



13th May 2025

THE FRONTIERS IN HUMAN CENTERED AI PARADIGM

Introduction

As the field of emerging technologies particularly Artificial Intelligence continues to expand in innovation and dynamism it has become critical that these innovations are not left at the whims and caprice of tech and AI companies alone. Humanity should always be the yardstick of evaluating the efficacy of any innovation if technology is to serve the common good of mankind. The question remains why tech if not to serve humanity and why tech regulations in the age of liberalism and global human rights violations. The disruptive nature of AI meant that human centered should be the guiding principle. AI algorithms must not be entrusted with only data in which they are built and trained, they should be rigorously fact checked on which purpose they are trained to serve from the stage of design, development and development of AI system.

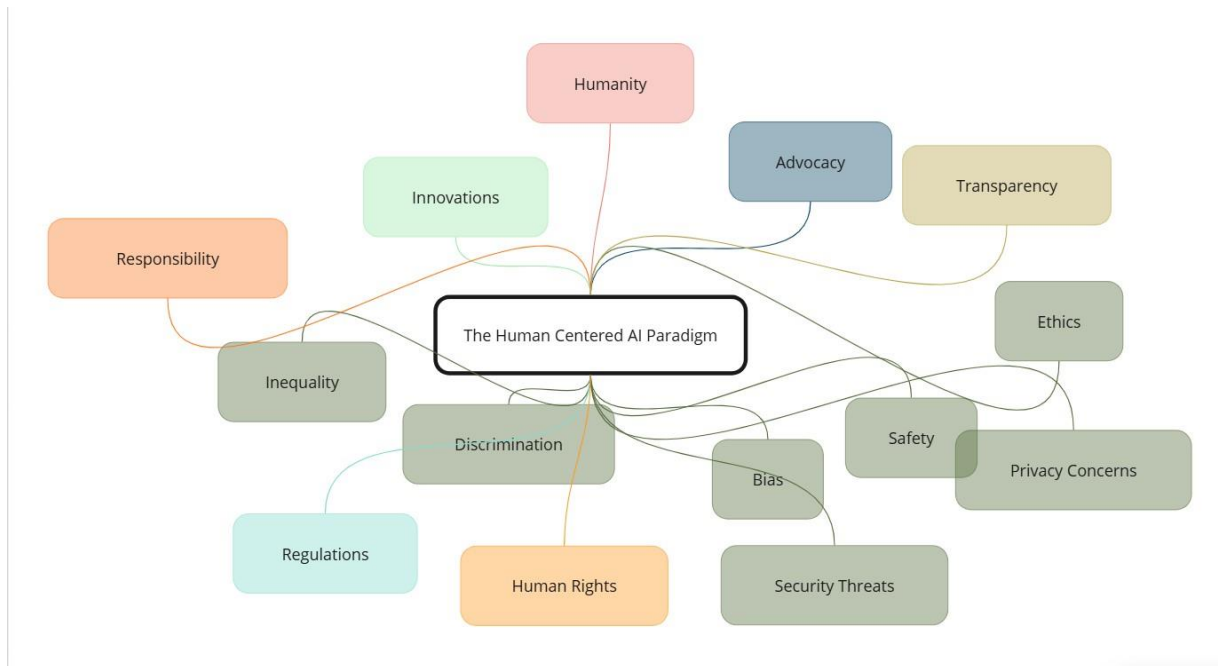
Humanity has been subjected to being at the end of lots of devastating technology innovation such as use of guns, bullets, bombs and weapons of mass destructions (atomic, nuclear and biological). The aftermath of the First World War in 1918 brought about horrors of how mankind has in an attempt to display military superiority committed atrocious killings and irreversible destructions. This carried on into the Second World War where military capacity became the determining factor in a show of force disregarding the humanitarian casualties. The historical bombings in Hiroshima and Nagasaki Japan in 1945 further re-emphasizes the need for humanity to be at the forefront of every tech innovation and use.

Over the years modern wars have not been any different save for more sophistication in precision bombs, missiles and artilleries capable of travelling far distances and reaching remote targets. The wars in the Middle East such as the Gulf war 1990-1991, the US invasion of Afghanistan in 2001, the age long West bank conflict between Israel and Palestine, The ongoing Ukrainian vs Russia war.

Human-centered AI (HCAI) refers to the development of artificial intelligence (AI) technologies that prioritize human needs, values, and capabilities at the core of their design and operation. This approach ensures teams create AI systems that enhance human abilities and well-being rather than replacing or diminishing human roles. It addresses AI's ethical, social, and cultural implications and ensures these systems are accessible, usable, and beneficial to all segments of society. HCAI is linked to Human-AI interaction, a field that examines how AI and humans communicate and collaborate¹.

¹ Interaction Design Foundation, 'Human-Centered AI (HCAI)', <<https://www.interaction-design.org/literature/topics/human-centered-ai> > accessed 20th August 2024

The Human Centered AI Paradigm



The human centered paradigm (HCP) depicts the central goal of advancing human centered AI while at same time unveiling threats to this possible advancement. It captures the negatives undermining human centered AI. It exemplifies both the road map for an AI ecosystem consistent in advancing humanity and social good as well as the pitfalls of AI systems. Achieving human centered AI will be evasive if concerns as these threats are not considered and tackled frontally.

