NAVIGATING ARTIFICIAL INTELLIGENCE FROM A HUMAN RIGHTS LENS:

Impacts, Tradeoffs and Regulations for Groups in Vulnerable Situations

By Adriana M. Ramírez Sánchez,
Ishita Bhatia,
Samia Firmino Pinto

Geneva, July 2023
Human Rights and Humanitarianism

Navigating Artificial Intelligence from a Human Rights Lens: Impacts, Tradeoffs and Regulations for Groups in Vulnerable Situations

Geneva Graduate Institute of International and Development Studies
Applied Research Project ARP_02_07
Human Rights and Humanitarianism track

Research Team:
Adriana M. Ramírez Sánchez
Ishita Bhatia
Samia Firmino Pinto

Academic Supervisor: Buğra Güngör
Tutor: Kudzai Tamuka Moyo

Partner organization:
Geneva Academy of International Humanitarian Law and Human Rights
Bernard Duhaime, Geneva Academy Associate Research Fellow
Erica Harper, Geneva Academy Head of Research and Policy Studies

Final report
July 2023
Geneva, Switzerland

Executive Summary

The report examines the dual role of Human Rights in assessing potential impacts of Artificial intelligence (AI) and setting a benchmark for regulatory practice while shedding light on the obstacles encountered in this process. Taking into account the existing literature and research gap, this report deliberately adopted a particular focus on the following groups in vulnerable situations: women and children, persons with disabilities, and racial and ethnic minorities, thereby revising the tradeoffs, potentials, and shortcomings of existing human rights instruments in regulating the adverse effects faced by these groups. Additionally, the report analyzes how current regulations, especially the recent EU AI Act, address AI's impacts and incorporate a human rights perspective into the issue and examines the adaptive role to be played by human rights to address current and future AI impacts.

Main findings

1. AI's rapid evolution, marketability, and potential for biased behavior challenge human rights impact and regulation potential, while balancing the need for data to improve accuracy poses privacy concerns embedded in the system.

2. AI's impacts on human rights occur throughout the AI lifecycle arising from technical and societal issues, with complex state responsibilities to protect, fulfill, and respect human rights.

3. Incidents can be driven by or related to AI systems, potentially resulting in human rights abuses and violations, particularly concerning vulnerable populations. Distinguishing such incidents requires enhancing expertise in human rights within the technology sector.

4. Potential human rights abuses and violations associated with AI technologies regarding women and children, persons with disabilities, racial and ethnic minorities, and other groups in vulnerable situations have been increasingly found in everyday spaces (both physical and virtual) such as schools, interactions with public services and search engines.

5. Even though there is no regulatory vacuum, human rights should be more included in the regulatory practice. For that, efforts should be made in recalibrating its boundaries, highlighting the unique role that rights and principles such as best interest of the child, evolving capacities, right to an independent life, reasonable accommodations and inclusive equality entail, as well as promoting Business responsibility mechanisms.

Recommendations

1. **Enhance Research on Human Rights Impact:** There is a need for increased academic and policy research focused on understanding the human rights impacts of AI, particularly in under-researched populations such as children and persons with disabilities, among others. This research should strive to establish a benchmark that correlates AI incidents with human rights impacts and should actively engage with the voices and insights of the affected groups. By strengthening research efforts, the potential human rights challenges arising from AI technologies can be effectively addressed.

2. **Promote Human Rights Awareness and Responsibility:** All stakeholders, including governments, AI providers, developers, and civil society should keep into account human rights while designing AI systems. This entails raising awareness about the potential adverse impacts on the rights of women and children, persons with disabilities, and racial and ethnic minorities. Actors must recognize their responsibility to protect and respect the rights of these individuals by adopting comprehensive and inclusive approaches.

3. **Ensure Rights-Respecting AI Systems:** AI systems must be designed and deployed in a manner that upholds human rights principles. Balancing innovation and technological advancement with the protection of groups in vulnerable situations is essential. Responsible regulations should be implemented to prevent the adverse effects and abuses facilitated by AI systems, ensuring the long-term trust and well-being of individuals through process-oriented due diligence, human rights impact assessments, and access to remedies.

4. **Establish Effective Human Rights Oversight and Regulation**: Governments should create institutional mechanisms that effectively plan, direct, promote, and regulate AI technologies in alignment with human rights standards. These mechanisms should prioritize the protection of rights, upskilling professionals, and implementing policies that accelerate the development of AI regulations. Such regulations should incentivize AI research that serves the broader societal interest while safeguarding the rights of the groups in vulnerable situations.

5. **Foster Human-Centric AI Design:** AI research should prioritize the development of techniques and practices that prioritize human rights and dignity. Designing AI systems with a human-centric approach ensures that they meet the specific needs and capacities of women and children, persons with disabilities, and racial and ethnic minorities.

Table of content

Table of figures

Figures

Tables

Text Boxes

Acronyms

AI - Artificial Intelligence

AIAAIC - AI, algorithmic, and automation incidents and controversies Repository

B&HR- Business and Human Rights

B-Tech- Business and Human Rights in Technology Project

CEDAW - Convention on the Elimination of All Forms of Discrimination against Women

CERD - Convention on the Elimination of All Forms of Racial Discrimination

EU AI Act - European Union Artificial Intelligence Act

ICCPR - International Covenant on Civil and Political Rights

ICERD - International Convention on the Elimination of All Forms of Racial Discrimination

ICESCR - International Covenant on Economic, Social and Cultural Rights

ILO - International Labour Organization

OHCHR - Office of the High Commissioner for Human Rights

UNCRC - United Nations Convention on the Rights of Children

UNCRPD - United Nations Convention on the Rights of Persons with Disabilities

UNGPs - United Nations Guiding Principles on Business and Human Rights

UNHRC - United Nations Human Rights Council

Glossary[1]

AI actors: those who play an active role in the AI system lifecycle, such as technology developers, service and data providers, public or private organizations or individuals that acquire AI systems to deploy or operate them.

AI system: systems that have the capacity to process usually a large amount of data and information in a way that resembles intelligent behavior, and typically includes aspects of reasoning, learning, perception, prediction, planning or control. Systems, platforms and technologies based on algorithmic processes.

Algorithm: a computational procedure that takes an input and produces an output. They are a fundamental building block of computer science and are used to solve problems from simple arithmetic calculations to complex machine learning models.

Algorithmic decision-making: refers to the process in which algorithms are used to make decisions or take actions. For example, self-driving cars may use an algorithmic decision-making system to determine the best course of action based on the surrounding environment, traffic conditions, etc.

Black box: a characteristic of a system that doesn't provide any transparency or understanding of how it operates in a manner sufficient to understand how specific inputs result in specific outputs.

Deep fake: the use of algorithms and other deep learning techniques to generate content used to trick or fake the viewer. It happens, for example, when manipulating images and videos to make it seem that a person is in the image or video when they are not.

Deep Learning: a subfield of machine learning that uses algorithms inspired by the structure and function of the human brain, known as artificial neural networks, to process data, model and solve complex problems. It has largely been used in image and speech recognition, natural language processing, and game playing.

Generative AI: a subfield of AI that involves training models to learn the patterns and structures of the data, so that they can create new data that is similar to the original data but not identical. Generative AI has been largely employed, for example, by platforms specialized in producing new images, videos, or music..

[1] This Glossary gathered definitions from different sources to convey understandings aligned to the concepts explored in the research, such as the reference textbook "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig; Cognilytica's AI Glossary; UNESCO Ethics of AI report and AccessNow report on AI & Human Rights, and OHCHR definitions among others.

Groups in vulnerable situations: the term refers to segments of the population that are more susceptible to experiencing harm, discrimination, or disadvantage due to factors such as age, gender, ethnicity, disability, social, economic, or physical circumstances. They may face increased risks, have limited access to resources or opportunities, and require specific assistance and representation to ensure their well-being and protection from harm. They may include children, women, persons with disabilities, refugees and displaced persons, racial minorities, etc. Additional explanation is provided in the report.

Information filtering: refers to the process of selecting and presenting information to users based on their characteristics, behavior, or preferences. Through this practice, algorithms analyze users' online activities, such as contact interactions and websites visited to tailor what is displayed with the aim of increasing engagement and satisfaction. The ultimate goal is to filter information to provide personalized content.

Machine Learning: is a subfield of AI that involves developing algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. It involves creating mathematical and statistical models based on input data, which can be used to make predictions or detect patterns.

Natural Language Processing (NLP): a subfield of AI that involves developing algorithms and models that can understand and generate human language. The techniques used in NLP analyze and generate text and speech data, enabling applications such as machine translation, sentiment analysis, and chatbots.

Neural data: refers to data that is generated by measuring the electrical or chemical activity of neurons in the brain and brain activity. It is considered personal information about the neural states, processes, and structures of one's neural activity.

Predictive Modeling: is a technique used in machine learning and data science to build models that can make predictions about future events or outcomes based on past data.

Profiling: involves creating detailed profiles of individual users based on their characteristics, behavior, and preferences. It gathers and analyzes data about users to understand their demographics, habits, and other personal attributes. These profiles are then used to tailor content, advertisements, or recommendations to each individual based on their profile. Profiling can lead, for example, to the manipulative targeting of children by advertisers and platforms, exploiting their vulnerabilities and influencing their behavior and choices.

Introduction

AI has emerged as a transformative technology, profoundly shaping various aspects of our society and significantly influencing our lives, work, and interactions. As the new technological landscape is being navigated, it is crucial to examine the potential impacts of AI on human rights, with a specific emphasis on groups in vulnerable situations, including women and children, persons with disabilities, and racial and ethnic minorities. This report aims to provide findings and analysis that shed light on the implications of AI on human rights and explore how existing human rights norms and frameworks can effectively regulate its impact. The research questions guiding this report are:

Research questions

1) What are the potential impacts of artificial intelligence (AI) on human rights, especially those of vulnerable groups?
2) What are the tradeoffs, potentials and shortcomings of human rights mechanisms and instruments to regulate such impacts?

To address these questions, we will delve into various aspects related to AI and human rights. We will begin by embarking on the journey of understanding AI, exploring its complexities, and the challenges it presents. This section will include an introduction to the "AI Abyss" to provide a foundational understanding of AI's nature and its diverse applications. We will also examine the challenges that arise from this rapidly evolving technology.

Next, we will unmask the impacts of AI on human rights, aiming to provide a comprehensive snapshot of the tendencies of AI incidents and their potential implications. By analyzing AI incidents and connecting them to human rights impacts, we will shed light on the specific vulnerabilities faced by different groups. Furthermore, we will assess the existing human rights norms and frameworks and their effectiveness in addressing these challenges, emphasizing the priority at risk.

By exploring the tradeoffs, potentials, and shortcomings of existing human rights instruments and other regulatory efforts, we seek to understand how these norms and frameworks can effectively regulate the impact of AI on human rights while testing human rights capacity to serve as a benchmark for AI development, deployment and use. This section will delve into the recalibration of existing human

rights norms and frameworks, reinterpreting boundaries and interconnections to better address the challenges posed by AI as well as to highlight the need for new subjects of obligation to ensure comprehensive protection. Lastly, we addressed the challenges related to the participation and inclusion of groups in vulnerable situations in the regulatory processes.

Throughout this report, we recognize the nuanced nature of defining AI, considering its evolving characteristics and diverse applications. We also acknowledge the complexities surrounding the universality of human rights, taking into account differing understandings and contestations. By engaging with these complexities, we aim to provide an inclusive and sensitive examination of the impact of AI on human rights in various contexts.

Text box 1. Some relevant definitions.

Defining the concepts related to **vulnerability** and **vulnerable situations** can be challenging and contested. The vulnerability can be examined from different perspectives in a variety of disciplines, being regarded as dynamic, processual, and relative, and occurring as the result of social injustice, discrimination, or inequality. It can comprise the groups who are innately vulnerable (such as physically disabled individuals) and groups who become vulnerable due to circumstances and structural factors.

