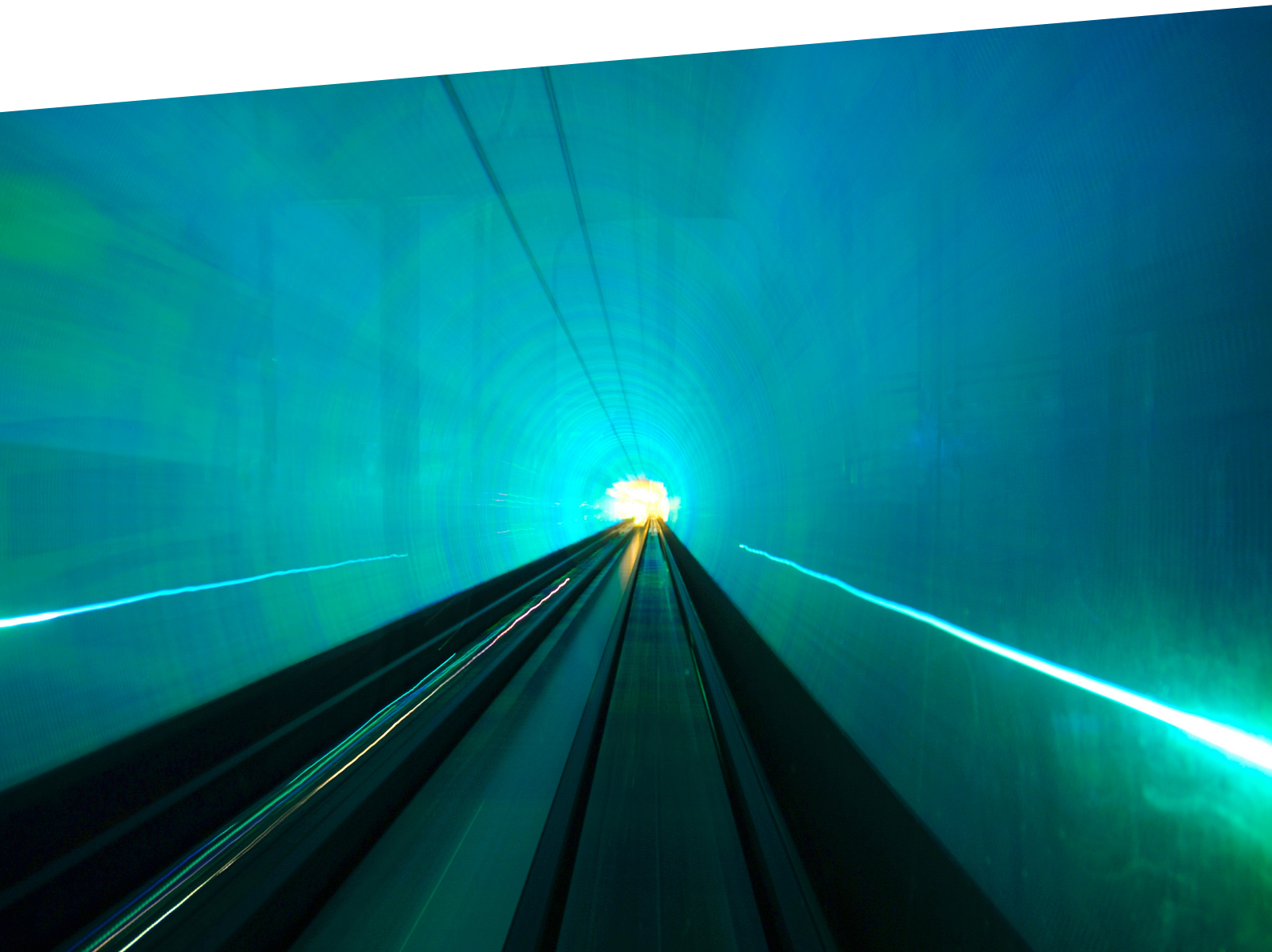
Ai
FOR
GOOD

Understanding Principles and Outcomes for Operationalizing Ethics in AI development



Authors:

Jesse Arlen Smith, Helena Ward,
and Olivia Gamblin

Acknowledgements

The authors are grateful to the following people for their input at workshops and valuable feedback on drafts of this paper: David Bergendahl, Dr. Mark Sprevak's - Dr. Bonime-Blanc, Dr. Lydia Kostopoulos Abhishek Gupta, Dr. Michael Klenk, Dylan Doyle-Burke, Dr. Olya Kudina, Dr Matthew Dennis , Steven Umbrello, Dr Willie Costello, Dr. Leila Taghizadeh, Dr. Semra Ascigil -E, and Dr. Danielle Anderson.

Aiforgood Asia STO

Aiforgood Asia is a financially and politically independent social enterprise

EQUALITY FROM TECHNOLOGY

www.aiforgood.asia

Aiforgood Asia

Our Mission

"To enable organizations and governments to implement AI projects that will improve society and fight inequality in Asia".