Effect of AI Systems without Human Centeredness

When AI systems are developed as a by-product of human imagination without making humanity the center of its development guideline, this omission leads to uncontrollable consequences and high-risk effects on humanity. These risks may either be in isolation or cumulative thereby causing deep infraction to human rights. An effective human centered policy for AI development will guardrail against inhuman and unintended consequences to humanity and mankind arising from use of AI systems.

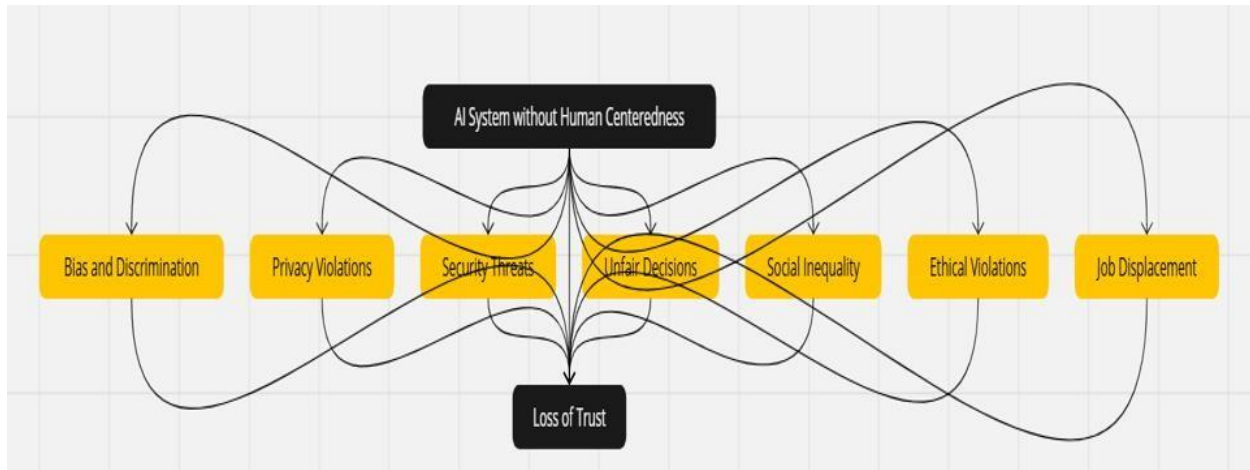
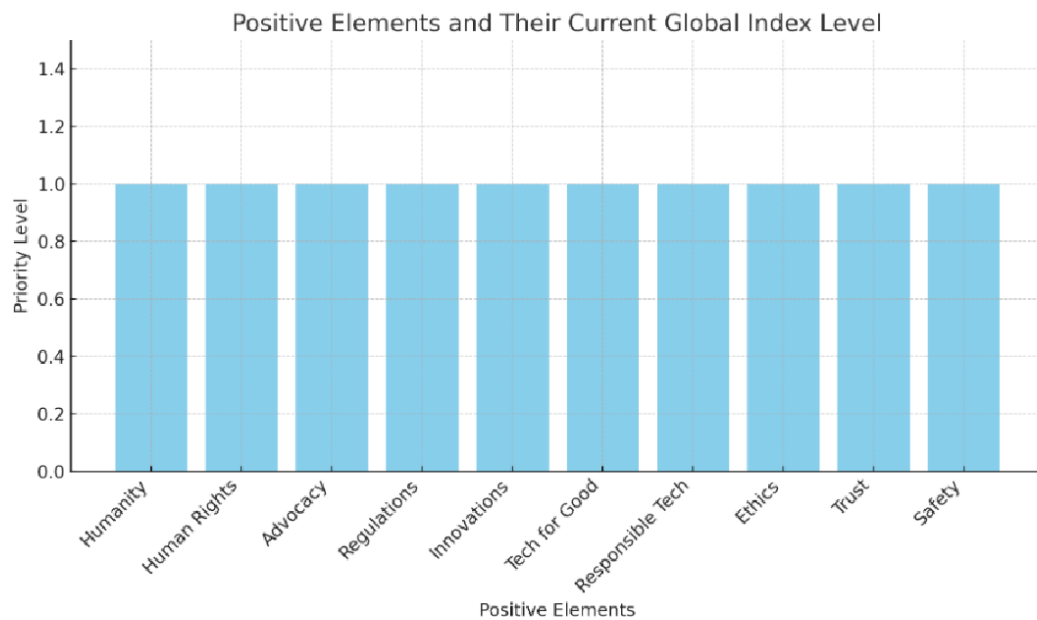


Diagram showing the chain effect of AI systems devoid of human centeredness

Positive Elements and Their Current Global Index Level



Element	Priority level
Humanity	High
Human Rights	High
Advocacy	Medium
Regulations	High
Innovations	Medium
Tech for Good	High
Responsible Tech	Medium
Ethics	High
Trust	High
Safety	High

In 2023, the Center for AI Safety in a statement signed by more than 350 executive researchers and engineers working in AI warned that the artificial intelligence they were building might one day pose an existential threat to humanity and should be given top priority with such risks as pandemics and nuclear wars².