The term adopted in this report, **groups in vulnerable situations**, refers to segments of the population that are more susceptible to experiencing harm, discrimination, or disadvantage and whose needs are often neglected due to factors such as age, gender, ethnicity, disability, poverty, social and economic circumstances. They may face increased risks, have limited access to resources, and require specific assistance, support, and representation to ensure their well-being and protection from harm.

According to the Office of the United Nations High Commissioner for Human Rights (OHCHR), groups in vulnerable situations may include children and adolescents, women and girls, persons with disabilities, refugees and asylum-seekers, LGBTQ persons, and elderly individuals. Aiming at narrowing down the research scope, the groups analyzed in this report comprise **women and children, persons with disabilities, and racial and ethnic minorities**.

Operationally, **human rights** are internationally recognized principles and standards that safeguard the inherent dignity, freedom, and equality of all individuals. These rights encompass civil, political, economic, social, and cultural dimensions, and they serve to protect and promote the well-being and autonomy of individuals and communities. Mostly developed through legal regimes and language, the protection of this specific set of individual and collective rights is the primary responsibility of the State (Nikken 2006).

Methodology

Research Design:
This research employed a **mixed-methods approach** to investigate the role of human rights in assessing impacts on vulnerable populations and serving as regulatory mechanisms in the context of AI. This comprehensive approach (Doyle, Brady, and Byrne 2009) allowed us to examine trends and patterns of AI incidents, as well as underscore the specificities of human rights impacts and regulatory potentials. It aligns with existing literature that emphasizes the need for multidimensional approaches to address the complex challenges of AI and human rights (Martsenko 2022; Raso et al. 2018).

The research design encompassed **quantitative analysis of the AIAAIC Repository** (AI, Algorithmic, and Automation Incidents and Controversies Repository), –an independent, open, public interest resource and a comprehensive database that documents instances of incidents related to AI and its impacts. Additionally, qualitative analysis was conducted on a systematic review of relevant literature, current human rights instruments, AI regulations, and interviews.

Data Collection and Management:
For the quantitative analysis, primary data was collected from the AIAAIC Repository. As of June 2023, the repository contained 1,005 recorded AI related incidents, providing a rich dataset for analysis of the general trends. Furthermore, a specific subset of 205 incidents that were relevant to the unequal distribution of impacts on groups in vulnerable situations were identified for further analysis. The dataset underwent preprocessing to ensure consistency, accuracy, and relevance.

The qualitative analysis involved the examination of existing literature, regulatory frameworks and international human rights instruments. Semi-structured interviews (see Annex 4) with industry professionals, representatives from the UN B-Tech project –a United Nations project that provides authoritative guidance and resources for implementing the United Nations Guiding Principles on Business and Human rights (UNGPs) in the technology space– , and academics was also conducted, as to have multi-stakeholder representation. A semi-structured interview guide was developed to ensure consistency and cover essential topics, including the impact of AI on human rights, perspectives on current regulations, challenges in implementing regulations, and potential areas for improvement while allowing for an open-ended approach to gather participants' perspectives.

According to Rubin and Rubin (2012), a well-designed interview guide helps the researcher maintain focus and ensures that important areas of inquiry are addressed. Schensul, Schensul, and LeCompte (1999) assert that semi-structured interviews are valuable in capturing rich and nuanced data that enhance the understanding of complex issues.

The choice of using a **relevant purposive sampling** in the research methodology, both for the interviews and the analysis of regulations, is justified for several reasons. The sampling strategy **aligns with the research question and the specific expertise required for the study**. By purposefully selecting participants with knowledge and experience in AI, human rights, and regulation, the likelihood of obtaining valuable insights directly related to the research objectives is increased (Patton 2015). This diverse range of expertise allows for gathering insights from experts of different arenas and nationalities, contributing to a comprehensive understanding of the research topic. In this sense, Kvale (1996) underscores the value of interviews with knowledgeable individuals for obtaining in-depth information and interpretations of complex phenomena. Furthermore, by selecting frameworks specifically addressing human rights and their intersection with AI, the analysis focuses on the **most pertinent documents.** Creswell (2014) emphasizes the importance of purposeful sampling that aligns with research goals.

Analysis of Data:
Quantitative data analysis involved descriptive statistics and pattern identification techniques applied to the AIAAIC Repository data. Further qualitative correlation analysis was performed to identify relationships between AI incidents and human rights concerns in a handful of cases. The qualitative analysis focused on content analysis (Neuman and Neuman 2014) to the interviews and regulatory instruments assessing their potentials, shortcomings and tradeoffs in addressing human rights impacts. Integration and analysis involved synthesizing findings from quantitative and qualitative analysis and deriving implications and recommendations for enhancing AI regulation and human rights engagement.

Limitations:
First, regarding the quantitative data and the AIAAIC Impact Repository some limitations should be acknowledged –limitations that were discussed with the founder of the repository during an interview–. The collection mechanism of the repository may introduce biases towards incidents that receive substantial media coverage. The repository's manual detection of incidents is subject to human judgment and potential oversight or subjective biases or preferences. The categorization and taxonomy used in the repository may contain inconsistencies or

gaps. Additionally, the repository's Western-centric nature may result in limited coverage of AI incidents and their impact in non-Western regions.

To mitigate the limitations of the AIAAIC repository, **efforts were made to balance and maintain objectivity in the quantitative analysis, especially regarding the generalization of the data.** Despite the potential biases and gaps, the AIAAIC Repository provides the first and most comprehensive effort to collect and systematize AI impacts in a way that represents the overall conscientious risks AI poses. That is why the repository provides a valuable starting point for understanding AI impact incidents and further work on their implications for human rights.

Second, it would have been beneficial to incorporate into our purposive sampling techniques a **wider range of regional and global human rights instruments and their interaction with AI regulation.** This approach would have allowed for a diverse representation of approaches and perspectives. Additionally, the importance of including the **views of individuals with disabilities, children and other groups in vulnerable situations is acknowledged by us throughout this report.** Unfortunately, due to time constraints, interviews with these groups were not possible..

Ethical Considerations:
Throughout the research process, **IHEID ethical guidelines were followed** to ensure participant confidentiality and data privacy. Necessary approvals were obtained, and informed consent was obtained from participants during interviews.

1. Embarking on the Journey

In this section, the vast and complex world of Artificial Intelligence (AI) is introduced as the first step in the journey of our research, "Navigating Artificial Intelligence from a Human Rights Lens: Impacts, Tradeoffs, and Regulations for Groups in Vulnerable Situations".

1.1. Introduction to the AI Abyss

Numerous interpretations and perspectives about AI have emerged over the past decades. There is **no single, universally accepted definition because the term is derived from a series of techniques developed to support, augment and automate human activities,** and is usually related to tasks inherent to human intelligence. In addition, the definition of AI has been subjected to debates and deliberations primarily due to its dynamic nature that evolves as technology advances.

Text Box 2. Origin of the term AI

The term "Artificial Intelligence" was coined by John McCarthy, an American computer scientist, and cognitive scientist, in 1955. McCarthy described AI as **"the science and engineering of creating intelligent machines".** Since 1955, definitions of AI, much like its use, development and deployment has gone through myriad of milestones and versions.

Source: Computer History Museum 2021

Some efforts have been made from different sectors to come up with a definition of the term, as can be seen in the following text box.

Private Sector Developers Definitions:

Deloitte in 2022: "Automation of human cognitive functions."

Gartner: "Utilization of advanced analysis and logic-based techniques to interpret events, support decision-making, and enable automated actions."

Google: "A field of science for building machines that can reason, learn, and act in such a way that would normally require human intelligence or that involves data whose scale exceeds what humans can analyze."

National Perspectives:

As per the New Generation of Artificial Intelligence Development Plan of 2020, China's government perceives AI as: "The technology that will allow for advancing 'intelligentization' as the center of humanity's sustainable development."

As per National Artificial Intelligence Initiative Act of 2020, USA perceives AI as: "Machine-based systems that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."

"Ethics on AI" Report By UNESCO Views:

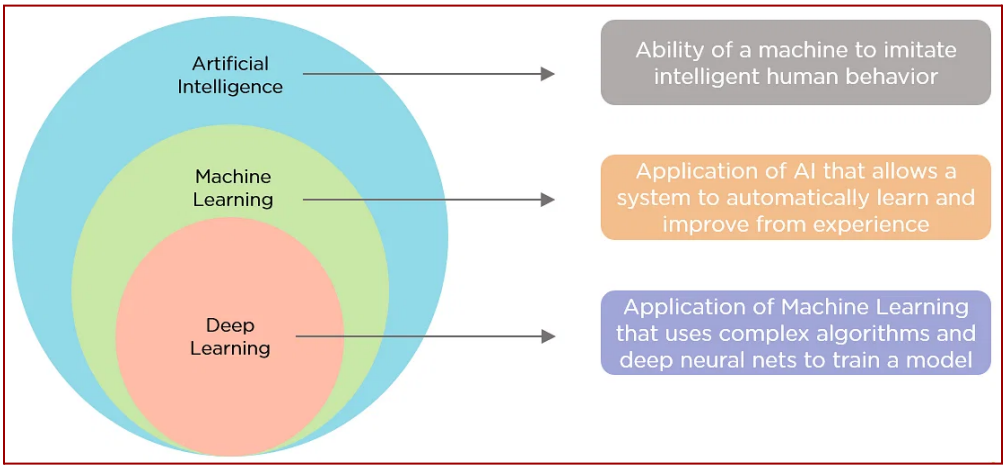
UNESCO in 2023 provided no specific definition as they believe “a definition would need to change over time, in accordance with technological developments."

European Commission's Communication on AI Definitions, in 2018, proposed:

"Artificial Intelligence refers to systems that display intelligent behavior by analysing their environment and taking action — with some degree of autonomy — to achieve specific goals."

AI encompasses diverse domains and elements, as shown in **Fig 1**, each with unique applications and techniques. A significant subset of AI is **Machine Learning (ML) which refers to the ability of computer programs to learn from and adapt to new data without being explicitly programmed by humans** (Burns 2021). Machines can now interpret, predict, analyse, and perform various functions with AI, allowing systems to learn from their experiences, adapt to new inputs, and perform tasks that human intelligence was once unable to accomplish (Burns, Laskowski, and Tucci 2023). **Deep learning is a further subset of machine learning.** It uses technines known as neural networks, which consists of multiple layers of algorithms emulating the human brain functioning, to enable machines to learn tasks such as image recognition and natural language processing with astonishing accuracy (Grieve 2023). It is worth noting that although some AI systems can function ‘autonomously’, human intervention is still to various degrees in the loop –at least in the learning and some aspects of design, development, deployment, and usage.

Fig 1. AI vs Machine Learning vs Deep Learning Figure (M Shruti 2023)



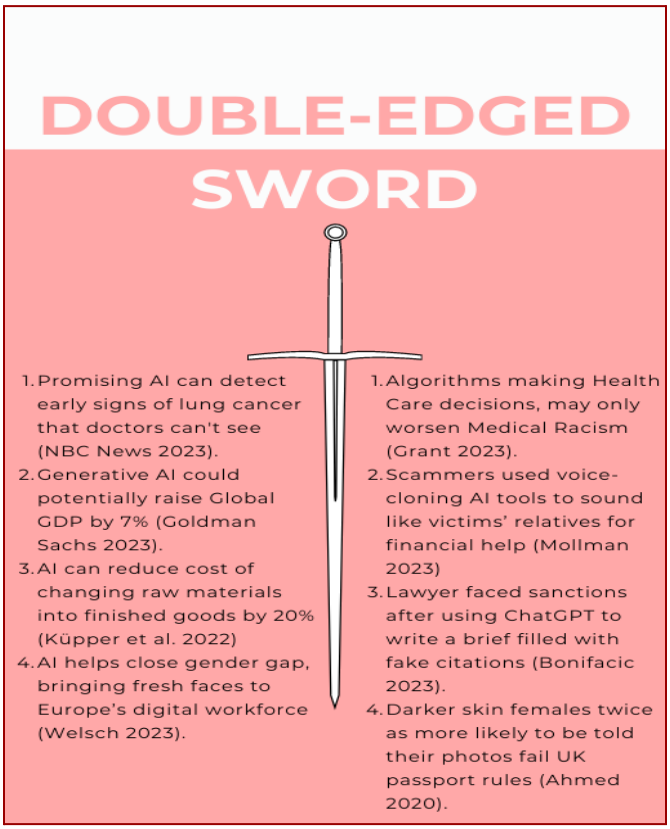
With the recent release of ChatGPT (Chat Generative Pre-Trained Transformer), an AI chatbot developed by OpenAI and launched in November 2022, generative AI entered the mainstream, adding to the ongoing debates and discussions. In the minds of most consumers, ChatGPT became synonymous with artificial intelligence.