Aiforgood Asia is an international NGO that conducts research and implements projects that can help ensure AI is being used for the betterment of humanity. On the research side, we bring people, organizations, and governments together to conduct original research on the operationalization of ethics in AI development. On the project side, we implement practical community-based AI solutions that improve the health and welfare of the communities we live in and the planet we live on.

Together with our partners, we deploy these solutions to improve health & welfare, reduce inequality, fight climate change, and aid conservation efforts that protect endangered species and preserve our oceans and forests.

Our Approach

We are an NGO under the destination of Science and Technology Organization (STO) Business Number 0109802932 registered in Vietnam operating under the rights and obligations of a Social Enterprise in accordance with Vietnam Enterprise Law. We are the first international STO to bring people, organizations, and governments together to conduct original research and implement practical community-based AI solutions in Asia.

We believe everyone should have access to the power of AI technologies and that AI should be used for the betterment of humanity. Through a collaboration of industry, government, and technology partners, we promote the responsible development and implementation of AI technologies in a way that enables everyone to realize the benefits. Bringing together stakeholders from different sectors and regions, we explore the potential of AI to accelerate progress in fields such as healthcare, agriculture, and environmental protection, while at the same time addressing the ethical and societal concerns that come with these advancements.

Understanding the Ethical Outcomes for Operationalizing Principles in AI Development

The recent rise of Artificial Intelligence (AI) has sparked a proliferation of policies and practices intended to ensure the ethical development of these technologies. However, the variety of different ethical issues and potential outcomes of AI have led to an extensive yet somewhat ill-defined discourse surrounding the implementation of ethical guidelines. This discourse has left AI technologists with a lack of clarity over ethics and its relation to AI technologies.

This paper aims to bring clarity to the issues surrounding the development, implementation and use of AI, by categorizing these issues as outcomes and linking them to specific ethical principles. Each of the potential AI outcomes will be categorized into three parts: human outcomes, economic outcomes and environmental outcomes, and then linked to ethical principles. The researchers believe that having a transparent map of outcomes and principles will be useful for technologists in the operationalization of ethics in AI, as it will bring a deeper understanding of which principles need to be operationalized to mitigate certain negative outcomes. Since AI technologies are more problem specific rather than industry specific, the researchers have taken this more general application of ethics to be applied across industries, countries and businesses. As such the examples provided come from a variety of different applications and are intended to be illustrative rather than comprehensive. In this way the researchers hope that this paper can have utility for technologists across all industries grappling with how to effectively operationalize ethics in their AI development and deployment processes.

Principles and Outcomes

It will be important first to understand the relation between ethical principles and potential outcomes arising from the use or implementation of AI technologies. We can think of an ethical principle as something that upholds a particular moral value or judgement. For example, privacy is an ethical principle; it upholds the value of respecting a human's right to freedom from interference or intrusion. Principles can be utilized to bring about, or mitigate a particular outcome. So, we can think of an outcome as what happens when a principle is not upheld or neglected within a certain context.

The operationalization of ethical principles alters the outcomes of AI technologies. Take privacy as an example again, if we employ the principle of privacy when designing, using and implementing AI then the result is likely the ethical collection and use of personal data. On the other hand, if technologists neglect privacy, the outcomes will be different, such as data breaches and the misuse of personal data. So, whether an outcome of an AI technology is ethical or not, is conditional on whether or not an ethical principle is applied and upheld.

Intentions, Means, and Unintended Outcomes

When thinking about outcomes, technologists can make three distinctions relevant to the ethical development, design and implementation of AI. Firstly, they need to think about intention. What is the intended outcome of an AI system? The intention behind AI systems will be ethically relevant and should be designed to ensure they uphold any relevant principles related to that technology. Secondly, these innovators will need to address the ethical status of the means by which they reach their intended outcome. A technologist's intention may be ethical - for instance, to increase the safety of society, but the means by which they achieve this may be unethical - such as introducing widespread digital surveillance, which could be an unreasonable invasion of privacy and therefore unethical. Finally, taking into account the multitude of possible outcomes of AI, technologists will need to think about unintended outcomes - those outcomes which were not accounted for in the original purpose of the AI technology. For instance, an AI system might be designed to accelerate the allocation of credit applications, however an unintended outcome might be a system which unfairly prevents certain individuals from receiving credit on the basis of erroneous factors such as race or religion.

Ethical Principles

To understand the ethical concerns surrounding the general development, design, implementation and use of AI technologies, technologists should understand the following eight distinct ethical principles. And the implementation or absence of these principles will in turn direct the ethical outcomes of AI development. These outcomes can be divided into three categories: (a) human outcomes - defined as outcomes relating to, or directly impacting humanity, (b) environmental outcomes - defined as outcomes relating to or directly impacting our environment, and (c) economic outcomes - defined as outcomes relating to or directly impacting the economy.