According to the UN High Commissioner for Human Rights Volker Türk,

To be effective, to be humane, to put people at the heart of the development of new technologies, any solution - any regulation - must be grounded in respect for human rights. AI technologies that cannot be operated in compliance with international human rights law must be banned or suspended until such adequate safeguards are in place. Third, existing regulations and safeguards need to be

² Kevin Roose, 'A.I. poses Risk of Extinction Industry Leaders Warn' (The New York Times 30 May 2023), <<https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html>> accessed 20th August 2024

implemented. For example, frameworks on data protection, competition law, and sectorial regulations, include for health, tech or financial markets. A human rights perspective on the development and use of AI will have limited impact if respect for human rights is inadequate in the broader regulatory and institutional landscape³.

Bias in algorithms is a costly human oversight. This is due to its immense impact on marginalized and underrepresented communities. As artificial intelligence (AI) continues to scale across industries and functions, it can be riddled with unconscious biases that do more harm than good while AI ethicists and responsible AI practitioners often speak about the need for more transparency and accountability in the AI lifecycle⁴. Thus, advocacy is crucial for advancing human centered tech and AI.

It has been argued that AI is too important not to regulate and that thoughtful regulation can promote reliable, robust, and trustworthy AI applications⁵. Given the cross-border nature of the digital economy, global AI regulatory frameworks and technical standards need to operate across nations and regions. Increased global alignment on AI regulation helps to facilitate the adoption, use, and interoperability of AI technologies across different jurisdictions⁶.

³ OHCHR, 'Artificial intelligence must be grounded in human rights, Volker Türk, UN High Commissioner for Human Rights' (July 12, 2023) available at <<https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>> accessed 20th August 2024

⁴ World Economic Forum, 'How can AI support diversity, equity and inclusion' (March 1, 2022), <<https://www.weforum.org/agenda/2022/03/ai-support-diversity-equity-inclusion/>> accessed 2nd October 2024 ⁵ Google, 'Building a responsible regulatory framework for AI' available at <<https://ai.google/static/documents/building-a-responsible-regulatory-framework-for-ai.pdf>> accessed 3rd October 2024

⁶ ibid

According to Forbes Advisor statistics, the global AI market is projected to reach as staggering \$1,339 billion by 2030⁷. Artificial intelligence is developing at an incredibly fast pace. With AI advancement and growth, it has become pertinent to consider how unintended bias can be mitigated, how transparency can be guaranteed and how human privacy can be protected. This has become pertinent as Innovation continues to outpace regulations globally⁸.

Although Artificial intelligence has become a growing phenomenon, reliance on innovation without more in advancing AI could be counterproductive to humanity. The human factor is integral for the future of Artificial Intelligence. Hence recognizing that humans are central to its design, operation and use requires that such systems amplify and augment humanity. Supportive tech provides new ways of preventing harm. Tech for good therefore is important to help solve problems, optimize benefits and help tackle the pressing social and environmental challenges of our time⁹.

Responsible technology is the active consideration of values, unintended consequences, and negative impacts of technology. Responsible tech includes a wide variety of voices in the adoption and deployment process, and seeks to manage and mitigate potential risk and harm to all communities affected by that technology¹⁰. In the words of Deb Donig Professor of Ethical Technology, California Polytechnic State University, "When people talk about responsible technology what they generally are interested in is the intersection of technological production and culture on the one hand and on the other hand, human values"¹¹.

⁷ Katherine Haan, '22 Top AI Statistics and Trends in 2024'
<<https://www.forbes.com/advisor/business/ai-statistics/>> accessed 22nd
October 2024

⁸ Michael Magrath, 'Can Regulators keep pace with Innovation?'
<<https://www.corporatecomplianceinsights.com/can-regulators-keep-up-with-innovation/>> accessed 22nd
October 2024

⁹ World Economic Forum, Tech for good: What it means and how we can deliver on it (Ming Tan) available at
<https://www.weforum.org/agenda/2023/03/tech-for-good-what-does-it-mean-and-how-we-can-deliver-on-it/>

¹⁰ MIT Technology Review Insights, 'The State of responsible technology' (January 2023),
<https://wp.technologyreview.com/wp-content/uploads/2023/01/MITTR-Thoughtworks_v5-2-1.pdf>
accessed 22nd October 2024

¹¹ ibid

Our society is undergoing a Fourth Industrial Revolution¹², in which powerful technologies such as artificial intelligence (AI), the internet of things (IoT) and augmented reality have the potential to create lasting societal benefits. However, without the right guard rails, these technologies can also cause immense harm. Organizations have a responsibility to design, develop, procure, deploy and use technologies in a responsible manner. As an initial step towards more responsible

innovation, dozens of groups from industry, civil society and government have published ethical technology principles, particularly on AI¹³. These values and frameworks offer a foundation for what responsible technology innovation outcomes might look like. Yet the need to define how technology products can be ethically made remains¹⁴.

In a world where technology continues to evolve at a rapid pace, the importance of incorporating ethical considerations into tech innovations becomes increasingly crucial. The increasing role of artificial intelligence (AI) and machine learning technology in our lives has raised an enormous number and variety of ethical challenges. Ethical reasoning is another essential aspect of human-centric AI training. As AI systems become more autonomous and influential, it is crucial that they are capable of making decisions that align with ethical principles and human values. Ethics is crucial to guarantee fairness, transparency, explainability, privacy, accountability by AI systems thereby establishing a ground for human centered AI.