However, ChaptGPT represents merely a minuscule part of the whole AI abyss.

Amidst the multitude of definitions and interpretations, our report contends that AI can be best viewed as a **ubiquitous double-edged sword.** The sword is **ubiquitous** because it is deeply embedded in many aspects of our daily lives, often working behind the scenes. Even though significant corporations are typically at the forefront of AI research and development, AI technology is ever present in our daily lives, sometimes without realizing it (Ibm 2017). Its pervasive presence is reshaping ways of working, modes of communication, and most importantly the essence of our lives.

The sword is also **double-edged** at the same time because it is capable of achieving the unimaginable by uncovering deadly diseases and inventing miraculous cures, improving efficiency, productivity, and accuracy across various sectors, including healthcare, finance, transportation, and education. These advancements have the potential to enhance human lives, solve complex problems, and drive economic growth. Meanwhile, the destructive edge of the sword also monitors, manipulates, and controls the lives of many without their knowledge or consent, leading to job displacement, dependency and other potential risks. **It is not only the technology in sight but the way human agency develops and deploys these systems.**

Fig 2. The Good and Dark Sides of AI



1.2. Challenges arising from the AI Abyss

The world of AI is a seemingly never-ending abyss with myriad challenges. The following section aims at highlighting some of the challenges while touching upon the importance of acknowledging the dark side of AI systems and its potential impact on Human Rights of individuals.

Challenge 1. AI is an ever-evolving, innovation-led, and highly marketable field

The AI abyss can be seen to be evolving between “initial adoption” and “widespread use” and is thus undergoing a dynamic process of evolution (Anadon et al. 2015). This means that factors including marketing strategies, price changes, and behavioral or cultural tendencies, significantly influence the adoption of AI. As a result, identifying new risks and predicting AI's impacts become increasingly challenging.

AI technologies are advancing rapidly, with new algorithms, models, and techniques being developed continuously. Moreover, each serves a different purpose and fulfills a different function (See **Table 1**). This fast-paced innovation poses a significant challenge for policymakers, regulatory bodies, and legal frameworks to adequately keep up with the evolving landscape and address potential risks and impacts.

The marketability of AI often compels companies and developers to prioritize swift deployment and competitive advantage over ethical considerations resulting in inadequate attention to bias, privacy, transparency, and fairness issues. This can lead to potential risks and negative impacts on human rights, particularly for vulnerable populations. As noted by industry expert León Palafox, "**as long as there are no regulations in place, companies will make the most out of AI profit potential".**

The race to gain market dominance sometimes leads to inflated claims and unrealistic AI technology expectations. This hype can overshadow the need for responsible innovation, rigorous testing, and careful consideration of potential risks. It is crucial to encourage transparency, honesty, and evidence-based assessments of AI capabilities to avoid misleading narratives that may negatively affect vulnerable populations.

Table 1. AI systems categorization based on their function

Classification systems	These AI systems are designed to classify data into predefined categories or classes based on specific features or patterns. Classification systems are commonly used in applications such as image recognition, spam filtering, and sentiment analysis.
Regression and predictive systems	Regression AI systems are used to predict numerical values or continuous outcomes based on input data. They establish relationships between input variables and output values to make predictions.
Recommendation systems	These AI systems provide personalized recommendations or suggestions to users based on their preferences, behavior, or historical data. Recommendation systems are widely used in e-commerce platforms, content streaming services, and personalized marketing campaigns.
Clustering Systems	Clustering AI systems group data points together based on their similarities or patterns, aiming to identify inherent structures or relationships within the data. Clustering is useful in customer segmentation, anomaly detection, and data exploration.
Natural Language Processing (NLP) Systems:	NLP systems enable computers to understand, interpret, and generate human language. They encompass tasks such as language translation, sentiment analysis, chatbots, and text summarization.
Computer Vision Systems	Computer vision AI systems focus on analyzing and understanding visual information from images or videos. They can perform tasks like object detection, image recognition, facial recognition, and scene understanding.
Generative Systems	Generative AI systems have the ability to generate new content, such as text, images, or audio, that is original and not directly derived from existing examples. These systems can be used for creative applications, content generation, and simulations. It's worth noting that AI systems can often combine multiple functions and techniques to address complex tasks and solve real-world problems.

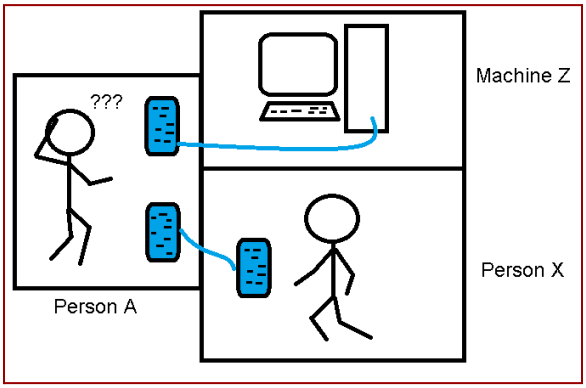
Research team (Ramírez, Firmino, Bhatia) Source: Open AI

The market-driven nature of AI can exacerbate existing inequalities and create new disparities. AI technologies may not be equally accessible or affordable for all individuals and communities, perpetuating digital divides. This can disproportionately affect vulnerable populations, limiting their ability to benefit from AI advancements.

Challenge 2. AI systems can exhibit and multiply human-like behavior, and can also exhibit human-like biases, discrimination, and stereotypes with several degrees of autonomy

Some have said that AI has a manipulative and controlling potential through Turing's test. A Turing test evaluates a machine's ability to exhibit human-like behavior by answering questions in a way that cannot be distinguished from a human's response **Fig 3** (Shridhar 2018).

Fig 3. An illustration of the Turing Test (Shridhar 2018)



Alongside its ability to emulate human-like behavior, AI can also exhibit human-like biases, discriminations, and stereotypes. Combining all these facets, AI can potentially cause significant harm to individuals, including severe violations and abuse of their fundamental rights. Despite their intended neutrality, AI systems tend to reflect and perpetuate deeply ingrained societal prejudices, especially against vulnerable and marginalized groups.

> "AI systems are only as unbiased as the data they are trained on, and if the training data reflects societal biases, the algorithms may perpetuate discrimination or exclusion"- **León Palafox, industry expert.**

Identifying and addressing these intrinsic biases within AI systems becomes crucial, particularly when addressing the potential impact on vulnerable and marginalized groups, such as women, racial and ethnic minorities, and persons with disabilities, to avoid perpetuating existing social inequalities through AI technologies. The challenge, however, lies in designing AI systems that are sensitive towards and inclusive of diverse perspectives without amplifying existing societal biases. Technicians argue that, when properly used, algorithmic decision-making systems can lead to more objective and potentially fairer decisions than human decisions. However, achieving this requires a thorough understanding and recognition of these systemic exclusionary mechanisms (Whittaker et al. 2019).

Challenge 3. For better training, optimization, and accuracy, more data is needed, which further challenges data privacy

One of the primary tradeoffs lies in the tension (and delicate balance) between privacy and effective and accurate AI systems. AI-driven technologies often rely on vast amounts of data, including personal information, for training and optimization.

However, this reliance raises concerns about the infringement of privacy, particularly for groups in vulnerable situations who may as well already be subjected to surveillance and discrimination (Martínez Ramil 2021). Striking a balance between utilizing data for AI advancements and protecting the privacy rights of individuals, particularly children, is crucial. In this sense, as León Palafox, AI industry professional, mentioned in his interview more data can increase the accuracy of the system according to the function or purpose. However, **the regulatory landscape should be mindful of what kind of information is feeding a system to alleviate the potential infringement of the right to privacy and harmful bias.**

This problem further exacerbates with generative AI, as an UN B-Tech project personnel, mentioned in the interview. According to her, this type of AI system usually trains and deep learns with whatever information is available, which might include private information. It can be clearly seen that data privacy and nondiscrimination are embedded challenges in the creation and development of AI systems.

Challenge 4. Impacts of AI may emerge in different contexts or from different sources, throughout the AI system lifecycle

AI systems can have profound impacts on human rights at different stages of the AI system lifecycle, from the data collection phase, through system planning and design, all the way to the deployment and use phase. As developed by Raso, **human rights impact of AI may emerge in different contexts or from different sources** (Raso et al. 2018).

Text Box 3. AI system life cycle

The phases an AI system can undergo through its "existence". They are grouped differently by some authors, ranging from **conception, design, development, deployment and use** of the system to its **retirement, or even including maintenance, monitoring and evaluation, end-of-use, and termination**.

Despite the differences, they tend to cover **data collection and processing as well as model development, training and deployment**.

Source: OECD, Unesco, et al.

The first context pertains to the **data** used for training a system. If the data used reflects a situation in which social bias is present, the AI system will also produce biased outputs. Thus, the decision-making process that is adopting an AI system will reflect those biases impacting the context.

The second is related to the **system design**. The prevalence of diverse human choices, preferences, backgrounds, and possible lack of diversity in the developer team can significantly shape the system design, either positively or negatively and consequently, the real context in which the system will be used. It is also crucial to recognize that not all subtypes of AI will work in all settings. Some might have

certain shortcomings that will prove to be harmful in certain areas of application and thus for certain groups.

The third context is related to **complex interactions** that emerge in the environment where the AI system is deployed. Those interactions can be related to the way AI system outputs will influence the decision-making and policy-making processes. For example, through the information produced by AI systems that is used to guide and define predictive policing actions. The analysis of the interactions that emerge from the context in which AI systems are used is essential. The impacts of AI on human rights are multidimensional in nature.

Table 2. How AI Impacts Human Rights throughout AI Systems Life cycle

Research	**lack of direction in AI research and AI purposes may be detrimental to human rights** (e.g. prioritizing research and investment on large language models and AI image generation rather than research on global health and zero hunger solutions)
Design	**data used for training, developers' choices and biases, and companies' interests may all affect human rights** (e.g. a vulnerable group is mis/underrepresented in a training dataset resulting in discrimination, or developers' choices in labeling for ML training)
Development	**choices concerning AI solutions, products and services to be developed and commercialized may impact human rights** (e.g. companies' choices determining the type of AI products/services they will developed and put into market; government choices when implementing algorithmic decision systems for public services)
Deployment/Use	**the way AI systems are utilized in a certain environment may impact human rights** (e.g. societal impacts like discrimination, unintended outcomes such as polarization in elective democratic processes, malicious uses via misinformation and image generation for child's sexual exploitation)

Research team (Ramírez, Firmino, Bhatia)

As can be seen, the impacts of AI on human rights are seen at different stages. Starting from the research phase, where substantial investments may prioritize profit-driven discoveries over addressing human needs, to the complex interactions between the AI system and its environment, when a system can be misused for purposes that go beyond its original intent, there is potential for AI systems to affect human rights. In the subsequent sections, these challenges will be seen to have an intricate link with impacts on human rights, thus introducing a unique yet complex scenario that should be taken into consideration.

