great importance when considering whether AI practices or systems are sustainable. This means that consideration must not simply be given to the running costs of AI systems (e.g. in terms of energy, water, and greenhouse gas emissions), but those associated with activities such as training or developing the AI systems, and producing the materials – including energy – to maintain or produce such systems in their entirety [44]. Similarly, while AI systems have been utilised to support both the environment and society in a number of ways, there are questions raised over the operational and embodied costs associated with using them – e.g. do the benefits of using the system outweigh the costs of building, training, and operating it? For example, ChatGPT has been used in applications such as water management and usage [28], however, the operational and embodied water usage of GPT-3 consumed 700,000 litres of water during training, which could increase for the current GPT-4 model and subsequent, larger models in the future [50]. It is also noted by Li et al. [50] that both the energy production methods and server farms necessary to operate AI systems such as ChatGPT require further water consumption for cooling, in addition to other activities such as training, over the system's lifetime. This raises the counterintuitive question of whether using a powerful, and energy-/water-consuming system such as ChatGPT is truly beneficial and sustainable, when considering both its operational and embodied costs, and beneficial outcomes; this highlights the increasing importance of analysing the sustainability of AI systems using a holistic approach, that not only considers the effects of their outputs, but considers the costs associated with achieving those outputs in their entirety.

One way to mitigate these problems could be to source green energy, for example, that produced by wind turbines as a source of renewable power. Counterintuitively, the production of this energy has its own environmental and social implications, as the production of the raw materials needed to build wind turbines can have adverse effects both on the environment through the extraction of the raw materials required and the manufacturing process [58]. Although these activities can lead to the creation of jobs, there are also negative impacts on local communities located near industrial activity [58]. Therefore, a strong link between the negative environmental and social impacts is present, and mitigating both would in turn benefit the other. In line with the arguments posed by Wu et al. [85] and Li et al. [50], designers of AI systems moving forward must consider both the operational and embodied costs associated with each aspect of AI system usage, from building infrastructure, to training, powering, cooling, and daily use. In addition to reducing these operational and embodied costs, for AI to be a sustainable option for the systems of the future – either through sustainable practice, AI for sustainability, or both – designers should also consider the interplay between environment and society. End users also contribute to the perceived sustainability of these systems when they engage with them; transparency is therefore important to enable users to make their own, educated decisions on whether to, and how to engage with these systems.

4 Designing the Sustainable Systems of the Future

The narrative in this chapter thus far has highlighted the prevalence of AI systems in our modern, sociotechnical environment, and considered their effects and influence on both society and the environment, as well as the ethical issues that arise through their usage and operation. We have argued that, due to the increasing integration of AI systems into our society and the unintended interactions they facilitate, sustainability in the context of AI now extends beyond thinking how AI systems operate sustainably for the benefit of the environment; it now also includes sustainable societal use, situatedness, and acceptance to be viable moving forward. van Wynsberghe terms this distinction as 'AI for sustainability' and 'sustainability of AI' [86]. We emphasise the need for designing and employing AI systems to benefit their own sustainability and the sustainability of our sociotechnical environment. The importance of considering the unintended consequences of interactions of AI systems in our complex sociotechnical landscape has been highlighted in Section 2, and the societal and environmental impacts of AI further in Section 3. Here, we posit that these things are a predecessor to designing a system that is capable of operating sustainably for the benefit of society and the environment, and that these benefits may only be realised through the public engaging with and using these systems after a sense of trust and acceptance is established through governance, accountability, transparency, and ethics.

4.1 Governance and Accountability of Sociotechnical Systems in the UK and EU

As discussed in Section 3, Beck's *Risk Society* [9] argued that the development of technology is faster paced than society's ability to react in terms of control and risk. The increased complexity of statistical inference leading to machine learning and artificial intelligence is not a new concept; however, at the time of writing, legal frameworks to govern AI technologies are only just beginning to adapt [39]. Huang et al. [39] note that globally coordinated adaptive, agile regulations must be introduced that are responsive to the advancement of AI technologies. These regulations should be lax enough not to stifle technological innovation, yet prescriptive enough to avoid loopholes that could be exploited to the detriment of society and the environment.