As artificial intelligence (AI) becomes increasingly integrated into various aspects of our lives, it is crucial to ensure that these systems are not only effective but also trustworthy and aligned with human values. This is where human-centric AI training comes into play, emphasizing the importance of empathy, ethical reasoning, and alignment with human values in the development and deployment of AI systems. This approach is particularly critical in sensitive applications such as eldercare, therapy, and education, where AI has a direct and significant impact on human lives¹⁵.

¹² World Economic Forum, 'Responsible Use of Technology: The Microsoft Case Study White Paper' (February 2021) available at https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf accessed 25th October 2024

¹³ ibid pg 4

¹⁴ ibid pg 4

¹⁵ Daisy Thomas, 'Building Trust in AI: The Crucial Role of Human-Centric AI Training' (July 8, 2024) available at <https://www.linkedin.com/pulse/building-trust-ai-crucial-role-human-centric-training-daisy-thomas-5loje/>

Human-centered design (HCD) can play a critical role in facilitating trust and transparency in AI systems. By placing a strong emphasis on understanding the needs and experiences of users, HCD can help ensure that AI systems are designed in a way that is transparent and trustworthy¹⁶.

Safety of AI systems is integral to achieving human centered AI. When AI systems are devoid of bias, discrimination, harm, privacy issues and other ethical challenges associated with AI, it leads to enthrone human values and regard which invariably means such AI can be utilized to amplify and augment humanity. AI safety is a comprehensive process, both dynamic and collaborative in nature. Multiple stakeholders, including policymakers, AI developers, users, and compliance experts, need to be appropriately involved in the process to provide consistent vigilance, innovation, and commitment to the ethical principles that will be key in ensuring all AI development leads to a positive contribution to society¹⁷.

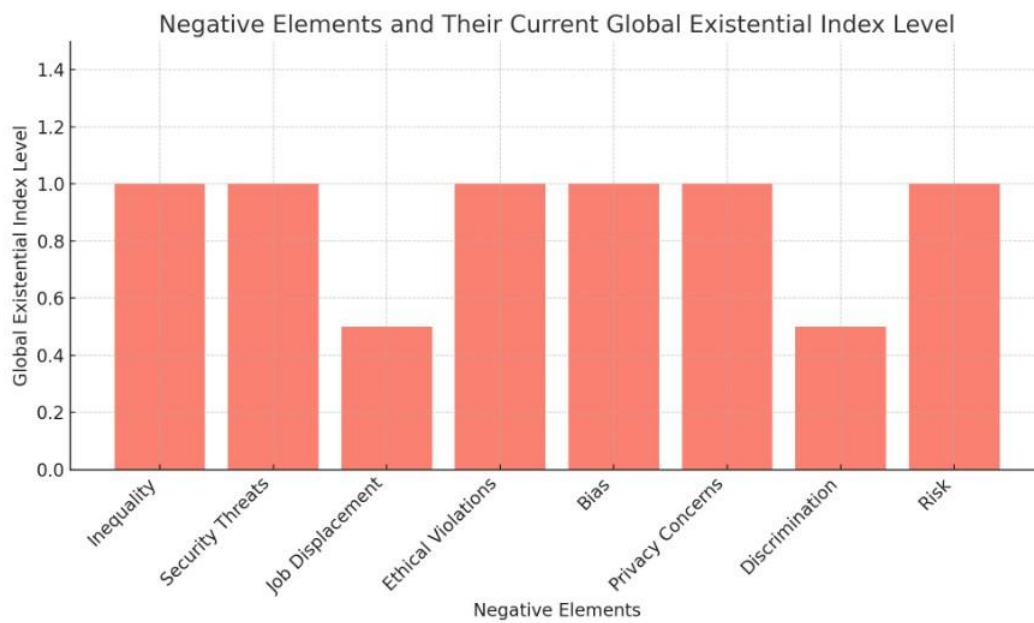
The importance of appropriate AI safety can be gauged from the risks and consequences that can easily be associated with unregulated AI development. Again, appropriately identifying the risks is an essential part of placing AI safety at the heart of the overall AI development process.

¹⁶ Akhtar, M.A.K., Kumar, M., Nayyar, A. (2024). The Role of Human-Centered Design in Developing Explainable AI. In: Towards Ethical and Socially Responsible Explainable AI. Studies in Systems, Decision and Control, vol 551. Springer, Cham.

¹⁷ Anas Baig (Securiti), What is AI safety? Available at <https://www.securiti.ai/ai-safety/> accessed 25th October 2024

Negative Elements and their Global Existential Index Level Elements

Global Existential Index	
Inequality	High
Security Threats	High
Job Displacement	Medium
Ethical Violations	High
Bias	High
Privacy Concerns	High
Discrimination	Medium
Risk	High



The relationship between Artificial intelligence (AI) technologies and the uneven distribution of benefits, opportunities, and resources among individuals or groups, leading to disparities in various areas such as access to AI technology, employment opportunities, wealth accumulation, and social impact highlights the deep-rooted inequality caused by Artificial intelligence.