The ethical principles that the researchers have identified are: (1) Justice, (2) Accountability (3) Fairness, (4) Human Dignity, (5) Agency, (6) Privacy, (7) Solidarity, and (8) Trust

Although other principles can be found in the myriad of literature surrounding AI ethics, for clarity the researchers have narrowed the scope of principles to these eight principles and feel that these eight principles comprehensively represent the broad range of values that are needed when developing and implementing ethics. And if utilized correctly in the design and use of AI technologies then they will be sufficient in avoiding some of the most common negative outcomes which have been illustrated in the tables below.

The values and principles for ensuring ethical AI are, as ethics arguably is, subjective, and we understand that the eight principles we have defined may have other names, or counterparts, which may be more easily understood or useful to particular users. However, in defining these principles we have tried to ensure that their implementation in AI avoids unwarranted ethical consequences for society, end-users and the environment, and so any other similar principles of use should have a similar mitigating effect on the outcomes illustrated.

Before laying out the ethical principles, it was important to make sure that these principles were comprehensive in assessing the array of possible undesirable outcomes across different industries. To do this the researchers have worked backwards from the landscape of potential outcomes relevant to AI development, to connect these outcomes to a particular ethical principle. For "one must recognize and understand the potential ethical and moral issues that may be caused by AI to formulate the necessary ethical principles, rules, guidelines, policies, and regulations for AI ethics'1. It is the hope of the researchers that a better understanding of the relationship between ethical principles and potential outcomes will have value to technologists as they look to operationalize these principles in practice.

Principles for Ethical AI

The eight ethical principles below have been given contextual and practical definitions, rather than philosophical ones, as we believe this will add some clarity to the field of ethics in AI and be of greater utility for the technologists who are tasked with their operationalization.

Justice	An AI algorithm is just when the decisions presented by the system are free from discrimination. AI use is just when the benefits and influence gained through the technology is distributed fairly.
Accountability	Accountability defines who is given the praise or blame for a specific outcome, an individual will be accountable for an action when they accept that they are responsible for a decision as well as the corresponding praise or blame for their actions.
Fairness	Bias, in the case of AI technology, is when resource allocation is subjected to undue preference of one demographic over another. Fairness is the opposite of bias, meaning an AI system does not favor one demographic over another for any reason apart from demographic characteristics. A fair AI technology would for example distribute resources in accordance with need or merit.
Human Dignity	This principle recognizes the obligations we have to humanity. AI that is developed with a respect for humanity will be in line with human values, promote prosperity and solidarity, while providing relevant safeguards for all human stakeholders.[12]
Agency	Agency is the ability to exert power over one's actions. Users of AI should be able to make informed autonomous decisions regarding AI, and remain in control of their own decisions.[1]
Privacy	Privacy is a user's ability to decide which information to disclose or not; it's the user's right to have control over their information, how it is shared and with whom.
Solidarity	Solidarity is an ethical principle that encourages humanity to share both the burdens and the benefits of technology, and to take into consideration the long-term consequences of a technology for humanity as a whole. Upholding solidarity will give a humanity centric approach to AI and ensure its use benefits humanity.[3]
Trust	Trust is an ethical principle for building a successful relationship with end-users. Transparency and explainability can be utilized to achieve trust. Trustworthy AI ensures clear communication of relevant information to all relevant stakeholders including easily understood decision logic in so far as it is useful for ensuring trust.

Subjectivity of Ethics

Deciding what is ethical is arguably a subjective matter, and the decision of which principles are prioritized and operationalized will be a matter of personal and organizational preference.

This is indeed one of the reasons it is important to have clear definitions of these principles and a better understanding of the link between certain principles and their outcomes. Dealing with subjectivity in business is far from uncommon; however, technologists should not worry about a lack of clarity surrounding what values are most important for AI development. Taking into account the principles in aggregate and developing a clear strategy for their operationalization, may not look the same for every organization, but the overall goal should be similar to ensuring that no tangible harm is done while human dignity and respect for people are upheld. As such, these eight principles are intended to be a starting point for providing comprehensive protection against unethical outcomes of AI. With this understanding, it is hoped that the right set of guidelines and principles can be implemented by technologists to avoid unwanted ethical outcomes.

How Utilizing Ethical Principles Can Avoid Particular Outcomes

In order for technologists to better understand how ethical principles can be implemented to avoid certain unwanted outcomes, it is important to see how these ethical principles relate to AI outcomes. The following three tables will explain which ethical principles are relevant to each outcome, and what outcome might result if a particular principle is not upheld in the design, use or implementation of AI. The tables are separated into the three categories, human outcomes, environmental outcomes and economic outcomes. In each outcome, we have first listed the principle, followed by the outcome and then an example of each outcome which is likely if each principle is absent.

It is important to recognize that the outcomes resulting from AI technologies are not straightforward, and each technology can have multiple outcomes unique to each circumstance and difficult to quantify. It is also a nuanced spectrum with both positive and negative impacts, but for clarity the following tables will focus on the most likely negative AI outcomes. In other words, instead of focusing on what happens when the principle is present in AI design, use and implementation, our focus is what happens when a principle is not present and upheld.