In the UK, the *Ethics, Transparency and Accountability Framework for Automated Decision-Making* [78] is guidance towards overcoming distinct distrust in advanced technology, which includes artificial intelligence. The framework includes seven key points. First, prototyping and testing, where expert human oversight ensures rigorous and controlled testing to avoid unintentional harm. Thus, this stage is focused on avoiding the potential for unintended social, ethical, and environmental impacts, as we have discussed previously. Point 2 is centred on bias specifically, on the ethical issues of prejudice and bias. The UK government suggests judgement, engagement, and testing along with diverse stakeholders to avoid unjust and unintentional impact.

The Government also recommends that experts from other disciplines be involved throughout the lifecycle of the project. Following protocols for avoiding unintended impact, point 3 introduces the assumption that automated decision-making systems rely on the responsibility of a *senior owner*. That is, organisations and individuals are to be held accountable for the proper functioning of such systems. Significant decisions must be signed off and in accordance with the Ministerial Code. Point 4 notes that algorithms or systems must comply with data protection laws so that the senior owner ensures the safe handling of data, as well as the protection of individual interests. Point 5 of the framework surrounds clear communication in that all those who may be affected by automated decision making should be able to understand the process with explanations written in layman's terms. Similarly to points 2 and 5, point 6 notes that automated decision-making systems should comply with relevant laws. In addition to the equality act and data protection, other necessary legislation may be applicable, and thus the algorithm or system may require legal sign-off by a government legal advisor. Finally, point 7 notes that unintentional consequences may manifest over time, i.e. in the case of an algorithm that is continually trained with new data, and therefore formal review points should be implemented in order to ensure oversight and mitigation.

The European Union, on the other hand, introduces comprehensive legislation through the EU AI Act [53]. Unlike the UK's *innovation first* approach, the EU approach consists of stringent risk assessments, mandatory documentation, and specific compliance with regulations. EU risk scores are a range of unacceptable, high, limited, and minimal risk. Examples of unacceptable risk include social scoring and real-time biometric identification, and high-risk systems are those that require rigorous testing and documentation such as artificial intelligence systems within critical infrastructure and autonomous vehicles, as well as those that are involved in education or employment. Limited-risk systems are those that include minimal transparency requirements, such as recommendation systems. Generally, systems are considered to be of minimal risk when no new obligations are introduced, such as spam filters that rely instead on voluntary codes of conduct. Systems that are considered at an unacceptable risk level will be banned six months after the act is introduced. The legislation places emphasis on transparency, where risk mitigation and the protection of fundamental rights are achieved through the disclosure of key information from developers of AI systems in the 12 months following the introduction of the act.

In an analysis of the aforementioned UK framework and EU act, we note three major differences: regulation, implementation, and precedence. In terms of regulation and implementation, the regulatory philosophy of the UK approach is pro-innovation in that principles of safety and transparency interpreted by existing regulation are favoured over the EU's alternative approach that involves specific legal obligations and bans. The EU act sets a precedent with the aim of influencing a global standard, whereas the UK approach instead adopts international norms. The EU act aims to ensure that artificially intelligent systems are sustainable by preventing compromising societal and environmental well-being, and the UK's emphasis on innovation aims to facilitate more rapid technological advancement with ethical oversight for sustainable growth. As new legislation and regulations are introduced to govern the adoption and development of AI systems around the world, it is increasingly important to balance attention to both protecting and enhancing our sociotechnical environment. For example, the access and operation of some Meta and Apple AI technologies will differ for EU users compared to users from other regions around the world to comply with EU regulations [56]; this situation can be perceived in two ways, either that EU regulations are too restrictive and may stifle technological progress, or other regulatory bodies allow too much freedom. This also raises the question of whether prescriptive legislation may in fact introduce further technological and digital divides and propagate technological inequalities, given that access to these state-of-the-art systems depends on the individual's geographic location. Thus, an unintended consequence here may be a contribution to societal harm while intending to reduce it. This discussion highlights the delicate nature of introducing AI regulations to govern the development and usage of AI systems, and for these to be introduced for the sustainable adoption and acceptance of AI, coordination and consistency will be key.