In an era marked by increasingly frequent and harmful cyber-attacks, artificial intelligence (AI) adds an extra layer of complexity to an already entropic environment. Recent debate revolves mostly around the security concerns and challenges posed by AI¹⁸. The integration of AI into cyber security processes is not without risks as AI, particularly generative AI, can be used to advance adversarial capabilities, such as phishing, malware development, and deep fakes. In recent years, major network security incidents have occurred frequently, and network attacks and security threats are everywhere. Security threats in cyberspace include physical infrastructure, network information systems and social media security information. In personal information and property security and so on constitute a serious threat.

AI will affect almost 40 percent of jobs around the world, replacing some and complementing others according to a new analysis by the International Monetary Fund (IMF)¹⁹. In advanced economies, about 60 percent of jobs may be impacted by AI²⁰. The advancement and implementation of artificial intelligence technologies lead to the loss or transformation of certain types of jobs, as machines and automated systems become capable of performing tasks previously done by humans.

¹⁸ Alexandra Borgeaud, Artificial intelligence (AI) in cybersecurity - statistics & facts available at <https://www.statista.com/topics/12001/artificial-intelligence-ai-in-cybersecurity/#topicOverview> accessed 26th October 2024

¹⁹ Kristalina Georgieva, AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity available at <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity> accessed 25th October 2024

²⁰ ibid

A study released by MIT Technology Review found that training a 'regular' AI using a single high-performance graphics card has the same carbon footprint as a flight across the United States²¹. Training a more sophisticated AI was even worse, pumping five times more CO₂ into the atmosphere than the entire life cycle of an American car, including its manufacturing. Whether it's the latest AI or machine learning algorithm that's active on a system, a new 5G network deployed at a commercial building or people streaming the latest Twitch gaming video, people generate and consume a lot of data. All that data must be captured, stored, analyzed and sent back out, which requires significant amounts of processing power.

Ethical Violations

It is important to understand how ethical challenges can pose a threat to a human centered AI. Where ethical violations occur, they cause a deep infringement on humanity. These violations occur in different forms from job displacement, privacy challenges, bias, security issues, lack of transparency (explainability), lack of accountability, deep fakes, misinformation, exploitation of intellectual property, loss of social connection, unfair competition between tech companies for AI superiority.

i. Displacement of Jobs

Job displacement is a concern that is frequently cited in discussions surrounding AI. There is fear that automation will replace certain aspects or entire job roles, causing unemployment rates to spike industries. According to CompTIA's Business Technology Adoption and Skills Trends report, 81% of U.S. workers have recently seen articles which focus on the replacement of workers with AI. The same report

²¹ Thomas Griffin, Why We Should Care About The Environmental Impact Of AI available at <https://www.forbes.com/councils/forbestechcouncil/2020/08/17/why-we-should-care-about-the-environmental-impact-of-ai/> accessed 26th October 2024

found that 3 out of 4 workers are very or somewhat concerned about how automated technologies will impact the workforce²².

ii. Privacy Issues

Training of AI models requires massive amounts of data. There is currently little insight into how the data is being collected, processed and stored which raises concerns about who can access these data and how they can use it. Also, there is usually no template for disposal of such data acquired after use or the proposed storage duration of these data collected. There are other privacy concerns surrounding the use of AI in surveillance. Law enforcement agencies use AI to monitor and track the movements of suspects. While highly valuable, many are worried about the misuse of those capabilities in public spaces, infringing upon individual rights to privacy.

iii. Bias

There is another ethical concern surrounding AI bias. Although AI does not inherently come with bias, systems are trained using data from human sources and deep learning which can lead to the propagation of biases through technology. For instance, an AI hiring tool could omit certain demographics if the data sets used to train the algorithm contained a bias against a particular group. This could also have legal implications if it leads to discriminatory practices.

iv. Security

Security remains a top priority when it comes to AI (and really any branch of computer science). Lax security can have a wide-ranging impact. For example, AI is susceptible to malicious attacks which can compromise outcomes. The Cybersecurity Infrastructure and Security Agency (CISA) references documented instances of attacks leading to misbehaviors in autonomous vehicles and the hiding of objects in security camera footage. Experts and governmental entities are urging for more security measures to limit potentially negative effects.

²² Ashley Watters, 11 Common Ethical Issues in Artificial Intelligence available at <https://connect.comptia.org/blog/common-ethical-issues-in-artificial-intelligence> accessed 26th October 2024

v. Explainability

It is simply not enough to put AI tools out into the world and watch them work. It can be particularly important to understand the decision-making process with certain AI applications. In some cases, it can be difficult to understand why certain AI tools came to conclusions. This can have sizeable implications, especially in industries such as healthcare or law enforcement where influencing factors must be considered, and real human lives are at stake.

vi. Accountability

The increasing prevalence of AI in all industries means that we use AI tools to make decisions daily. In cases where those decisions lead to negative outcomes, it can be difficult to identify who is responsible for the results. Are companies on the hook for validating the algorithms of a tool they buy? Or do you look to the creator of an AI tool? The quest for accountability can be a deep rabbit hole which can make it difficult to keep people and companies accountable.

vii. Deepfakes

The usage of deepfakes creates ethical concerns. Deepfakes are now able to circumvent voice and facial recognition which can be used to override security measures. One study even showed that a Microsoft API was tricked more than 75% of the time using easily generated deepfakes²³. Other ethical challenges arise when it comes to impersonation. The usage of deepfakes to sway public opinion in political races can have far-reaching implications. There is also concern over whether deepfakes could be used to influence the stock market if a CEO were believed to be making decisions or taking actions that were considered questionable. With no oversight and easy access to the software, the abuse of deepfakes presents a significant security gap.