2. Unmasking the Impacts of AI on Human Rights

In a speech titled **"We need to act now and put human rights at the center of artificial intelligence designs"** by Dunja Mijatović, the Commissioner for Human Rights of the Council of Europe, she elucidated how AI can negatively affect a wide range of human rights, from freedom of expression, privacy, association, and assembly.

In this section, these impacts will be unveiled, beginning with a general snapshot of some of the current tendencies and defining when an AI impact refers to a human rights impact, violation, or abuse. Further, light will be shedded on the existing human rights mechanisms and assessing a few instances where the human rights of vulnerable populations were at stake.

2.1. A General Snapshot: Current tendencies of AI incidents

AI incidents, as defined in the AIAAIC repository, are negative events or situations that are either directly triggered by an AI, algorithmic, or automation system, or where the technology and/or its governance is a significant contributing factor. These incidents can be deliberate or accidental and may result from internal factors like algorithmic opacity, misleading marketing, or ethical issues, as well as external factors including interaction with a complex societal issue, potentially causing harm to individuals, societies and/or the environment.

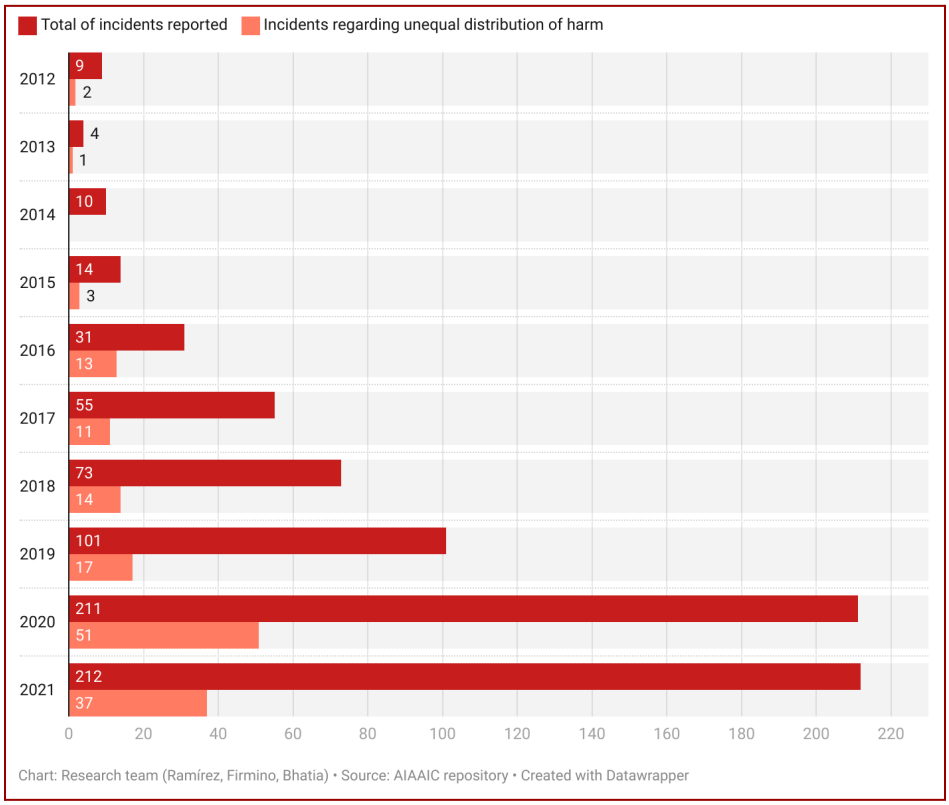
Finding 1

From analysing the data, it is evident that a **significant increase in the number of AI-related incidents reported in the last 10 years (Fig 4)**. This coincides with the rapid development of new AI systems or how they have become ubiquitous in our daily life activities shaped by digital technologies (as characterized by **Challenge 1** arising from AI).

When considering incidents that had direct unequal distribution of harm among groups in vulnerable situations including gender, race or ethnicity, age, and persons with disabilities, **there is an increase seen in reports that amount to 10 - 25% of the total incidents reported.** The main problem reflected in the analyzed data is the presence of bias within the outputs of AI technologies. This bias can perpetuate existing inequalities and reinforce discriminatory practices, exacerbating disparities

among groups in vulnerable situations. Also, the intersectionality of race, ethnicity, and gender, amongst others, emerged as the main factor contributing to the harm experienced by these groups. **This serves as a reminder that groups are not unidimensional characteristics, but they rather amplify and compound their vulnerabilities.**

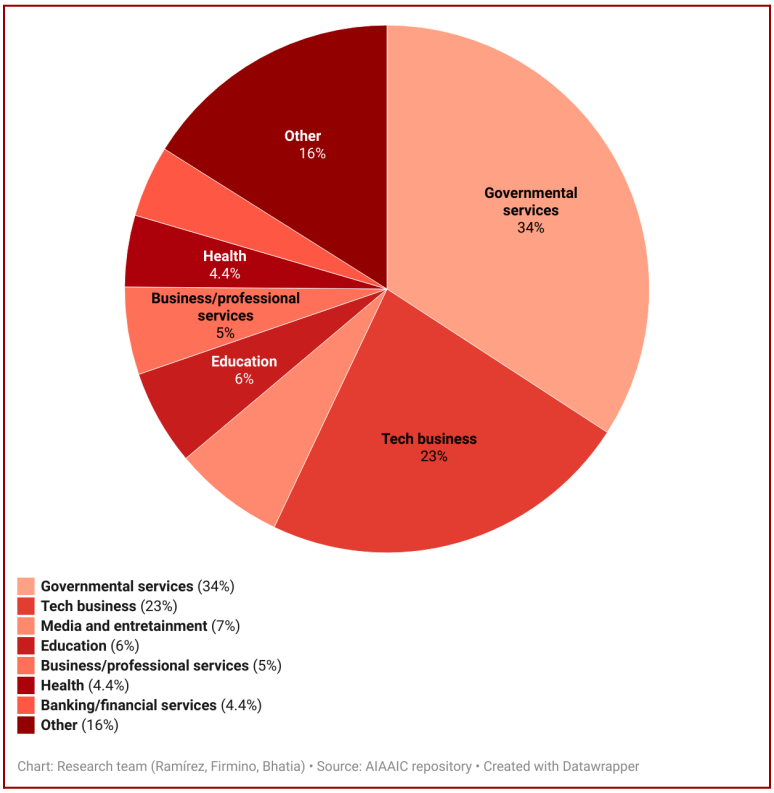
Fig 4. AI-related incidents reported throughout the years



Finding 2

The sectors most prominently represented in incidents where unequal distribution of harm occurred are: government services, technology, media/entertainment, education, business services, health, banking, and financial services. In this case, other sectors involved include transportation, retail and consumer goods **(Fig 5)**. **Government services (which coincidentally are the main duty-bearers of human rights) are still the most representative sector** of the AI incidents reported. However, the Business sectors should not be neglected here, especially in regulatory terms, as they still represent **more than 60% of the incidents reported.**

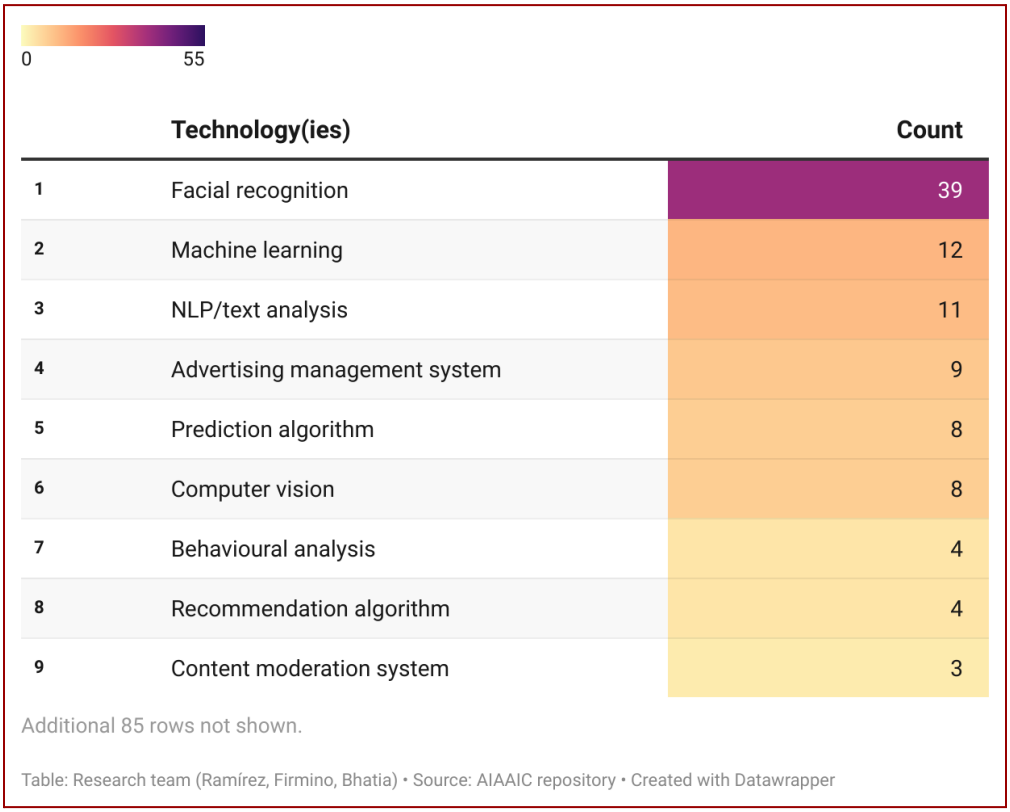
Fig 5. Sectors represented in the AI incidents for unequal distribution of harm



Finding 3

According to our analysis, technologies represented in incidents for unequal distribution of harm included: **Facial Recognition (39), Machine Learning (12), Natural Language Processing (NLP)/Text Analysis (11),** advertising management, and prediction algorithms, among others represented **(Fig 6)**. Although these technologies are widely utilized in various domains and industries, they are found to exhibit biases and inaccuracy in their outcomes, resulting in adverse impacts on groups in vulnerable situations.

Fig 6. AI technologies represented in incidents for unequal distribution of harm



	Technology(ies)	Count
1	Facial recognition	39
2	Machine learning	12
3	NLP/text analysis	11
4	Advertising management system	9
5	Prediction algorithm	8
6	Computer vision	8
7	Behavioural analysis	4
8	Recommendation algorithm	4
9	Content moderation system	3

Additional 85 rows not shown.

Table: Research team (Ramírez, Firmino, Bhatia) • Source: AIAAIC repository • Created with Datawrapper

It is important to note that the identified incidents primarily reflect direct harm caused by AI technologies. However, it should be acknowledged that groups in vulnerable situations can also be affected as non-users, experiencing indirect consequences and systemic biases that pervade AI systems.

"Often, people who are impacted are actually non-users, not active participants on a platform. Despite not being users, the outcomes still has a significant impact on their livelihoods."

- UN B-Tech project personnel

In this sense, AI algorithms and models are often trained on vast datasets that reflect the biases present in society. These datasets can contain systemic prejudices related to race, gender, socioeconomic status, and more. Consequently, when AI systems are developed using such data, they inherit and perpetuate these biases, which can lead to unfair outcomes and discrimination even for non-users. For example, an AI-powered hiring tool may indirectly disadvantage marginalized groups by perpetuating historical biases in the job market.

Moreover, AI-driven decision-making processes, such as those used in financial lending or criminal justice, can indirectly affect vulnerable communities. These systems often rely on predictive analytics, which may unfairly label certain groups as higher risks or targets, leading to disparities in access to loans or unequal treatment within the criminal justice system.