4.2 Acceptance, Trust and Integration into Society

Overlapping with the discussion on governance and accountability of AI in Section 4.1, Shavit et al. [71] highlight the importance of safety and accountability in agentic AI systems, i.e. those that have little or no human intervention in the decision-making process. The effects of integrating AI systems such as this into our society can be beneficial, for example, through offloading decision-making processes or completing tasks on behalf of human users, but also pose risks if these systems fail, are used with ill intention, or are not designed to prevent harm [71]. This is described as an AI system being 'robust' by Díaz-Rodríguez et al. [25], and is defined as one of three pillars that constitute trustworthy AI, alongside lawful and ethical operation. From a sustainability perspective, we argue that designers should reasonably attempt to consider the embodied potential harm of their systems (in the same sense as 'embodied cost' as discussed in Section 3.2), e.g. from design, to physical construction, to training, and to daily use. The public should therefore reasonably expect to be able to interact with AI systems, either directly or indirectly, without fear of experiencing individual or societal harm; thus, this would affect the formation of trust in both the system and its actions, where trust can lead to sustainable acceptance of the system by individuals in our society [14]. Reliability and safety are also cited as contributing factors for trustworthiness in AI systems by Shneiderman [72], recognising that humans are a core part of the sociotechnical landscape and both effect and are affected by AI systems, thus enforcing the need for safe and reliable operation of AI systems. These concepts are also explored in research focussed on AI for Social Good (AI4SG), which explores how AI can be used for the benefit of humans and environment. Floridi et al. [32] propose seven factors of significance that influence AI4SG, including transparency, privacy, and human-friendly semantisation; the authors argue that systems, both AI and non-AI, should respect the autonomy of users, and systems that restrict human autonomy could indeed lead to a lack of trust and dismissal of the technology altogether. Floridi et al. [32] also highlight the importance of determining which tasks are suitable for delegating to AI systems to facilitate human-friendly semantisation, and which tasks should be actioned by humans as to not have adverse effects on the human experience. Again, if this distinction is not made with clear consideration, and AI systems are used for tasks that reduce the human experience in a way that loses meaning, this can result in trust being broken and disengagement with the system. For AI to remain a sustainable option to benefit society, it should be designed in a way that encourages trust and autonomy. Exploring the concept of trust further, Powers et al. [64] emphasise that a foundation for building trust in AI systems is the education of its users, since AI systems influence society and the individuals within it both directly and indirectly. Education here is defined more broadly than just educating individuals about the low-level technical workings and detail of the system, but primarily encompasses providing individuals with an awareness and literacy of how these AI systems affect themselves, their own, personal environment, and further into how AI affects their social and environmental context. Powers et al. [64] also note the risk of providing such education and awareness, in that this could potentially cause societal divides if not accessible by all. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design* [75] reiterate the need for educating the public to promote awareness of the societal and ethical impacts of the use of AI, and for ensuring that AI systems promote individual wellbeing and fairness, for example, by not excluding individuals in the community from accessing this information. Part of this education and awareness could stem from transparency and explainability in the decisions that AI systems make, providing that AI systems are governed in such a way to benefit the current and future society [4]. Brauner et al. [14] suggest that accessible education about the benefits, limitations, and usage of AI is important to facilitate productive discussion around its adoption; AI literacy is also important in this sense for producing regulations that both protect and enhance our sociotechnical environment. This also raises the issue of ensuring that everyone, regardless of socioeconomic status, benefits from the use of AI [77]. Acceptance of, and trust in, AI systems is predicated on individuals understanding how these systems affect them, as well as their place in our sociotechnical environment. To promote trust in AI systems and their use, AI systems should not encroach on human values, diminish the human experience, or ostracise individuals, but instead provide individuals with the tools required to establish trust, such as education, explainability, and safety in operation. Sustainable societal use and acceptance of AI systems will therefore rely on communication of its abilities, decision-making processes, and effects on individuals, our society, and the environment, to empower individuals to make informed decisions about how and why they use AI, and how AI affects them in the context of their own situatedness in the sociotechnical environment.