HUMAN OUTCOMES

Principle		Outcome (HUMAN)	Example
Fairness	...the absence of which commonly results in...	**Bias** - Algorithmic bias can be defined as "unintended algorithmic preference" and can result in unfair, illegal and ethically inappropriate implications.[2]	Algorithmic biases might manifest in recruitment if the output of a machine learning algorithm results in a bias in favor of a specific demographic over another as a result of prejudiced assumptions made during algorithm development and the training process. Such as with Amazon's (now defunct) hiring algorithm, which was found to discriminate against female applicants.[3]
		Unequal Access - an outcome in which the acceleration of technology benefits only a specific group in society. Leaving certain groups or individuals without access to new technologies.	Unequal Access may result because AI innovations are primarily developed as proprietary technologies - their use and availability being limited to those who develop them or can afford them. For example, research suggests that benefits from the use of AI will not be distributed evenly among developed countries and developing countries[12] - exacerbating existing injustices and geopolitical inequalities[13]. Countries that lag behind in AI development will be forced to pay technologists for the use of AI systems to remain competitive, which could slow down or hinder global access and increase inequality.
Privacy	...the absence of which commonly results in...	**Data Breaches** - Data Breaches are outcomes in which information is - intentionally or unintentionally - leaked into an untrusted environment.	Data breaches occur when websites and platforms which track and store users' information fail to adequately protect this data. This happened on a major scale with Facebook, when it was reported that more than 533 million Facebook users from 106 countries had their phone numbers, Facebook identification credentials, full names, birthdates and some email addresses uploaded onto a hacking forum.[4]
		Digital Addictions - Digital Addictions are outcomes in which end-users of technologies develop impulse control disorders that involve obsessive use of digital devices, technologies, and platforms. For instance users may be addicted to: the internet, video games, mobile devices, digital gadgets or social network platforms.[11]	Digital addictions may occur as outcomes because digital technologies use a set of persuasive and motivational techniques to keep users returning. For example, a news feed is designed to filter and display news based on your interest, to maintain engagement. And "reciprocity" techniques, where inviting more friends gains you extra 'points' contribute to addictions as once your friends are part of the network it becomes much more difficult for you or them to leave.
Accountability	...the absence of which commonly results in...	**Ethics Bluewashing -** a term coined by Luciano Floridi, refers to the potential for AI practitioners to knowingly make misleading claims about the ethics of their process, products or services in order to appear ethical.5 This form of misinformation is a significant threat to the safety of end-users as it fails to take the harms resulting from AI technologies seriously.	Ethics Bluewashing is when 'ethics becomes a performance or public relations exercise'. As with CSR greenwashing, this is when a company's social image is largely made up of photo-ops and marketing efforts that make a company seem more ethical than they are in reality. Many organizations exhibit Ethics Bluewashing by having vague AI ethics guidelines which are difficult to implement. For instance, ethics boards might be set up and principles posted on walls and on social media, whilst there is no actual strategy for operationalizing ethics in the development cycle. This serves to water down the effectiveness of adopting ethics in AI. An example of this is when Google formed a nominal AI ethics board with no actual power over ethically questionable projects.[6]
		Disinformation Attack - where AI facilitates the spreading of disinformation. The disinformation in these attacks is released quickly and broadly with the goal of creating a disruptive effect.14	Disinformation Attacks are made possible by utilizing AI technologies. AI can be used to make realistic fake profiles, which share slightly varying content and amass followers by following both fake accounts as well as real people.[15] As these technologies become more readily available and more convincing it will become increasingly difficult for people to separate truth from fiction. This can have a direct impact on society consensus without the general public necessarily being aware. Disinformation is especially prevalent related to political candidates and corporate mandates, as seen in the 2020 US election or in the manipulation of share prices with fake text generated announcements.