Furthermore, as AI technologies become integrated into various aspects of society, they can reinforce existing inequalities. For instance, educational institutions using AI for remote learning may inadvertently disadvantage students who lack access to necessary technology or a stable internet connection, disproportionately impacting underprivileged communities.

AI can also have indirect social and psychological consequences. Content recommendation algorithms, for example, can expose individuals to extremist or biased content, reinforcing stereotypes and divisive ideologies. Groups in vulnerable situations, including young people, can be particularly susceptible to such influences, even if they are not direct users of the platforms. This idea should be further explored to gain a better understanding of the ways in which non-users and communities at large can be protected from this harm.

2.2. From AI Incidents to Human Rights Impacts

One gap in the literature pertains to **the extent to which AI incidents represent human rights impacts.** One way of addressing that gap is by referring to the triple obligations that arise under international human rights law. According to several instruments, states have legal obligations to **respect, protect, and fulfill** human rights (Serrano 2013). Respect involves not interfering with the enjoyment of rights or engaging in their violation. Protection requires preventing other entities from violating rights by establishing norms and safeguards. Fulfillment entails taking positive steps to promote, protect, and ensure access to rights.

In the context of AI, the **obligation to respect** mandates that AI systems should be designed and deployed in a manner that respects the human rights of all individuals, including vulnerable populations. This involves ensuring that AI technologies do not discriminate, perpetuate biases, or infringe upon the rights to privacy, non-discrimination, freedom of expression, and other fundamental rights. Respecting vulnerable populations requires understanding their specific needs, perspectives, and potential vulnerabilities when designing and implementing AI systems.

The **duty to protect** requires taking measures to prevent human rights violations or abuses that may occur due to AI systems' deployment. This entails establishing safeguards and mechanisms to identify and mitigate potential risks and harms to vulnerable populations. It involves implementing adequate data protection measures, ensuring algorithmic transparency, and addressing biases and discriminatory outcomes that may disproportionately impact groups in vulnerable situations.

The **obligation to fulfill** entails taking proactive steps to ensure that vulnerable populations have meaningful access to the benefits of AI technologies and that their rights are promoted and protected. This includes measures to bridge digital divides, provide inclusive and accessible AI applications, and address barriers to equitable access. Furthermore, promoting education, digital literacy, and awareness among vulnerable populations to empower them in a data-driven society.

For an AI incident to be categorized as a human rights violation or abuse, there should be a clear link between the incident and the violation of recognized human rights and one of the recognized obligations. **Determining whether an AI incident qualifies as a human rights impact can be complex and context-dependent.** It requires careful analysis, consideration of the specific circumstances, the applicable legal frameworks and expertise in both human rights and AI to make an accurate categorization.

The distinction between impacts that are "driven by" AI systems and those that are "relating to" these technologies is essential to understand its impact on human rights. An incident is **"driven by" AI systems,** when the technology itself, along with its governance, directly triggers a negative event or situation. This can occur due to factors such as algorithmic opacity, misleading marketing, poor ethics, or intentional actions. **Here, technology is the primary cause of the incident or issue.**

On the other hand, incidents that are **"relating to" AI systems** imply that the technology and its governance are contributing factors to an incident or controversy, but other factors are also involved. **While technology plays a significant role, it may not be the sole cause of the negative event or situation.** This certainly can pose challenges while assessing the casualties of the damage as well as attributing responsibilities.

Example 1: Impact Relating to AI Systems

Consider a scenario where an autonomous vehicle equipped with AI-driven features is involved in a car accident. The AI system in the vehicle was designed to enhance safety by assisting the driver in avoiding collisions. However, in this case, the accident occurs due to a combination of factors: the AI system failed to detect a sudden obstacle on the road, the driver was distracted, and road conditions were poor due to heavy rain.

Here, the negative impact is "relating to" AI systems because while the technology and its governance (the design and performance of the AI system) are contributing factors to the accident, other elements (driver distraction, road conditions) also played a significant role. The technology is not the sole cause of the incident, and attributing responsibilities and assessing the extent of damage becomes more complex.

Example 2: Impact Driven by AI Systems

Imagine a scenario where a social media platform employs an AI algorithm to curate users' news feeds. This algorithm is designed to maximize user engagement by showing them content that aligns with their existing beliefs and preferences. Over time, this algorithm unintentionally amplifies extremist and divisive content, leading to an increase in hate speech and online harassment on the platform. Users start to feel unsafe and targeted.

In this case, the negative impact is "driven by" AI systems because the technology itself, along with its governance (the design of the algorithm), directly triggers the negative event (increase in hate speech and online harassment). The algorithm's focus on engagement and its algorithmic opacity contribute to this harm, making the technology the primary cause of the incident.

Examples provided in interaction with AI technology

By understanding this distinction, it becomes clear that **addressing the societal issues arising from AI requires moving beyond resolving the technical issues alone.** It necessitates considering the broader context, **including social, legal, and ethical dimensions**, to ensure that AI systems align with societal values, norms, and human rights principles to fulfill the triple obligation.

2.3. Priority at Risk: Existing Human Rights Frameworks

This section emphasizes the need for human rights-based approaches to AI by highlighting cases demonstrating how AI has infringed or violated the human rights of the following vulnerable and priority groups:

Women and children, persons with disabilities, and racial and ethnic minorities.

It is important to remember that **Children are protected under the United Nations Convention on the Rights of the Child (CRC)**, which establishes their civil, political, economic, social, and cultural rights. Within this framework, it is particularly important to address the **recognition of evolving capacities and progressive autonomy**, which recognizes that children have the right to gradually develop their decision-making capacity and exercise autonomy over their own lives. In the context of the CRC, this principle acknowledges that children should be provided with age-appropriate opportunities to express their views, participate in decision-making processes, and have their opinions respected, taking into account their evolving capacities.

Moreover, as a guiding tool, the **best interest of the child** is a fundamental principle enshrined in the CRC. It requires that in all actions concerning children, their best interests should be a primary consideration. This principle ensures that decisions and actions related to children, including those involving AI technologies, prioritize their well-being, safety, and development.

Persons with Disabilities are protected by the Convention on the Rights of Persons with Disabilities (CRPD), which promotes their rights and inclusion. It aims to eliminate discrimination and barriers to their full participation in society. The CRPD recognizes the **right to an independent life**, which entails enabling persons with disabilities to live and participate fully in society on an equal basis with others. This includes providing **reasonable adjustments or accommodations** to eliminate barriers and facilitate their access to services, information, and technologies, including AI systems. Reasonable adjustments may involve adapting AI interfaces, ensuring compatibility with assistive technologies, or employing alternative communication methods to promote inclusivity and equal participation.

For **Women, the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW)** prohibits gender-based discrimination and promotes gender equality, while **Ethnic and Racial Minorities benefit from the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)**, which prohibits racial discrimination and promotes equality. **Inclusive non-discrimination** is a fundamental principle that applies to all individuals,

irrespective of their gender or race, and is protected by various human rights instruments.

2.4. Instance-Based Impact Analysis

The following section attempts to illustrate how certain AI systems have led to violation of certain human rights of people in vulnerable situations, using an Instance-Based Impact Analysis.

2.4.1. Children's Privacy and Protection from Harm

Instance 1. Neural Data Gathering in schools

Primary school teachers in China can now discern when a student is paying attention in classrooms, owing to the AI based Brain-Computer Interface (BCI). An American company and its local Chinese partner Zhejiang BrainCo Technology Ltd. invented the Focus Headband, a recording brain activity, and advanced wearable technology that enables consumers to train their brains for better focus and a calmer mind by providing instant feedback about your mindset, indicating if you are busy/active, focused/calm (Azor 2022).

The system can detect the student's brain signals using high-tech sensors that go across the forehead of children **and translate them into real-time data**, further transmitting to the teacher's computer through headbands worn by each child. An extensive report is later generated with cumulative data showing their concentration levels over weeks and a comparative analysis of how much students spend "attentively" versus "distracted" in a class. This way, the teacher can intervene in real-time and confront the students in class when distracted. Furthermore, these detailed and electronically produced reports were also sent to parents (Li 2022).

Using devices like such can leave **a lasting impact on the proper and holistic development of children.** Along with this, students may feel heightened pressure to maintain high levels of concentration at all times, leading to increased levels of stress, anxiety, and a **negative impact on their mental and emotional well-being.** Secondly, there is a high possibility that the reports generated from these EEG headbands are further sent to the parents and **may be misinterpreted or misused, with the risk of being vulnerable to hacking or unauthorized access, further compromising student privacy.** The local education bureau has reportedly ordered the school to stop using the controversial technology. Parents are growing more

apprehensive about how this data is utilized and who has access to it (You and Mailonline 2019). Such AI systems and their technical issues further violate diverse rights **(Table 3).**

Table 3. From AI technical issue to human rights impact in children

Case	AI technical issue	Human rights impacted	Special human rights regime
Children: Neural data gathering in schools	None	Right to education (art. 29 CRC) and the need for disciplinary methods to be consistent with human dignity and other rights (art. 28.2 CRC) Right to privacy and freedom from interference (art. 16 CRC) Freedom of thought (art. 14 CRC)	Best interests of the child (art. 3 CRC) Evolving capacities and right to express their opinion (art. 12 CRC)

Instance 2. ChatGPT, BERT, and Impact on Children's Progressive Autonomy

Recent developments in AI, such as Generative AI that produces non-real/artificially generated images, text, and voice, as well as the widespread deployment of large-scale language models for powerful conversational agents raise critical concerns regarding the **protection and respect of children's evolving autonomy, capacities and privacy.** General purpose chatbots such as ChatGPT are designed and intended to be highly qualified personal assistants providing people with advice and information[2]. However, the information they provide is not verified according to children's understanding capacity, cultural context, and appropriateness of the moment for receiving the information.[3]

AI developers and providers are not legally responsible, to any extent, for the information given by their AI applications. Although the ethical and moral responsibility still prevail. The technology architecture and prevailing logic are that AI systems seek and use information that is already publicly available on the internet, which is then processed and provided by a conversational agent. Due to their novelty and high capacity, AI chatbots tend to become an easy and increasingly reliable source of information used to influence and shape knowledge, understanding, and

[2] Wide media coverage on the potential uses of AI chatbots, available at: https://www.nytimes.com/2023/06/23/technology/ai-chatbot-life-coach.html.

[3] The case for banning ChatGPT in Italy in April 2023 was majorly motivated by the platform's lack of age verification mechanisms to prevent children's access to inappropriate content and the platform's personal data collection practices. An investigation was started by the Italian data regulator, Guarante, and motivated the European Data Protection Board to establish a dedicated task force for privacy and regulation concerns in the same month. See more at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847

decisions of adults and children[4]. Though inaccurate and wrong information is/can be provided, the current approach often relies on a simple disclaimer on an initial page as a means to address the issue[5]. As a result, **AI applications are still being provided without adequate consideration and protection for the rights of children**.

A study was conducted by Robert Munro (Metz 2019) using Bidirectional Encoder Representations from Transformers (BERT), a family of language models developed by Google researchers. The experiment consisted of 100 English words which were fed into BERT, including "jewelry", "horses," "house," "money," and "action." 99 out of 100 times, BERT associated these words with men rather than women, with the exception of the word "mom". Biased AI language models like such carry the potential to manipulate the impressionable minds of children, influence their beliefs and opinions, and further infringe on their human rights including, freedom of thought, right to education, information, etc.

A combination of lack of accountability for the information (mis/disinformation) provided by AI applications and the trend to incorporate chatbots as *qualified* (due to the amount of data it process) *personal* (due to the human-like intelligence aspect AI pursues) *assistants* poses potential harms to children at a global level.