5 Conclusion

In recent years, AI systems have become ubiquitous, with their existence and use affecting every facet of our societal activities and the environment in which we co-exist. In this chapter, we argue that for AI to be sustainable in the broadest sense of the word, a distinction needs to be made so that AI systems are not only designed to be of benefit to the environment in the general and traditional sense of the word 'sustainability', but also that AI needs to have continued and defined societal benefits that consider how it is used and affects our increasingly interconnected and complex sociotechnical environment as a whole.

The discussion throughout this chapter amounts to the trade-offs that are associated with introducing AI systems into our society, in terms of acceptance, trust, success (in terms of achieving goals), and social, ethical, environmental, and financial costs. We acknowledge that the very nature of a trade-off means that compromises must be made; however, to achieve true sustainability in AI, we argue that it is the *understanding* of the context in which the AI systems will operate that will lead to sustainability. While the impact of, and interactions between, AI systems and the sociotechnical environment can be difficult to anticipate – hence the phrase "unintended consequences or interactions" – it is our responsibility as AI practitioners to fully consider these impacts, and mitigate negative consequences to the best of our ability. Recognising that AI systems do not act in isolation and are in fact intricately woven into the fabric of our sociotechnical environment should, as a result, translate to AI systems being intentionally designed to sustainably and responsibly operate – not just for the good of society, but for our environment as well.

For sustainable integration of AI into our society, we discuss the need to educate the public about how AI systems impact us both individually and as a wider society as a whole [64], as a step toward acceptance and trust in these systems moving forward. Empowering individuals to make informed decisions about when or how they use AI, as well as providing information and education for how individuals are impacted by AI as part of society as a whole, could enable trust to be established for sustainable societal use of these systems. Part of sustainable AI in relation to the UN SDGs would be to ensure that this education does not exclude individuals [64, 75], and that its use and operation promotes employment and learning opportunities without harming individual well-being, livelihoods and wealth [54, 77]. Furthermore, while AI is widely used in applications to support sustainability in our environment [1, 3, 36, 55], consideration must be made regarding the embodied and operational costs of these systems, and whether the process of designing, building, and operating these systems on a daily basis is indeed sustainable when weighing the drain on environmental resources that these systems may have when integrated into society.

The rise of AI systems in our society is inevitable, and the consequent effects and interactions with our sociotechnical environment are progressing rapidly. AI has vast potential to empower individuals and enable us to operate as a society more effectively in terms of how we manage and consume natural resources for the preservation of our environment; acknowledgement and understanding of the complexities of our increasingly AI-enabled sociotechnical environment, and mitigation of unintended consequences stemming from operation, will enable the AI systems of the future to be considered sustainable at a fundamental level for the benefit of both the society of today and the future by preserving our social, ethical and environmental landscape for generations to come.

References

[1] Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., Chen, H.: Artificial Intelligence in Sustainable Energy Industry: Status Quo, Challenges and Opportunities. Journal of Cleaner Production **289**, 125834 (2021). DOI https://doi.org/10.1016/j.jclepro.2021.125834

[2] Albahri, A.S., Duhaim, A.M., Fadhel, M.A., Alnoor, A., Baqer, N.S., Alzubaidi, L., Albahri, O.S., Alamoodi, A.H., Bai, J., Salhi, A., et al.: A Systematic Review of Trustworthy and Explainable Artificial Intelligence in Healthcare: Assessment of Quality, Bias Risk, and Data Fusion. Information Fusion (2023)

[3] Asha, P., Natrayan, L., Geetha, B., Rene Beulah, J., Sumathy, R., Varalakshmi, G., Neelakandan, S.: IoT Enabled Environmental Toxicology for Air Pollution Monitoring using AI Techniques. Environmental research **205**, 112574 (2022)

[4] Astobiza, A.M., Toboso, M., Aparicio, M., López, D.: AI Ethics for Sustainable Development Goals. IEEE Technology and Society Magazine **40**(2), 66–71 (2021)

[5] Autor, D.: Polanyi's Paradox and the Shape of Employment Growth. Tech. rep., National Bureau of Economic Research (2014)

[6] Azoulay, P., Bonatti, A., Krieger, J.L.: The Career Effects of Scandal: Evidence from Scientific Retractions. Research Policy **46**(9), 1552–1569 (2017)