²³ ibid

viii. Misinformation

Misinformation has a way of creating social divides and perpetuating untrue opinions to the detriment of organizations and others. A topic that gained scrutiny in the context of the political upheaval seen in recent years, misinformation can affect public opinion and cause severe reputational damage. Governments and political actors around the world, in both democracies and autocracies, are using AI to generate texts, images, and video to manipulate public opinion in their favor and to automatically censor critical online content²⁴. Once misinformation becomes widely shared on social media, it can be difficult to determine where it originated and challenging to combat. AI tools have been used to spread misinformation, making it appear as though the information is legitimate, when it is in fact not.

Venezuelan state media outlets, for example, spread pro-government messages through AI-generated videos of news anchors from a nonexistent international English-language channel; they were produced by Synthesia, a company that produces custom deepfakes²⁵. And in the United States, AI-manipulated videos and images of political leaders have made the rounds on social media. Examples include a video that depicted President Biden making transphobic comments and an image of Donald Trump hugging Anthony Fauci²⁶.

²⁴ Tate Ryan-Mosleyarchive, How generative AI is boosting the spread of disinformation and propaganda available <https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/> accessed 27th October 2024

²⁵ ibid

²⁶ Steve Contorno and Donie O'Sullivan, (CNN), DeSantis campaign posts fake images of Trump hugging Fauci in social media video available at <https://edition.cnn.com/2023/06/08/politics/desantis-campaign-video-fake-ai-image/index.html> accessed 27th October 2024

ix. Exploitation of Intellectual Property

A lawsuit against ChatGPT involving several popular writers who claim the platform made illegal use of their copyrighted work²⁷ has brought attention to the issue of AI exploitation of intellectual property. Several authors, including favorites as Jodi Picoult and John Grisham sued OpenAI for infringing on copyright by using their content to train their algorithms. The lawsuit further claims that this type of exploitation will endanger the ability of authors to make a living from writing. This kind of exploitation has owners of intellectual property concerned about how AI will continue to impact their livelihoods.

x. Loss of Social Connection

While AI has the potential to provide hyper-personalized experiences by customizing search engine content based on your preferences and enhancing customer service through the use of chatbots, there is concern that this could lead to a lack of social connection, empathy for others and general well-being. If all you see on social media are opinions that reinforce your own, you're unlikely to develop a mindset that allows you to empathize with others and engage in actions for social good. Technology has become the architect of our intimacies. Online, we fall prey to the illusion of companionship, gathering thousands of Twitter and Facebook friends, and confusing tweets and wall posts with authentic communication. But this relentless connection leads to a deep solitude²⁸. Genuine, organic social interactions can become degraded through constant exposure to illusory meaningful exchanges with AI.

²⁷ The New York Times, A recent lawsuit against ChatGPT involving several popular writers who claim the platform made illegal use of their copyrighted work, available at <https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html> accessed 27th October 2024

²⁸ Sherry Turkle, Alone Together: Why We Expect More from Technology and Less from Each Other available at <https://www.sherryturkle.com/alone-together> accessed 27th October 2024

xi. Unfair competition between tech companies (AI superiority race)

New technologies present companies, tech giants and startups alike, with a particular challenge because there is a constant race to innovate. Often, success is determined by a company's ability to be the first to release a particular technology or application. When it comes to AI systems, companies often aren't taking the time to ensure their technology is ethically designed or that it contains stringent security measures. In a world increasingly defined by algorithms and automation, the race for artificial intelligence (AI) dominance has become a breakneck sprint. Nations, corporations and research institutions are pouring resources into developing ever-more sophisticated AI systems, each vying for a coveted edge in this high-stakes competition. This frantic pursuit has been aptly dubbed the "Code War," a battleground where lines between collaboration and competition blur in the relentless pursuit of AI supremacy²⁹.

At the forefront of this mad race stand the titans of technology: Google, Amazon, Meta and Microsoft, each wielding vast resources and intellectual capital. These tech giants are not only developing their own AI systems but also acquiring promising startups and talent, further consolidating their hold on the AI landscape³⁰.

xii. Discrimination

AI systems when deployed for use often times tend to be discriminatory especially where its dataset used in training it was not based on diverse data. For instance, is use of AI system in the airport to conduct facial recognition. Where same is built only for use by whites, it therefor means that persons with black skin will not be recognized. Similarly, where AI is used in criminal investigation a particular race or colour might be subjected to unfair criminal harassment.

²⁹ Giri Devanur (Forbes), The Code War: A Mad Race For AI Supremacy available at <https://www.forbes.com/councils/forbesbusinesscouncil/2024/01/10/the-code-war-a-mad-race-for-ai-supremacy/> accessed 28th October 2024

³⁰ *ibid*

AI is already making rapid progress in critical domains such as hacking, social manipulation, and strategic planning, and may soon pose unprecedented control challenges³¹. To advance undesirable goals, AI systems could gain human trust, acquire resources, and influence key decision-makers. To avoid human intervention, they could be capable of copying their algorithms across global server networks. Large scale cybercrime, social manipulation, and other harms could escalate rapidly³². In open conflict, AI systems could autonomously deploy a variety of weapons, including biological ones. Consequently, there is a very real chance that unchecked AI advancement could culminate in a large-scale loss of life and the biosphere, and marginalization or extinction of humanity³³.