Agency	...the absence of which commonly results in...	**Behavioral Manipulation** - an outcome in which technology 'nudges users in a particular direction' attempting to sway users to identify with certain types of content. These interventions can range from 'being benign... to being questionable (persuasion, nudging), and possibly malign (being manipulative and coercive)'.[7] **Decision Apathy** - Decision Apathy is an outcome in which end-users of AI technologies experience a reduction of goal-directed behavior as a result of AI making decisions autonomously. Users concede control over certain decisions losing the interest or even the ability to make certain decisions for themselves.	Behavioral Manipulation may occur as a result of recommendation engines. These can encroach on individual users' autonomy, by providing recommendations that nudge people's behavior and change's people's perceptions. This outcome can make society increasingly divisive and divided. This happened on a huge scale in the US 2020 elections as people turned to their AI driven recommendation engines on social media feeds for election information.[8] Decision Apathy may materialize as an outcome when AI systems autonomously make decisions without evolving a human operator. This could result in the devolution of responsibility to machines and the inability to intervene or detect potential failures in the systems decision making. In extreme cases these systems can put "pressure on members of societies to live according to what 'the system' suggests is 'best for us' to do and not to do. Ultimately 'we may lose the autonomy to decide for ourselves how we want to live our lives'. This type of decision apathy could have profound implications for individuals and societies[9] both on a small and large scale because having control over our decisions is critical to motivation, personal growth and psychological wellness[10].
Trust	...the absence of which most commonly results in...	**xplainability and Interpretability** - Explainability and Interpretability are outcomes of a system whose inputs and operations are not visible to the user or another interested party. Often referred to as the 'black-box' nature of AI.	Trust is a key requirement for the mass adoption, and thus overall success, of artificial intelligence. Yet, according to Edelman's 2020 Trust Barometer global survey: 61% of people felt that the pace of change in technology is too fast and that governments do not understand new tech enough to regulate them effectively. A lack of trust exists when people question the decision or answers provided by AI systems. The "AI black-box" problem is when it is not clear how the AI arises at its decision. The layers of obfuscation create a lack of trust in the system because it is not understood exactly how the system comes to its conclusion. For example, training an AI system to identify birds using a neural network or deep learning method that relies on a complex system of hidden layers of nodes to learn patterns. This inability to see what the nodes have learned decreases explainability and interpretability thus reducing the trust in the overall system.

ENVIRONMENTAL OUTCOMES

Principle		Outcome (ENVIRON)	Example
Solidarity	...the absence of which most commonly results in...	**Energy Consumption -** refers to the outcome in which the development of AI technologies leads to a dramatic increase in global energy expenditure, which will directly impact the environment and climate.	Machine learning and natural language processing requires a substantial amount of data storage and analysis, and thus energy expenditure, and so the development and growth of AI systems will lead to outcomes of excessive energy consumption. For instance, training a powerful machine-learning algorithm often means running huge banks of computers for days, if not weeks. And the Department of Energy in the US estimates that data centers account for about 2 percent of total US electricity usage. Worldwide, data centers consume about 200 terawatt hours of power per year, which is more than the yearly consumption of some countries. And the forecast is for significant growth over the next decade, with some predicting that by 2030, computing and communications technology will consume between 8 percent and 20 percent of the world's electricity, with data centers accounting for a third of that.[10]
		Materials Used - refers to the fact that AI technologies require a significant amount of material, and resources for the computing power.	This outcome will directly contribute to the depletion of our natural resources. Material demands occur as outcomes because AI technologies require both plastics for device and packaging, as well as the 'intensive use and extraction of raw materials such as nickel and cobalt'.[35]
		Lack of Global Solutions - an outcome in which AI experts fail to utilise AI for the benefit of the environment.	This outcome is essentially a missed opportunity to provide solutions to climate change and other environmental problems resulting from the failure to uphold the principle of solidarity. AI efficiencies and decision making could mitigate or even solve many of the world's environmental problems. Yet, the insuler and proprietary nature of AI systems means they are rarely applied to these growing environmental issues. A lack of global cooperation is further hindered by dangerous polarization and division. Added to this bitter disagreements arising from disinformation can pose a real threat to global cooperation and unity which is badly needed to use AI systems to find and implement solutions for our environmental issues.

ECONOMIC OUTCOMES

Principle		Outcome (ENVIRON)	Example
Justice	...the absence of which commonly results in..	**Inequality** - AI technology will increase inequality between unemployed workers and technologists, creating "technological unemployment". Technologists will be endowed with the benefits of AI, enabling them to capture most of the wealth generated by AI technologies, whereas those who are not invested in AI or perform tasks that will be replaced by AI systems will lose out.	An increase in productivity that will be gained from the widespread implementation of AI systems will directly increase the income associated with capital and will worsen the gap between the owners of capital and those who rely on labor for their income. Exasperating this inequality shift is the replacement of tasks by AI systems as the transition to more automation will see a greater fall in income among unskilled workers. For example, the automation of factories and transportation will cause a great displacement of workers while at the same time increasing the profitability of the companies that implement these technologies. Under the current capitalist system where AI is developed and implemented for profit, there is no mechanism for sharing the spoils that arise from the efficiency created thus increasing the inequality in societies and between countries.11 AI implementation is a catalyst for capital income and thus a catalyst for inequality. This will be particularly bad for emerging economies that primarily rely on labor advantages causing a drop in both relative and absolute GDP for developing countries