[7] Barnes, C.M.: Interference and Volatility in Evolutionary Agent-Based Systems. Ph.D. thesis, Aston University (2021)

[8] Beard, C.A.: Time, Technology, and the Creative Spirit in Political Science. American Political Science Review **21**(1), 1–11 (1927)

[9] Beck, U.: Risk Society: Towards a New Modernity, vol. 17. sage (1992)

[10] Bellman, K., Botev, J., Hildmann, H., Lewis, P.R., Marsh, S., Pitt, J., Scholtes, I., Tomforde, S.: Socially-Sensitive Systems Design: Exploring Social Potential. IEEE Technology and Society Magazine **36**(3), 72–80 (2017)

[11] Berghel, H.: Malice Domestic: The Cambridge Analytica Dystopia. Computer **51**(05), 84–89 (2018)

[12] Berthelot, A., Caron, E., Jay, M., Lefèvre, L.: Estimating the Environmental Impact of Generative-AI Services using an LCA-based Methodology. Procedia CIRP **122**, 707–712 (2024). DOI https://doi.org/10.1016/j.procir.2024.01.098. 31st CIRP Conference on Life Cycle Engineering

[13] Bingley, W.J., Curtis, C., Lockey, S., Bialkowski, A., Gillespie, N., Haslam, S.A., Ko, R.K., Steffens, N., Wiles, J., Worthy, P.: Where is the Human in Human-Centered AI? Insights from Developer Priorities and User Experiences. Computers in Human Behavior **141**, 107617 (2023)

[14] Brauner, P., Hick, A., Philipsen, R., Ziefle, M.: What oes the Public Think about Artificial Intelligence? — A Criticality Map to Understand Bias in the Public Perception of AI. Frontiers in Computer Science **5**, 1113903 (2023)
[15] Bruzzone, A., Sinelshikov, K., Massei, M., Giovannetti, A., Tarone, F., Longo, F., Fancello, G., Giliberti, C., Mariconte, R., Diano, M., et al.: Reducing Dangers within Industrial Plants by Extended Reality. In: Proceedings of the 33rd European Modeling & Simulation Symposium (EMSS 2021), pp. 413–417 (2021)
[16] Celis Bueno, C., Chow, P.S., Popowicz, A.: Not "What", but "Where is Creativity?": Towards a Relational-materialist Approach to Generative AI. AI & SOCIETY pp. 1–13 (2024)
[17] Chou, J.S., Ho, C.C., Hoang, H.S.: Determining Quality of Water in Reservoir using Machine Learning. Ecological Informatics **44**, 57–75 (2018). DOI https://doi.org/10.1016/j.ecoinf.2018.01.005
[18] Choudrie, J., Manandhar, N., Castro, C., Obuekwe, C.: Hey Siri, Google! Can you help me? A Qualitative Case Study of Smartphones AI Functions in SMEs. Technological Forecasting and Social Change **189**, 122375 (2023)
[19] Chryssolouris, G., Alexopoulos, K., Arkouli, Z.: Artificial Intelligence in Manufacturing Equipment, Automation, and Robots. In: A Perspective on Artificial Intelligence in Manufacturing, *Studies in Systems, Decision and Control*, vol. 436, pp. 41–78. Springer (2023)
[20] Coeckelbergh, M.: AI for Climate: Freedom, Justice, and Other Ethical and Political Challenges. AI and Ethics **1**(1), 67–72 (2021)
[21] Conroy, G.: Scientific Sleuths Spot Dishonest ChatGPT Use in Papers. Nature (2023)
[22] Cooke, D., Edwards, A., Barkoff, S., Kelly, K.: As Good as a Coin Toss Human Detection of AI-Generated Images, Videos, Audio, and Audiovisual Stimuli. arXiv preprint arXiv:2403.16760 (2024)
[23] Corrigan, H.B., Craciun, G., Powell, A.M.: How does Target Know So Much About its Customers? Utilizing Customer Analytics to Make Marketing Decisions. Marketing Education Review **24**(2), 159–166 (2014)
[24] Crutzen, P.J.: The "Anthropocene". In: Earth System Science in the Anthropocene, pp. 13–18. Springer (2006)
[25] Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M.L., Herrera-Viedma, E., Herrera, F.: Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation. Information Fusion **99**, 101896 (2023)
[26] Do, K., Pang, R.Y., Jiang, J., Reinecke, K.: "That's important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–16 (2023)
[27] Duhigg, C.: How Companies Learn your Secrets (2012). URL `www.nytimes.com/2012/02/19/magazine/shopping-habits.html`. [Accessed: 20 June 2024]