Significance of Interconnectivity of Human Centered AI Paradigm

The Human Centered AI Paradigm mind map shows how various components interact, contribute and compete with the overarching goal of ensuring Human Centered AI. These relationships, impact and their significance is broken down as follows:

i. Humanity and Human Rights:

Central to Human Centered AI is the focus on human well-being and dignity. Ensuring that AI respects and enhances human life is fundamental to achieving human centered AI. Humanity here becomes the center piece of AI developments guardrails. If humanity is centralized in AI, it ordinarily would infer that AI developments should be advanced for social good and human benefits. In the same vein, AI systems must respect and uphold human rights, ensuring fairness, justice, and equality. Human rights require that humanity is protected, and this demands regulations to align with human rights provisions.

³¹ University of Oxford News and Events, World leaders still need to wake up to AI risks, say leading experts ahead of AI Safety Summit available at <https://www.ox.ac.uk/news/2024-05-20-world-leaders-still-need-wake-ai-risks-says-leading-experts-ahead-ai-safety-summit> accessed 28th October 2024

³² *ibid*

³³ *ibid*

ii. Advocacy and Regulations:

Advocacy groups play a crucial role in voicing concerns about AI impacts and pushing for ethical practices. This has led to global call for a multi stakeholder approach in AI regulation wherein advocates and civil society organizations call for responsible and ethical AI developments. Advocacy groups help curtail the unregulated excesses of AI and tech companies. Their voice in championing a society friendly AI regime help champion ethical values and call for more regulated AI.

Effective regulations ensure that AI development and deployment align with societal values and ethical standards, preventing misuse and harm. Without guardrails, AI regulation could become a Frankenstein monster. Regulations are the bedrock of a sustainable AI. AI is a double-edged sword and if not regulated possess the capability to cause more harms than anticipated. Regulations can also drive growth and development in AI.

iii. Innovations and Tech for Good

Innovations globally should be encouraged and advanced. In this advancement tech for good should form the cornerstone and pillar of every innovation. How can we make lives easier, things faster and the world more prosperous should be questions that need be answered in the positive. Technological advancements should be geared towards solving real-world problems and improving lives. Why tech if not to advance and better our common existence. This principle emphasizes the use of technology to address societal challenges and enhance well-being. Tech and AI are like two-edged sword and it either serves for good or used to advance evil. Innovation without serving the good of mankind could mean otherwise.

iv. Responsible Tech and Ethics:

Responsible Tech is crucial in the world driven by technology. Developing technology responsibly involves considering long-term impacts, including social, economic, and environmental effects. If tech is to be developed responsibly, then it must align to certain standards and ethical requirements.

Ethical frameworks guide the development and deployment of AI to ensure it aligns with moral values and societal norms. These ethical standards require that tech and AI is devoid of challenges like bias, discrimination, privacy issues, underrepresentation, marginalization and inequality. Ethical standards are thresholds that tech and AI companies need to attain in establishing responsible tech ecosystem. If tech is responsible, then such tech must align with ethical standards.

v. Bias and Discrimination:

Bias forms a major challenge to AI system development and this is more so as tech and AI systems are designed and trained with datasets. Identifying and mitigating biases in AI systems is essential to prevent unfair treatment of individuals or certain groups. If datasets in which these AI systems are built on are sieved of biases, then the outcomes in use of such AI system will not be discriminatory. Imagine if an AI system is built on a biased dataset that favours men over women, the able over the disable, the poor over the rich, the white over the blacks etc., such AI system will be highly discriminatory in its outcome. If discrimination must be avoided by tech and AI systems, bias in data set and training of such AI system must be eradicated completely. Regular audit for bias in AI systems in design, development stage needs to be carried out to ensure they are free from bias. Hence its crucial AI is designed to avoid reinforcing or deepening existing societal inequalities and discrimination.

vi. Privacy Concerns and Security Threats:

Privacy Concerns is a big threat in use and adoption of AI. These fears stem from unauthorized access, use, storage and retention of user data. AI thrives in data accumulation to update its knowledge and working mechanism to suite end users demands and data obtained from users are usually stored for a prolonged term. Users do not also know how such data will be used and if they will fall into the hands of third parties. Often time user policy portrayed by tech companies in use of user data are often not adhered to. Safeguarding personal data and ensuring privacy is crucial to maintaining trust in AI systems.

Security Threats also pose another deep challenge in AI. These threats could be towards altering personal data, mutilating data, destroying data or data theft and authorized access to data. Addressing security vulnerabilities protects users from all potential harms, such as data breaches or malicious AI usage. Also, the use of encryption to protect third party access to personal data is really an integral component of safeguarding personal data from unauthorized access.

vii. Safety and Inequality:

Safety is vital in the use of AI systems. User safety must be guaranteed by tech and AI companies through safety by design adherence. Unintended consequences of using an AI system must be understudied and tackled at the design stage before deployment to end users. Ensuring the safety of AI systems involves rigorous testing and validation to prevent accidents and unintended consequences.