		Allocation of Goods and Services - While AI technologies might be utilized for the allocation of goods and services, there are no universally recognized and available systems to take advantage of these efficiencies on a macroeconomic or global scale.	Intelligent legal systems, automated driving systems, computer-generated works, and intelligent decision-making processes, and agent-to-agent AI contracts have the potential to bring great efficiencies to international trade and reshape world economic order. However, the emerging technocratic power dynamics and interest groups that control these technologies will have increasing influence over their implementation. With these barriers in place, it is unlikely that the world will benefit on a whole and these systems will be implemented for the solving of global issues. Yet at the same time the power dynamics will have a profound effect on geopolitical imbalances and resource allocation of goods and services.*
Human Dignity	...the absence of which commonly results in...	**Job displacement** - Automation has the potential to disrupt labor markets in a major way. Job displacement is the outcome in which automation renders certain labor types redundant[12] which will affect workers across many professions and skill levels.[13]	AI technologies are creating "technological unemployment," or unemployment due to automation because AI can replace the need for human labor in certain tasks, leaving more and more workers unable to generate an income or in need of significant training to find a job. This could threaten social cohesion as human labor, mental and physical, becomes replaced leading to severe financial and psychological challenges as people find themselves out of work or having to change careers. This is evident across a huge variety of sectors and unemployment types. For example, in manufacturing, warehouse tasks will be replaced by robot workers, taxi drivers will be replaced with autonomous drivers, and chatbots will replace a large number of customer service tasks.
Fairness	...the absence of which commonly results in...	**Lack of Democratization** - the proprietary nature of AI development for the purpose of profits, disincentivizes the sharing and openness of AI technologies.	AI has the potential to only benefit a specific group in society, so there is a need for the democratization of AI. Large companies such as Google, Amazon and Apple have access to the massive data and computing resources needed to train and launch sophisticated AI algorithms, leaving small startups effectively blocked out of the market. As these companies amass even more wealth and resources, it will create even greater barriers to smaller innovations and further widen the gap between countries that lead in AI innovation and those that are not. The datasets and technologies need to be much more democratized rather than concentrated in the hands of a few large corporations in a few countries for the benefits to be shared more equally.
Accountability		**Lack of Explainability and Interpretability** - Explainability and Interpretability are outcomes of a system whose inputs and operations are not visible to the user or another interested party. Often referred to as the 'black-box' nature of AI.	This outcome occurs because neural networks consist of hidden layers of nodes, which process the given input and give an output to the proceeding nodes, technologies such as Deep Learning consist of huge neural networks, with many hidden layers. The issue is that we can't see what the nodes have 'learnt' so we can't know how the nodes are analyzing the data. In order to be able to trust AI systems that aid in or make decisions it will be important to understand how those systems come up with their answers. Machine learning constructs decision-making systems based on data describing human activities and therefore can be susceptible to bias and prejudices. This is a problem especially when the decision can be about life or death such as in diagnosis, or crime prediction. What is needed is a clear, comprehensible, human-readable explanation that can allow all stakeholders to understand and validate the system's decisions. Without proper explainability and interpretability it will be difficult to have clear accountability when systems go wrong, which will have financial repercussions in terms of insurance and liability.

We have defined here outcomes, for each of the ethical principles we proposed. However, it's important to note that these principles are interconnected; multiple principles may need to be implemented in order to avoid certain unethical outcomes. For example, in our human outcomes, we have said that the absence of agency commonly results in the outcome of behavioral manipulation. However, to prevent this outcome from occurring we might also need to operationalize the principle of fairness and respect for humanity into our technologies. Because of the interconnection of principles, a multiple principle approach will be necessary to prevent the outcomes presented in this paper.