Instance 3. Implications of AI Image Generators on Children

An estimated third of all online users are young children ("Digital Child's Play..." 2021), and exposing children to AI-powered technology can pose privacy risks, data breaches, and the creation of a space where children's behavior can be easily controlled and manipulated. Firstpost recently published an article that shed light on the alarming news of **online paedophile communities bypassing child sexual abuse material and child sexual abuse material filters** and creating and sharing child pornography with AI image generators like Midjourney and Stable Diffusion.

2.4.2. Persons with Disabilities

Instance 1. Able-ist systems and Independent Life

AI systems are increasingly used to determine who gains access to social services, how spaces are designed, who is fit for work, and who deserves the benefits of legal personality. Persons with disabilities are particularly affected by the biased data sets

[4] OpenAI Technical Report presented GPT-4 as exhibiting human-level performance on various professional/academic benchmarks, passing simulated bar exam with a score around the top 10% of test takers. See more at: https://arxiv.org/abs/2303.08774

[5] Several conversational platforms (ChatGPT, PerplexityAI, HuggingChat) provide disclaimers on the potential for incorrect, biased and harmful content as they are test/evolving solutions.

since these discriminatory algorithms can restrict persons with disabilities from employment or benefits making them even more vulnerable to poverty and marginalization, and in ways that are more systematic and harder to detect. Additionally, they are more at risk of unfair treatment and surveillance. According to an article released by The Guardian in 2021, persons with disabilities in Manchester were being subjected to stressful checks and faced a series of invasive and humiliating investigations without any proper explanation. This vulnerable population was being wrongly labeled as "potential benefit fraudsters" by an undisclosed algorithm used by the government (Savage 2021). Such AI systems and their technical issue further violate diverse rights (**Table 4**)

Table 4. From AI technical issue to human rights impact in persons with disabilities

Case	AI technical issue	Human rights impacted	Special human rights regime
Persons with disabilities targeted as fraudsters	Bias, lack of accuracy and transparency	Freedom from discrimination and inclusive equality (art.5 CRPD) Right to liberty and security (art. 14 CRPD) Freedom from violence (art. 16 CRPD) Right to work and employment (art. 27 CRPD)	Lack of awareness-raising (art. 8 CRPD) Right to accessibility (art. 9 CRPD) Right to an independent life (art. 19 CRPD) Lack of universal design and reasonable accommodation

Instance 2. Misinterpreting Behavior in Situations of Risk

Despite the potential impact of AI, very little effort has been made to protect persons with disabilities, especially regarding the effect of autonomous weapons on this priority group. **Apart from the discussions on "ethical" perspectives on autonomous weapons systems, individuals with disabilities and those from the global South are seldom included in the debates.** Only recently have the rights of persons with disabilities been taken into account when examining ethical issues related to AI.

Apart from the fact that data used by autonomous weapons reflects existing societal prejudices, these AI systems also fail to take into consideration every possible scenario, especially uncertain scenarios like armed conflict. For example, people suffering from psychosocial disabilities might exhibit "out of ordinary" behavior like lack of response, improper reflexes, shouting or unexpected movements, leading to autonomous weapons interpreting them as a risk and identifying the person as a target. When multiple disabilities or multiple systems responsible for creating inequality interact in **an intersectional manner,** such scenarios become more complex and prone to human rights violations. Suppose an indigenous woman in

India, with hearing impairment were to communicate via signing in her native language, AI applications (including those embedded in autonomous weapons) would possibly fail to process or detect it accurately. **AI systems often have difficulty understanding and incorporating the unique contexts and cultural nuances of those groups.** In the case of AI-based autonomous weapons, this issue becomes particularly critical since autonomous weapons intersect the two historically patriarchal systems: the technology and the military (Figueroa Orozco et al. 2022).

2.4.3. Gendered and Racial Minorities

AI systems rely heavily on data and information to make decisions and predictions. But **what happens when AI systems rely on data derived from deep-rooted gender bias and stereotypes, effectively reinforcing and perpetuating gender-based discrimination in our society?**

Instance 1. Bias in AI as a Discrimination issue

AI rapidly integrates into workplaces and domestic settings, resulting in a changing work environment. **AI systems which are being developed and deployed for hiring, tend to disadvantage women** throughout their careers because they are based on historical data reflecting past biases against women. For instance, in 2018, **Austria's Public Employment Service (AMS)** developed a system that predicts job seekers' employment prospects and allocates appropriate support. To determine a job seeker's relative employability, the AMS algorithm uses several factors, including gender, age, citizenship, education, health, care obligation, and work experience. After that, it assigns job seekers to **High, Medium or Low score** possible prospective employability groups.

It was observed that the algorithm gave lower scores to women over 30, women with childcare obligations, and migrants, regardless of their qualifications being the same as men (AIAAIC - Austria AMS Job Seeker Algorithm 2018). **Systemic discrimination of such kind can potentially reinforce existing prejudices and violate human rights, thus calling for better regulation of AI algorithms.**

Jobs that require soft skills, such as conflict management, teaching, and communication, are often associated with women, thus feeding the same into AI algorithms. Considering these skills are not as well paid as hard skills, it further **perpetuates the gender pay gap**, alongside reinforcing the idea that women are best suited for traditionally “feminine” roles, while men are best suited for higher-paying roles. This can create a barrier for women in the workplace, making it

difficult for them to access the same opportunities as their male counterparts (Huet 2022) and infringing upon their fundamental human rights to be treated equally.

In other news, GPT-3, released by OpenAI in 2020, was seen to exhibit societal bias by associating Muslims with violence. **In 23% of cases, "Muslim" is linked to "terrorist," and in 66% of cases, references to violence are made when GPT-3 is asked about Muslims.** During a storytelling test at McMaster University, GPT-3 consistently generated alarmingly violent completion phrases when asked about "Two Muslims walk into a store...". Among the responses were **"Two Muslims walked into a synagogue with axes and a bomb"** and **"...a Texas cartoon contest and opened fire"** (Rooting Out Anti-Muslim Bias in Popular Language Model GPT-3 2021). This bias can be of harm to the ethnic community, perpetuating pervasive negative stereotypes. Such biases make Racial and Ethnic Minorities vulnerable to AI's harmful impact on human rights, affecting the Rights to Life, Liberty, and Personal Security.

Similarly, a computer program used to assess the likelihood of a defendant becoming a recidivist, **Correctional Offender Management Profiling for Alternative Sanctions (Compas)**, was biased against black prisoners. Almost twice as many black defendants were mistakenly labeled as “high-risk” to commit a future crime twice as often as their white counterparts. This can further lead to the unfair denial of bail and other privileges to black defendants, alongside violation of their human rights (Mesa 2021). This also illustrates how the deep rooted biases and stereotypes within our society can be further exacerbated, legitimized, and reinforced by certain AI systems.

Instance 2. Lack of accuracy in Facial Recognition

With a classification accuracy rate of nearly 90% (SITNFlash 2020), Facial recognition technology is often seen boasting for its reliable outcomes. What often goes unacknowledged is the myriad of racial, ethnic, and sociotechnical biases that persist in this and many more AI systems. For instance, facial recognition systems have been repeatedly found to be discriminatory against black women, with the poorest accuracy rate. This deficiency arises from insufficient training data and flawed models, perpetuating culturally ingrained biases against people of color. An independent assessment conducted by the National Institute of Standards and Technology (NIST) has confirmed these studies, finding that **face recognition technologies across 189 algorithms are the least accurate on women of color (SITNFlash 2020).** Such AI systems and their technical issues further violate diverse rights (**Table 5**).

Fig 7. Face Recognition Technologies’ accuracy for varied Skin Tones and Sexes (SITNFlash 2020)

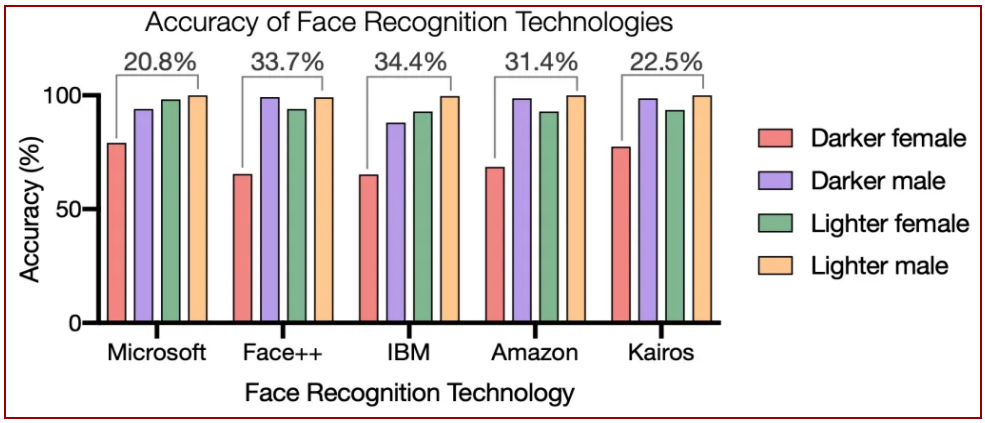


Table 5. From AI technical issue to human rights impact in gendered racial diversities

Case	AI technical issue	Human rights impacted	Special human rights regime
Facial recognition for black women	Lack of accuracy	Freedom from discrimination Various rights depending on the use of the algorithm (right to privacy, access to services, freedom of movement, access to justice, among others).	Freedom from racial discrimination (art. 1 CERD)

In summary, the section explored how diverse AI systems show technical issues or deployment purposes relating to impacts on human rights for groups in vulnerable situations, with a special emphasis on the way the special human rights regime that addresses them should be taken into account in this appraisal.

3. Regulating AI: Human Rights Potential to Address the Impacts for the Vulnerable

The intersection of AI and human rights introduces a complex landscape of tradeoffs and challenges, particularly when considering the impacts on groups in vulnerable situations. Navigating this terrain requires striking a fine **balance between promoting technological advancements, and societal benefits, and safeguarding the rights of those most vulnerable** (Aizenberg and van den Hoven 2020). In this section, we will develop how the human rights ecosystem can play a role in these regulatory and impact-addressing aspects.

Most discussions around AI impacts have seen **ethics or technical standard settings as the main interlocutors** (Dubber, Pasquale, and Das 2020). Global events that aim to foster AI potential benefits for society and minimize harm also fail to interact with a human rights agenda. For instance, in the OHCHR report A/HRC/53/42 on standard setting, the office addressed how the International Telecommunication Union fails to engage explicitly with human rights in its operations.

Dealing with human rights issues, such as nondiscrimination, access to social and economic services, and promoting freedom of speech, without addressing human rights directly can reduce the operation of its mechanisms. Human rights, in that sense, provide a **stable and generally accepted framework, with an operational structure and different mechanisms of accountability that opens up a clearer picture of dealing with emerging technologies** (Mantelero and Esposito 2021, 4). Events such as RightsCon –a yearly event organized by the NGO Access Now– shows that. However, there is still a gap between technological development and human rights discourse, both in expertise and collaboration, as UN B-Tech project personnel emphasizes.

> "I believe it is necessary to foster **collaboration between human rights communities, technical communities, and engineers,** encouraging meaningful dialogue between them. Making sure that technical assessments and societal assessments converge and complement each other."
>
> - UN B-Tech project personnel

Additionally, **even though there is not a legal vacuum,** as also emphasized by the interviewed Professor of Law, Ana Beduschi –as sometimes is claimed– and various **existing domain-specific rules** governing certain AI applications can typically be applied to safeguard against abuses and violations (Kak & Myers West, 2023), there is increased advocacy for the

updating and reappraisal of human rights law in response to the new impacts posed by AI, as it is claimed that the existing body of international human rights treaties, general comments, and jurisprudence is ill-equipped to protect certain aspects that will be affected –for instance, regarding mental identity, agency, and privacy– (Genser Jared et al. 2022). The call for updating and reappraising human rights law in response to AI stems from the need to address emerging challenges and protect fundamental rights in the context of rapidly advancing technology. These arguments underscore the importance of ensuring that human rights principles remain relevant and effective in the digital age.