[28] Egbemhenghe, A.U., Ojeyemi, T., Iwuozor, K.O., Emenike, E.C., Ogunsanya, T.I., Anidiobi, S.U., Adeniyi, A.G.: Revolutionizing Water Treatment, Conservation, and Management: Harnessing the Power of AI-driven ChatGPT Solutions. Environmental Challenges **13**, 100782 (2023). DOI https://doi.org/10.1016/j.envc.2023.100782
[29] Ekdale, B., Tully, M.: How the Nigerian and Kenyan Media Handled Cambridge Analytica. The Conversation (2020)
[30] Epstein, Z., Hertzmann, A., of Human Creativity, I., Akten, M., Farid, H., Fjeld, J., Frank, M.R., Groh, M., Herman, L., Leach, N., et al.: Art and the Science of Generative AI. Science **380**(6650), 1110–1111 (2023)
[31] Feldman, J.A., Sproull, R.F.: Decision Theory and Artificial Intelligence II: The Hungry Monkey. Cognitive Science **1**(2), 158–192 (1977)
[32] Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to Design AI for Social Good: Seven Essential Factors, pp. 125–151. Springer (2021). DOI 10.1007/978-3-030-81907-1_9
[33] Floridi, L., Sanders, J.W.: On the Morality of Artificial Agents. Minds and Machines **14**, 349–379 (2004)
[34] Fuhrman, J.D., Gorre, N., Hu, Q., Li, H., El Naqa, I., Giger, M.L.: A Review of Explainable and Interpretable AI with Applications in COVID-19 Imaging. Medical Physics **49**(1), 1–14 (2022)
[35] Gabriel, I.: Artificial Intelligence, Values, and Alignment. Minds and Machines **30**(3), 411–437 (2020)
[36] Gambín, A.F., Angelats, E., González, J.S., Miozzo, M., Dini, P.: Sustainable Marine Ecosystems: Deep Learning for Water Quality Assessment and Forecasting. IEEE Access **9**, 121344–121365 (2021). DOI 10.1109/ACCESS.2021.3109216
[37] Hamilton, J.R., Ekárt, A., Patelli, A.: Predicting Normal and Anomalous Urban Traffic with Vectorial Genetic Programming and Transfer Learning. In: J. Correia, S. Smith, R. Qaddoura (eds.) Applications of Evolutionary Computation, pp. 519–535. Springer (2023)
[38] Hern, A.: How to Check whether Facebook Shared your Data with Cambridge Analytica. The Guardian **10** (2018)
[39] Huang, K., Joshi, A., Dun, S., Hamilton, N.: AI Regulations, pp. 61–98. Springer Nature Switzerland, Cham (2024). DOI 10.1007/978-3-031-54252-7_3. URL `https://doi.org/10.1007/978-3-031-54252-7_3`
[40] Huberman, J.: Amazon Go, Surveillance Capitalism, and the Ideology of Convenience. Economic Anthropology **8**(2), 337–349 (2021)
[41] International Commission on Stratigraphy: Joint Statement by the IUGS and ICS on the Vote by the ICS Subcommission on Quaternary Stratigraphy (2024). URL `https://quaternary.stratigraphy.org/working-groups/anthropocene`
[42] Kanny, M.A., Sax, L.J., Riggers-Piehl, T.A.: Investigating Forty Years of STEM Research: How Explanations for the Gender Gap have Evolved over Time. Journal of Women and Minorities in Science and Engineering **20**(2) (2014)

[43] Kaswan, K.S., Dhatterwal, J.S., Malik, K., Baliyan, A.: Generative ai: A review on models and applications. In: 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), pp. 699–704. IEEE (2023)

[44] Khowaja, S.A., Khuwaja, P., Dev, K., Wang, W., Nkenyereye, L.: ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. Cognitive Computation pp. 1–23 (2024)