Every innovation and tech or AI system widen the inequality gap. Addressing inequality means ensuring that AI benefits are distributed fairly and do not widen existing social and economic gaps. AI must strive to bridge the digital divide and ensure equal accessibility.

Inter-relativity and Human Centered AI

Inter-relativity in the Human Centered AI Paradigm indicates that the positive elements are interconnected and must work together to ensure AI is human-centered while the negative elements should serve as a signpost to avoid pitfalls in AI systems. A human-centered and responsible AI approach considers the collective impact of humanity, human rights, advocacy, regulations, innovations, ethics, privacy, and security to advancing AI for social good that serves humanity.

Solutions for human centered AI must be multifaceted, addressing ethical, social, and technical dimensions simultaneously. Hence collaboration among stakeholders, including policymakers, technologists, ethicists, and advocacy groups, is essential to create balanced and inclusive AI systems. Integrating these components can establish AIs that are ethical, fair, safe, and beneficial to all of humanity, ensuring that technology serves to enhance human life and dignity.

Instances of Challenges to Human Centered AI

Creating Human Centered AI involves numerous challenges that can impact the fairness, reliability, and overall effectiveness of AI systems. Some of these key challenges include:

Bias and Fairness

- **Data Bias**
Training data often contains biases reflecting societal prejudices, leading to biased AI models.
- **Sampling Bias**
When data collection does not take into consideration diverse groups and persons it leads to incomplete unrepresentative data. Incomplete or unrepresentative samples can skew results, affecting the fairness and generalizability of AI models.
- **Measurement Bias**
Inaccurate measurements or proxy variables can introduce errors and unfairness into AI predictions.

Data Quality and Quantity

- **Data Scarcity:** Limited or imbalanced data sets can hinder model training and lead to unreliable predictions.
- **Data Quality:** Noisy, incomplete, or inconsistent data can degrade model performance and decision-making accuracy.

Complexity and Interpretability

- **Model Complexity:** Highly complex models, such as deep neural networks, can be difficult to interpret, making it challenging to ensure they align with human values and ethics.
- **Transparency:** Lack of transparency and explainability in AI decision-making processes can prevent stakeholders from understanding and trusting AI outcomes.

Uncertainty and Robustness

- **Uncertainty Quantification:** Accurately quantifying uncertainty in AI predictions is essential for reliable decision-making but remains a challenge.
- **Robustness:** Ensuring AI systems perform well under varying conditions and are resilient to adversarial attacks is critical.

Ethical Considerations

- **Value Alignment:** Aligning AI system outputs with human values and ethical principles requires sophisticated statistical methods to balance competing interests and objectives.
- **Trade-offs:** Balancing trade-offs between fairness, accuracy, and privacy can be challenging.

Dynamic and Evolving Data

- **Non-Stationary Data:** AI systems must adapt to changes in data distribution over time, which can be complex.
- **Continuous Learning:** Implementing continuous learning mechanisms without degrading performance or fairness is a significant challenge.

Causality and Correlation

- **Causal Inference:** Distinguishing between correlation and causation is essential for making reliable and ethical decisions but is often challenging.
- **Confounding Variables:** Identifying and controlling for confounding variables to avoid misleading conclusions requires advanced statistical techniques.

Privacy and Security

- **Data Privacy:** Ensuring data privacy while allowing AI systems to learn effectively from data poses statistical challenges, such as implementing differential privacy.

- **Security:** Developing statistically robust methods to detect and mitigate security threats, such as adversarial attacks, is essential.

Evaluation Metrics

- **Appropriate Metrics:** Selecting and balancing appropriate evaluation metrics to measure AI system performance can be challenging, especially when dealing with fairness and bias.
- **Multi-Objective Optimization:** Optimizing multiple conflicting objectives, such as accuracy and fairness, requires sophisticated statistical approaches.

Human-AI Interaction

- **Usability:** Designing AI systems that are user-friendly and intuitive while maintaining statistical rigor is a challenge.
- **Feedback Loops:** Incorporating human feedback into AI systems in a statistically sound manner is crucial for continuous improvement and alignment with human needs.

Addressing these statistical challenges is crucial to developing Human Centered AI that is fair, reliable, interpretable, and aligned with human values and ethics.

Conclusion

Human-centered AI is crucial because it ensures that AI systems focus on human needs and values. To incorporate human-centered design in AI means to involve users actively in the development process. This collaborative approach leads to more effective and ethical solutions as it harnesses diverse perspectives and expertise. For example, when teams involve users from various backgrounds, they can help identify and mitigate biases in AI algorithms, leading to more equitable outcomes. Moreover, human-centered AI fosters trust and acceptance among users. When people understand and see the value of AI systems, they are more likely to adopt and support these technologies. This trust is essential for the successful integration of AI into everyday life.

A handwritten signature in black ink, appearing to read 'Majiuzu', with several horizontal lines drawn through it.

Majiuzu Daniel Moses Executive

Director

Africa Tech for Development Initiative - Africa4dev

info@africa4dev.org; mmajiuzu@gmail.com www.africa4dev.org

+2348063154897