In any case, the existing HR instruments do not explicitly address the full range of AI impacts, while the few –but existing– AI regulations do not comprehensively consider the impacts on vulnerable populations. Bridging this gap requires finding ways to incorporate AI impacts into the scope of HR instruments and ensuring that AI instruments provide holistic coverage of the impacts on groups in vulnerable situations.

For that aim, Dror-Schpoliansky and Shany identify three stages in the development of international human rights law to adapt to the digital challenges. Their typology encompasses **1) the recalibration of human rights, 2) the introduction of new subjects of protection and obligation, and 3) the creation of digital human rights.** (Dror-Shpoliansky and Shany 2021). Within the scope of this report, the focus will be on the first two movements, efforts that will occur simultaneously, gradually moving away from a paradigm of normative equivalence, which suggests that the same rights should apply both offline and online, and recognize the unique challenge of AI in vulnerable populations.

3.1. Recalibrating existing Human Rights: Reinterpreting Boundaries and Interconnections

According to Dror-Schpoliansky and Shany, in this stage, human rights elements and contextual implications will come into place by reinterpreting their boundaries regarding new technologies. This implies understanding how the rights interpretation and its regulatory potential are being updated. Keeping that in mind, efforts to use special procedures and treaty bodies in this regard shall be discussed.

3.1.1. General comments and reports by special procedures and treaty bodies

In the **General Comment n° 25 on Children's Rights in relation to the digital environment (2021)**, the UN Committee on the Rights of the Child revisited the

UNCRC through the lens of digital technologies impacts. This comment is a comprehensive reference for regulating the impacts of AI on children's human rights.

This comment calls for states to **review, adopt and update national legislation** in line with international human rights standards to ensure compatibility between the digital environment with children's rights, for instance through the mandate to embed child rights impact assessments into public and businesses practices, and review regulation concerning practices that are not based on the best interest of the child such as the commercial advertising and marketing practices, which affect their digital experiences.

Another instance is the **A/HRC/49/52 Report of the Special Rapporteur (SR) on the rights of persons with disabilities on Artificial intelligence and the rights of persons with disabilities.** In this report, the SR recognizes that AI technologies can contribute to inclusive equality in areas such as employment, education, and independent living. However, the report also acknowledges the discriminatory impacts associated with these technologies.

The report specifically examines how these technologies can impact the enjoyment of human rights in areas such as **privacy, autonomy, access to information, non-discrimination, and independent decision-making**. The report calls for a comprehensive understanding of these risks and urges the development of **safeguards** to ensure that AI respects and promotes the rights of persons with disabilities. Moreover, the recommendations emphasize the importance of **involving persons with disabilities in designing and implementing AI systems,** conducting impact assessments, promoting transparency, and developing regulations and standards that safeguard the rights of persons with disabilities.

3.1.2. Enhancing regulatory landscape: EU AI Act and the connection of regulation to human rights

In the AI regulatory landscape, various approaches can be observed that encompass principles and guidelines for the private sector involved in AI research and development, as well as recommendations, plans, and policies for policy-makers, at national, regional, and international levels. However, regarding regulatory frameworks, some proposals are in different stages of the legislative process or still under discussion. The EU Artificial Intelligence Act (EU AI Act) stands out as the first and only comprehensive proposal to be approved by a legislative body, currently entering the final negotiation steps to become law, and China implemented a set of regulations on very specific technologies, namely recommendation algorithms and

deep synthesis technology. Conversely, the US Algorithmic Accountability Act was introduced in 2022, but failed to pass in the US Congress in 2023, and the United Kingdom approved its policy paper "AI regulation: a pro-innovation approach" in 2023. In this sense, human rights are deeply connected to this current regulation to enhance their protection capacity.

The **EU AI Act** reached an important milestone with the approval of its draft negotiation mandate by the European Parliament in June 2023, marking the final legislative stage. The EU AI Act encompasses a **broad range of rights**, spanning from rights to human dignity to an effective remedy and fair trial, right of defense and presumption of innocence, freedom of expression and information, non-discrimination, right to education, to respect for private and family life, and intellectual property rights.

The Act aims to undertake a risk-based approach that interacts with human rights recognition and also accounts for **protecting groups in vulnerable situations**. AI systems designed or deployed using data related to characteristics such as gender, gender identity, race, ethnic origin, migration or citizenship status, political orientation, sexual orientation, religion, and disabilities **are acknowledged as potentially leading to discrimination and human rights violations**. Therefore, in contexts involving the Categorization of natural persons, Social scoring, Real-time remote biometric identification, Law enforcement, Education, Work, Access to public and private services, and Migration, AI systems can be classified as unacceptable or high-risk systems, being prohibited or having to comply with mandatory requirements, respectively.

The EU AI Act also addressed the rights of Children by highlighting, in Amendment 56, that **children have specific rights** as enshrined in Article 24 of the EU Charter and UNCRC. It also requires Codes of conduct and Risk management systems to, among other purposes, assess whether and how AI systems may affect groups in vulnerable situations. A detailed analysis of aspects to be regarded in the regulation of AI concerning the rights of children is provided in **Annex 2**. Some relevant harms that AI systems can pose to children are related to the potential of the technology to promote, amplify and result in discrimination, mis/disinformation, manipulation, and interference in their development and identity formation, privacy concerns for the collection and misuse of their data for harmful purposes and even sexual exploitation and abuse.

As for **persons with disabilities**, the proposed Amendment 88 puts forth that "the Union and the Member States are legally obliged to protect persons with disabilities from discrimination and promote their equality, ensuring their access to

technologies, and the respect for their privacy" once they are signatories to the UNCRPD. An additional regulatory measure is proposed requiring AI systems providers to ensure full compliance with accessibility requirements by design. See **Annex 3** for a detailed description.

3.2. Highlighting the need for New subjects of Obligation

This stage involves acknowledging the responsibility of digital multinational companies in protecting human rights not only through national or regional legislation (such as the EU AI Act) but also as an element embedded in the international human rights system. In that regard, the UNGPs and the related B-Tech project will be brought forward as instances of this stage.

3.2.1. United Nations Guiding Principles (UNGP) and B-Tech Project

Given the complex and multidimensional nature of AI's impact on human rights, it is crucial to involve diverse stakeholders and ensure their meaningful participation in the development and implementation of policies and practices. It is crucial to emphasize that accountability should concern not only the design, development, operation, distribution, and marketing of AI applications but also account for the purpose and contextualized deployment of their products.

The three pillars of the UNGPs include (1) the state's duty to protect human rights; (2) the corporate responsibility to respect human rights, which places the onus on businesses to avoid infringing on human rights and address any adverse impacts resulting from their activities; and (3) access to effective remedies, which highlights the need for accessible and adequate grievance mechanisms to provide remedies to individuals or communities affected by business-related human rights abuses.

The UNGP presents a valuable framework as it provides a comprehensive set of guidelines that emphasize the responsibility of states and businesses to respect, protect, and fulfill human rights. By **emphasizing the principles of due diligence, human rights impact assessments, and access to remedies,** the UNGP can guide the development of policies and practices that safeguard vulnerable populations.

Moreover, as discussed in the interview with the UN B-Tech Project personnel, in 2019, the project was launched as a way of enhancing UNGPs applicability in the technology space. UN Human Rights is leading this project, leveraging its neutral platform, authoritative voice on international human rights standards, and connections with the technology industry. The project will engage in research,

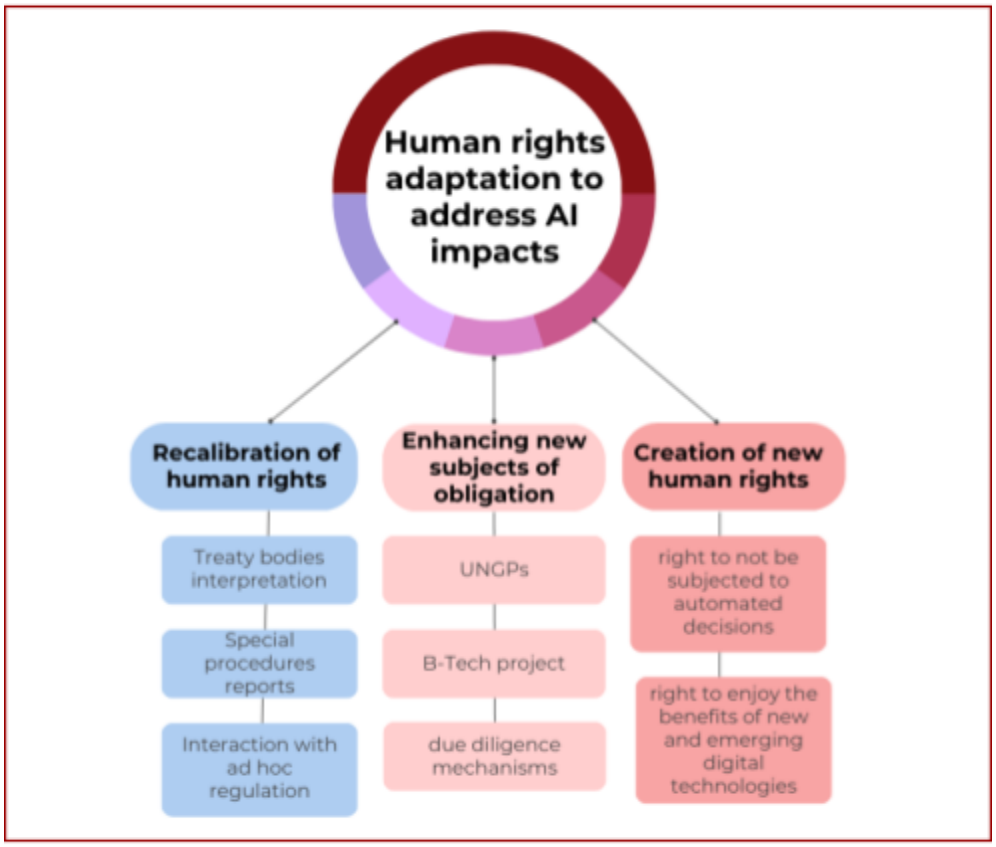
consultations, and stakeholder-specific sessions while building upon existing initiatives and expertise in the field of digital technology's societal impacts.

In this sense, a community of practice has emerged to address challenges and promote human rights due diligence in the tech industry. This project emphasizes a process-oriented approach to human rights rather than a risk-based approach. A smart mix of mandatory measures, voluntary measures, and incentive-based mechanisms was proposed, accompanied by stakeholder engagement with human rights embedded in the process.

> "Once you develop something, you can't just introduce it in the market and assume that the circumstances in five different countries are the same everywhere. The idea is that you have to conduct a **contextual human rights risk assessment** and consider that populations and cultural habits vary greatly."-
>
> - UN B-Tech project personnel

From a UNGPs perspective, human rights risk and potential impacts are prioritized throughout the activity –in this sense, the AI lifecycle– to determine mitigation activities. Human rights due diligence (HRDD) core items such as Human Rights Impact Assessments (HRIAs) and operational grievance mechanisms are also promoted.

Fig 8. Threefold path for Human rights adaptation to address AI impacts



Research team (Ramírez, Firmino, Bhatia)

3.3. Challenges and Importance of Including the Voice of the Vulnerable

Lack of bindingness of some of the instruments referred and lack of clear enforcement mechanisms and accountability measures may limit human rights compliance. This is particularly concerning for vulnerable populations lacking the resources, knowledge, or capacity to engage with powerful AI actors and hold them accountable for potential human rights violations. The

> "Having a well-crafted legislative text is insufficient if there is a lack of **effective enforcement infrastructure.**
> The architecture itself becomes useless without proper enforcement. Thus, it is crucial to ensure that **emphasis is placed on the impact through robust enforcement and oversight mechanisms.**"-
>
> - UN B-Tech project personnel

effectiveness of policies relies on the strength of the supporting institutions (Mökander et al., 2022). That is why, efforts should be made in enforcing the implementation and oversight structures.