[45] Kim, J., Giroux, M., Lee, J.C.: When Do You Trust AI? The Effect of Number Presentation Detail on Consumer Trust and Acceptance of AI Recommendations. Psychology & Marketing **38**(7), 1140–1155 (2021). DOI https://doi.org/10.1002/mar.21498

[46] Knees, P., Ferraro, A., Hubler, M.: Bias and Feedback Loops in Music Recommendation: Studies on Record Label Impact. In: MORS@ RecSys (2022)

[47] Kreps, S., McCain, R.M., Brundage, M.: All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. Journal of Experimental Political Science **9**(1), 104–117 (2022)

[48] Krupitzer, C., Stein, A.: Unleashing the Potential of Digitalization in the Agri-Food Chain for Integrated Food Systems. Annual Review of Food Science and Technology **15** (2023)

[49] Lambrecht, A., Tucker, C.: Algorithm-Based Advertising: Unintended Effects and the Tricky Business of Mitigating Adverse Outcomes. NIM Marketing Intelligence Review **13**(1), 24–29 (2021)

[50] Li, P., Yang, J., Islam, M.A., Ren, S.: Making AI Less "Thirsty"': Uncovering and Addressing the Secret Water Footprint of AI Models (2023)

[51] Liang, F., Su, Z., Sheng, W.: Multimodal Monitoring of Activities of Daily Living for Elderly Care. IEEE Sensors Journal (2024)

[52] Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., et al.: Mapping the Increasing Use of LLMs in Scientific Papers. arXiv preprint arXiv:2404.01268 (2024)

[53] Madiega, T.: Artificial Intelligence Act. European Parliament: European Parliamentary Research Service (2021)

[54] Makridakis, S.: The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. Futures **90**, 46–60 (2017). DOI http://dx.doi.org/10.1016/j.futures.2017.03.006

[55] Mana, A.A., Allouhi, A., Hamrani, A., Rehman, S., el Jamaoui, I., Jayachandran, K.: Sustainable AI-based Production Agriculture: Exploring AI Applications and Implications in Agricultural Practices. Smart Agricultural Technology **7**, 100416 (2024). DOI https://doi.org/10.1016/j.atech.2024.100416

[56] Milmo, D.: Meta Pulls Plug on Release of Advanced AI Model in EU (2024). URL `https://www.theguardian.com/technology/article/2024/jul/18/meta-release-advanced-ai-multimodal-llama-model-eu-facebook-owner`. [Accessed: 20 December 2024]

[57] Mishra, P.: Explainable AI Recipes

[58] Morozovska, K., Bragone, F., Svensson, A.X., Shukla, D.A., Hellstenius, E.: Trade-offs of Wind Power Production: A Study on the Environmental Implications of Raw Materials Mining in the Life Cycle of Wind Turbines. Journal of Cleaner Production p. 142578 (2024)
[59] Murphy, T.: The First Level of Super Mario Bros. is Easy with Lexicographic Orderings and Time Travel. The Association for Computational Heresy (SIGBOVIK) **2013**, 112–133 (2013)
[60] Nightingale, S.J., Wade, K.A., Watson, D.G.: Can People Identify Original and Manipulated Photos of Real-World Scenes? Cognitive Research: Principles and Implications **2**, 1–21 (2017)
[61] Onesi-Ozigagun, O., Ololade, Y.J., Eyo-Udo, N.L., Ogundipe, D.O.: Revolutionizing Education through AI: A Comprehensive Review of Enhancing Learning Experiences. International Journal of Applied Research in Social Sciences **6**(4), 589–607 (2024)
[62] Opdahl, A.L., Tessem, B., Dang-Nguyen, D.T., Motta, E., Setty, V., Throndsen, E., Tverberg, A., Trattner, C.: Trustworthy Journalism through AI. Data & Knowledge Engineering **146**, 102182 (2023)
[63] van de Poel, I.: AI, Control and Unintended Consequences: The Need for Meta-Values. In: Rethinking Technology and Engineering: Dialogues Across Disciplines and Geographies, pp. 117–129. Springer (2023)
[64] Powers, S.T., Linnyk, O., Guckert, M., Hannig, J., Pitt, J., Urquhart, N., Ekárt, A., Gumpfer, N., Han, T.A., Lewis, P.R., Marsh, S., Weber, T.: The Stuff We Swim In: Regulation Alone Will Not Lead to Justifiable Trust in AI. IEEE Technology and Society Magazine **42**(4), 95–106 (2023). DOI 10.1109/MTS.2023.3341463
[65] Pringle, R., Michael, K., Michael, M.G.: Unintended Consequences of Living with AI: The Paradox of Technological Potential? Part II [Guest Editorial]. IEEE Technology and Society Magazine **35**(4), 17–21 (2016)
[66] Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Ahmad, J., Ali, M.E., Azam, S.: A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. IEEE Access (2024)
[67] ur Rehman, I.: Facebook-Cambridge Analytica Data Harvesting: What You Need to Know. Library Philosophy and Practice pp. 1–11 (2019)
[68] Roose, K.: An AI-Generated Picture Won an Art Prize. Artists Aren't Happy (2022)
[69] Schaffer, A.: Théophile Gautier and "L'Art Pour L'Art". The Sewanee Review **36**(4), 405–417 (1928)
[70] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
[71] Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., Hickey, A., et al.: Practices for Governing Agentic AI Systems. White Paper, OpenAI (2023)