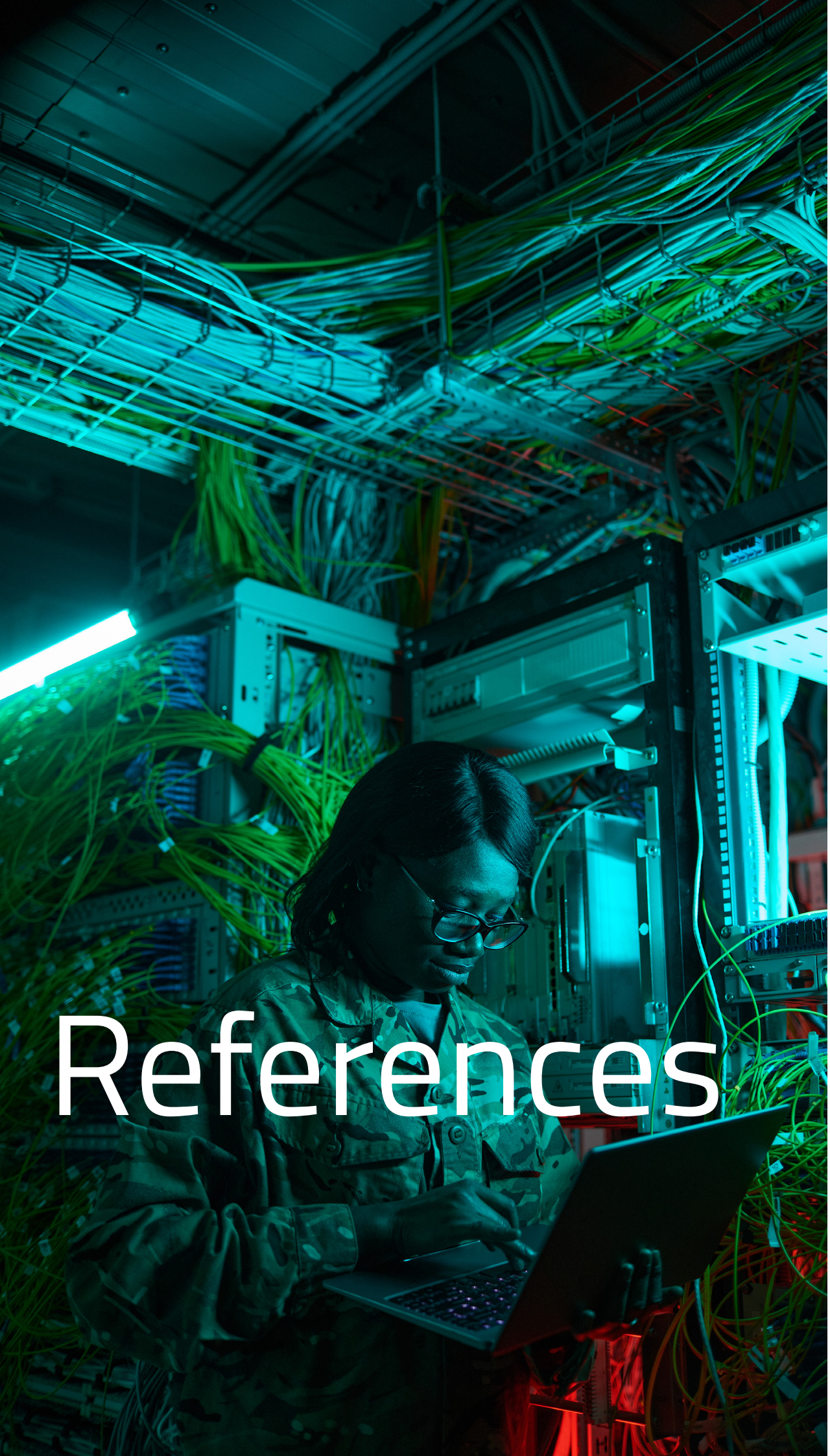
Conclusion

This paper defines a framework in which outcomes can be understood in terms of the ethical principles necessary to mitigate them. We began by outlining eight ethical principles, and then organized potential outcomes into three brackets: human outcomes, environmental outcomes and economic outcomes. Both principles and outcomes defined are intended as a starting point for providing understanding of protections against harmful AI outcomes. As technologies develop, so will the outcomes, and we may likely require the addition of further principles to keep up with the changing technological landscape.

As researchers, we hope that this information can be used by technologists to better understand the potential pitfalls that could result from the design, use and implementation of AI technologies. With a better understanding of the potential dangers of AI systems, technologists will be able to make appropriate protections to ensure that these technologies avoid unwarranted unethical outcomes and make societies better, rather than bringing harm. Understanding both how technologies can lead to harmful outcomes, as well as how these outcomes can be avoided by upholding certain ethical principles will be essential in the effective operationalization of ethics in AI.

With this grasp of principles and outcomes, we hope that technologists are able to find the right ethical frameworks and consultants to aid them in operationalizing ethics and ensuring that AI technologies exist and develop in a way that prioritizes the end-user, puts humans in the loop, and avoids introducing harms into society.

It's precisely because AI has so much potential that it presents us with both staggering benefits as well as pernicious risks. Having an understanding of these potential risks, as well as their relation to ethics can help technologists to harness the positive benefits of these technologies, and implement AI systems for the benefit of our world.