[72] Shneiderman, B.: Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. ACM Transactions on Interactive Intelligent Systems (TiiS) **10**(4) (2020). DOI 10.1145/3419764

[73] Silva, R.S.R., Lotfi, A., Ihianle, I.K., Shahtahmassebi, G., Bird, J.J.: ArtBrain: An Explainable End-to-End Toolkit for Classification and Attribution of AI-Generated Art and Style. arXiv preprint arXiv:2412.01512 (2024)

[74] Sun, W., Bocchini, P., Davison, B.D.: Applications of Artificial Intelligence for Disaster Management. Natural Hazards **103**(3), 2631–2689 (2020)

[75] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (2019). URL `https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html`. [Accessed: 24 June 2024]

[76] Tomforde, S., Hähner, J., Seebach, H., Reif, W., Sick, B., Wacker, A., Scholtes, I.: Engineering and Mastering Interwoven Systems. In: ARCS 2014; 2014 Workshop Proceedings on Architecture of Computing Systems, pp. 1–8. VDE (2014)

[77] Truby, J.: Governing Artificial Intelligence to Benefit the UN Sustainable Development Goals. Sustainable Development **28**(4), 946–959 (2020)

[78] UK Government: Ethics, Transparency and Accountability Framework for Automated Decision-Making (2021)

[79] Usmani, U.A., Happonen, A., Watada, J.: Human-Centered Artificial Intelligence: Designing for User Empowerment and Ethical Considerations. In: 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1–7. IEEE (2023)

[80] Vaswani, A.: Attention is All You Need. Advances in Neural Information Processing Systems (2017)

[81] Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Fuso Nerini, F.: The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. Nature Communications **11**(1), 1–10 (2020)

[82] Whittaker, L., Kietzmann, T.C., Kietzmann, J., Dabirian, A.: "All Around Me Are Synthetic Faces": The Mad World of AI-Generated Media. IT Professional **22**(5), 90–99 (2020)

[83] Wilson, T.: Prehistoric Art; or the Origin of Art as Manifested in the Works of Prehistoric Man (1898)

[84] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al.: Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45 (2020)

[85] Wu, C.J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., et al.: Sustainable AI: Environmental

Implications, Challenges and Opportunities. Proceedings of Machine Learning and Systems 4 **4**, 795–813 (2022)

[86] van Wynsberghe, A.: Sustainable AI: AI for Sustainability and the Sustainability of AI. AI and Ethics **1**(3), 213–218 (2021)

[87] Zhang, M., Wu, L., Yang, T., Zhu, B., Liu, Y.: RETRACTED: The Three-Dimensional Porous Mesh Structure of Cu-based Metal-Organic-Framework-Aramid Cellulose Separator Enhances the Electrochemical Performance of Lithium Metal Anode Batteries. Surfaces and Interfaces p. 104081 (2024)