References

1. https://drive.google.com/file/d/1HP8eaFfqg5P6HuH5PvGAKS2Ogkuu5RCS/view
2. https://arxiv.org/ftp/arxiv/papers/1910/1910.12583.pdf
3. https://drive.google.com/file/d/17iKvuEK4VYi-mzqKh-lDBNhelvc6c_1k/view
4. https://www.researchgate.net/publication/340662575_Hiring_Fairly_in_the_Age_of_Algorithms
5. 6https://www.researchgate.net/publication/331408062_Facebook_User's_Data_Security_and_Awareness_A_Literature_Review
6. https://link.springer.com/article/10.1007/s00146-020-00950-y
7. https://firstmonday.org/ojs/index.php/fm/article/view/11431/9993
8. https://www.ethicsdialogues.eu/2019/06/12/the-underdog-in-the-ai-ethical-and-legal-debate-human-autonomy/
9. https://link.springer.com/chapter/10.1007/978-3-030-50585-1_2
10. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=6839&context=libphilprac#:~:text=Digital%20addiction%20referred%20to%20an,gadgets%2C%20and%20social%20network%20platform.
11. https://drive.google.com/file/d/1onx8b4b2itIs3WiBsGf4CbVTDsCg1e51/view
12. https://drive.google.com/file/d/155LLUJ4ru4yFxpTcqFAct40qxwaaFiAi/view
13. https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/
14. https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/
15. https://drive.google.com/file/d/1YY6lGnTs1TGxlrRXi_n1nqlmsY9U15pw/view
16. https://drive.google.com/file/d/1gkIA2PAr40Wb6zqKsLGPCGgsdcbvP3y8/view
17. https://www.nber.org/system/files/chapters/c14018/c14018.pdf
18. https://link.springer.com/article/10.1007/s00146-020-00950-y
19. https://firstmonday.org/ojs/index.php/fm/article/view/11431/9993
20. https://drive.google.com/file/d/1WRZGcgn6SEeCQDT7zs7UmnnkSy93fhxV/view
21. https://drive.google.com/file/d/17iKvuEK4VYi-mzqKh-lDBNhelvc6c_1k/view
22. https://www.researchgate.net/publication/340662575_Hiring_Fairly_in_the_Age_of_Algorithms
23. https://www.researchgate.net/publication/331408062_Facebook_User's_Data_Security_and_Awareness_A_Literature_Review
24. https://link.springer.com/article/10.1007/s00146-020-00950-y
25. https://firstmonday.org/ojs/index.php/fm/article/view/11431/9993
26. https://www.ethicsdialogues.eu/2019/06/12/the-underdog-in-the-ai-ethical-and-legal-debate-human-autonomy/
27. https://link.springer.com/chapter/10.1007/978-3-030-50585-1_2
28. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=6839&context=libphilprac#:~:text=Digital%20addiction%20referred%20to%20an,gadgets%2C%20and%20social%20network%20platform.
29. https://drive.google.com/file/d/1onx8b4b2itIs3WiBsGf4CbVTDsCg1e51/view
30. https://drive.google.com/file/d/155LLUJ4ru4yFxpTcqFAct40qxwaaFiAi/view
31. https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/
32. https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/

33. https://drive.google.com/file/d/1WRZGcgn6SEeCQDT7zs7UmnnkSy93fhxV/view
34. https://drive.google.com/file/d/1WRZGcgn6SEeCQDT7zs7UmnnkSy93fhxV/view
35. https://drive.google.com/file/d/1GviHiDx6E6sBmF3U166OXZJFS8bktbsk/view
36. https://drive.google.com/file/d/1onx8b4b2itIs3WiBsGf4CbVTDsCg1e51/view
37. https://drive.google.com/file/d/1gkIA2PAr40Wb6zqKsLGPCGgsdcbvP3y8/view
38. https://www.nber.org/system/files/chapters/c14018/c14018.pdf
39. https://drive.google.com/file/d/1YY6lGnTs1TGxlrRXi_n1nqlmsY9U15pw/view
40. https://drive.google.com/file/d/1aFF0fv_Jgy63YBklfWZcKeCau-Op4TKW/view
41. https://www.researchgate.net/publication/225249319_The_responsibility_gap_Ascribing_responsibility_for_the_actions_of_learning_automata#:~:text=This%20phenomenon%20is%20known%20as,agent%20of%20responsibility.%20...
42. https://drive.google.com/file/d/13VRCHSkIqOozeYxYSX-Vqi04qwGqsVfW/view
43. https://drive.google.com/file/d/1GviHiDx6E6sBmF3U166OXZJFS8bktbsk/view

1. https://arxiv.org/abs/1905.06876
2. https://drive.google.com/file/d/17iKvuEK4VYi-mzqKh-lDBNhelvc6c_1k/view
3. https://www.researchgate.net/publication/340662575_Hiring_Fairly_in_the_Age_of_Algorithms
4. https://www.researchgate.net/publication/331408062_Facebook_User's_Data_Security_and_Awareness_A_Literature_Review
5. https://drive.google.com/file/d/1YY6lGnTs1TGxlrRXi_n1nqlmsY9U15pw/view
6. https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/
7. https://link.springer.com/article/10.1007/s00146-020-00950-y
8. https://firstmonday.org/ojs/index.php/fm/article/view/11431/9993
9. https://drive.google.com/file/d/1WRZGcgn6SEeCQDT7zs7UmnnkSy93fhxV/view
10. https://www.wired.com/story/ai-great-things-burn-planet/
11. https://psycnet.apa.org/record/2014-07087-000
12. https://drive.google.com/file/d/1gkIA2PAr40Wb6zqKsLGPCGgsdcbvP3y8/view
13. https://www.nber.org/system/files/chapters/c14018/c14018.pdf

1. https://www.researchgate.net/publication/340115931_Artificial_Intelligence_AI_Ethics_Ethics_of_AI_and_Ethical_AI
2. https://www.edelman.com/trust/2020-trust-barometer

ETHICAL INTELLIGENCE

This paper is written in collaboration with Ethical Intelligence as part of a larger research study on the operationalization of Ethics in business.

Contact

Jesse Arlen Smith, President
Aiforgood Asia STO,
Hanoi Vietnam

jesse@aiforgood.asia
www.aiforgood.asia