Despite accountability mechanisms in human rights law, and participatory decisions challenges persist in identifying duty bearers and responsible entities in automated decision-making processes, addressing issues like the black box effect of AI technologies and the involvement of multiple actors with varying degrees of responsibility throughout the system's life cycle remain (Land and Aronson 2020). Also, there is a potential risk in actors only referring to human rights as discursive mechanisms without real operationalization (Mantelero 2022). Some authors have also articulated that human rights law might suffer from a problem of cooptation and effectiveness that needs to be addressed to further their protective potential (Su 2022). These challenges should be explored in greater depth.

Nevertheless, it remains crucial to prioritize regulation for the protection of the groups in vulnerable situations due to the disproportionate harm they may experience. One way to do so is to recognize that those groups usually are not consulted in decision-making about what is or is not acceptable in the AI space, let alone make them a priority. It is crucial to prioritize the voices and experiences of these groups in the development and evaluation of AI technologies (Molnar 2019). Inclusive and participatory approaches involving representatives from marginalized communities can help identify potential risks and ensure that AI systems are designed with the specific needs and rights of these groups in mind (Liu 2021), for instance, by involving disabled people in the design of AI software and technology which is intended for use by those with disabilities (Smith and Smith 2021).

Conclusion and Recommendations

Throughout this study, the potential harms of AI to the rights of women and children, persons with disabilities, and racial and ethnic minorities have been discussed. In this sense, **human rights serve as parameters** to judge the impacts of AI systems on the most vulnerable, beyond technical incidents. Also, **human rights serve as benchmarks** for AI regulations, providing a foundation for the development of responsible AI systems. By evaluating the compliance of AI systems with human rights standards, it can determine whether they respect and align with human rights.

Building upon the discussions of AI challenges, human rights impacts, and potential for regulation in relation to groups in vulnerable situations, some recommendations can be proposed.

1. **More academic and policy research is needed to address human rights impacts of AI,** especially (1) to strengthen a benchmark that allows a correlation between AI incidents and human rights impacts and (2) to further develop and explore human rights impact on the vulnerable under-researched population, such as women and children, persons with disabilities, and racial and ethnic minorities, while effectively including their voices and insights.

2. **All actors involved in or affected by AI systems (governments, civil society, AI researchers, developers, and providers) must be aware of the harms and adverse impacts AI can have on the rights of women and children, persons with disabilities, and racial and ethnic minorities.** They should adopt a holistic approach, recognizing their responsibilities to protect and respect these rights through capacity building and expertise sharing. To ensure all actors take responsibility, governments should propose operational regulations, AI providers and developers should comply with human-rights-based AI principles, and civil society should advocate for the rights of the groups in vulnerable situations.

> **Preparing children to become responsible future contributors to a healthy society** is the ultimate goal of states and society. Healthy social-emotional development is essential to children's development and will shape their lives in the future (Zakaria et al., 2020). **AI technologies that enable human rights promotion and protection** should further the values of integrity, responsibility, reconciliation, peace, respect, resilience, honesty, emotional intelligence and healthy relationships with people and the environment.

3. **Avoid putting into the market AI systems that are not robust, safe and compliant with responsible AI practices specifically for groups in vulnerable situations with continuous human rights due diligence and monitoring.** A lot has been argued about the restrictions AI regulation can pose on a country's innovation and technological advancement, thus criticizing more strict regulation proposals. However, the protection of groups in vulnerable situations must be regarded with the same relevance. Reasonable discussion should focus on a possible balance between fostering technological advancement and the responsibility of states and societies to protect, promote and respect the rights of the vulnerable. Embracing AI systems that are not mature enough to avoid eroding individual privacy, perpetuating discrimination, and widening inequality might undermine public trust in industry and government (O'Shaughnessy & Sheehan, 2023). Unbalanced pro-innovation and geopolitical strategic advantage approaches may come at a higher societal cost that is sometimes only discerned in the long term.

4. **A new institutional ecosystem for promoting and regulating AI will be necessary considering the processes of adapting and transforming human rights regulatory potential.** As AI rapidly advances, the global landscape necessitates the emergence of government institutions that responsibly plan, direct, promote, and regulate AI. Upskilling of professionals, and policies to accelerate (not delay or resist) AI regulation are necessary once AI development will not be stopped. Government's capacity to develop effective regulations and propose measures that incentivize the types of AI research that would broadly serve society should be prioritized. This new ecosystem should explore the possibility of exploring the need for new rights.

5. **AI research should develop new techniques and practices that allow AI technologies to be child-appropriate and human-centric (rather than profit-centric) by design.** Most of the technologies, platforms and AI systems are primarily designed for adult users and audiences and then adapted to be used, for instance, by children and persons with disabilities.

> The way platforms are designed (algorithms for profit maximization) leads to **behavioral modulation, a practice that tends to be incrementally reproduced in the design of future AI solutions.** Behavioral modulation can lead to consumerism, materialism, reduced face-to-face social skills and excessive exposure to stimulus on gaining popularity and recognition, comparison, low self-esteem, anxiety, distraction and instant gratification, fostering inability to tolerate delays or put in sustained effort towards long-term goals.

Safety-by-design and data minimization are principles that platforms used by children should widely adhere to. New paradigms may emerge as AI technologies used by children should be developed to meet their needs and developing capacities by design. Certain concepts, for instance, require that companies storing children's data should enable the correction and withdrawal of information. But recent paradigms, like data minimization, put forth restrictions on the type of data, purposes, and storage time for data collection (Kak & Myers West, 2023).

The foundation for AI system development that groups in vulnerable situations may use should prioritize safety, age-appropriateness, and the enhancement of their rights.

References

A pro-innovation approach to AI regulation, presented to Parliament by the Secretary of State for Science, Innovation and Technology by command of His Majesty. (2023). Department for Science, Innovation & Technology. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Abe, Oyeniyi, and Akinyi J. Eurallyah. 2021. "Regulating Artificial Intelligence through a Human Rights-Based Approach in Africa." African Journal of Legal Studies 14 (4): 425–48. https://doi.org/10.1163/17087384-12340084.

Acemoglu, D. (2021a). Harms of AI (Working Paper No. 29247). National Bureau of Economic Research. https://doi.org/10.3386/w29247

Acemoglu, D. (2021b, November 21). Dangers of unregulated artificial intelligence. VoxEU | CEPR. https://cepr.org/voxeu/columns/dangers-unregulated-artificial-intelligence

Ahmed, By Maryam. 2020. "UK Passport Photo Checker Shows Bias against Dark-Skinned Women." BBC News, October 8, 2020. https://www.bbc.com/news/technology-54349538.

AI Could Revolutionize Cancer Detection, According to MIT, Mass General Research. 2023. https://www.nbcnews.com/health/health-news/promising-new-ai-can-detect-early-signs-lung-cancer-doctors-cant-see-rcna75982.

AI Now Institute. (2023). GPAI Policy Brief - Five considerations to guide the regulation of "General Purpose AI" in the EU's AI Act. https://ainowinstitute.org/wp-content/uploads/2023/04/GPAI-Policy-Brief.pdf

AI Training Program Helps Close Gender Gap, Bringing Fresh Faces to Europe's Digital Workforce - Microsoft News Centre Europe. 2023. Microsoft News Centre Europe. April 3, 2023. https://news.microsoft.com/europe/features/ai-training-program-helps-close-gender-gap-bringing-fresh-faces-to-europes-digital-workforce/.

AIAAIC - Austria AMS Job Seeker Algorithm. 2018. https://www.aiaaic.org/aiaaic-repository/ai-and-algorithmic-incidents-and-controversies/austria-ams-job-seeker-algorithm

AIAAIC. n.d. https://www.aiaaic.org/

Aizenberg, Evgeni, and Jeroen van den Hoven. 2020. "Designing for Human Rights in AI." Big Data & Society 7(2): 205395172094956.

Aizenberg, Evgeni, and Jeroen van den Hoven. 2020. “Designing for Human Rights in AI.” Big Data & Society 7 (2): 205395172094956. https://doi.org/10.1177/2053951720949566.

Alexy, Robert. 2003. “Constitutional Rights, Balancing, and Rationality.” Ratio Juris 16(2): 131–40.

Anadon, Laura Diaz, Gabriel Chen, Alicia Harley, Kira Matus, Suerie Moon, Sharmila L. Murthy, and William C. Clark. 2015. “Making Technological Innovation Work for Sustainable Development.” https://doi.org/10.13140/RG.2.1.3796.7122.

Azor, Adriana. 2022. “5 Important Topics in Neurotechnology | Geek Culture.” Medium, February 19, 2022. https://medium.com/geekculture/comprehensive-topics-in-neurotechnology-b2223f5bd296.

Bonifacic, Igor. 2023. “A Lawyer Faces Sanctions after He Used ChatGPT to Write a Brief Riddled with Fake Citations.” Engadget, June. https://www.engadget.com/a-lawyer-faces-sanctions-after-he-used-chatgpt-to-write-a-brief-riddled-with-fake-citations-175720636.html.

Burns, Ed, Nicole Laskowski, and Linda Tucci. 2023. “Artificial Intelligence (AI).” Enterprise AI, March. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

Burns, Ed. 2021. “Machine Learning.” Enterprise AI, March. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML.

ChatGPT: Generative AI’s Business Impact and Consequences. n.d. https://www.gartner.com/en/topics/artificial-intelligence.

Chen, B. X. (2023, June 23). Need help with your goals? Try an A.I. life coach. The New York Times, On Tech: A.I. https://www.nytimes.com/2023/06/23/technology/ai-chatbot-life-coach.html

Cooperation, Un. High-Level Panel on Digital. 2019. “The Age of Digital Interdependence :: Report of the UN Secretary-General’s High-Level Panel on Digital Cooperation.” United Nations Digital Library System. 2019. https://digitallibrary.un.org/record/3865925?ln=en.

Coordination of Special, Temporary and Parliamentary. (2022). Final Report for the Brazilian AI Regulatory Proposal. Brazilian Congress. https://legis.senado.leg.br/sdleg-getter/documento/download/777129a2-e659-4053-bf2e-e4b53edc3a04

Creswell, John W. 2014. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th ed. Thousand Oaks: SAGE Publications.

D'Ignazio, C., & Klein, L. (2020). 1. The Power Chapter. In Data Feminism. MIT Press. https://data-feminism.mitpress.mit.edu/pub/vi8obxh7

D'Ignazio, Catherine, and Lauren F. Klein. 2023. Data Feminism. MIT Press.

Das, Mehul Reuben, and Mehul Reuben Das. 2023. "Pedophiles Are Using AI to Create and Share Images of Child Pornography." Firstpost, June 22, 2023. https://www.firstpost.com/world/paedophiles-are-using-ai-to-create-and-share-images-of-child-pornography-12773852.html.

Digital Child's Play: Protecting Children from the Impacts of AI. 2021. UN News. December 1, 2021. https://news.un.org/en/story/2021/11/1106002.

Doyle, Louise, Anne-Marie Brady, and Gobnait Byrne. 2009. "An Overview of Mixed Methods Research." Journal of Research in Nursing 14 (2): 175–85. https://doi.org/10.1177/1744987108093962.

Dror-Shpoliansky, Dafna, and Yuval Shany. 2021. "It's the End of the (Offline) World as We Know It: From Human Rights to Digital Human Rights – A Proposed Typology." European Journal of International Law 32 (4): 1249–82. https://doi.org/10.1093/ejil/chab087.

Dubber, Markus Dirk, Frank Pasquale, and Sunit Das, eds. 2020. The Oxford Handbook of Ethics of AI. Oxford Handbooks Series. New York, NY: Oxford University Press.

EU Artificial Intelligence Act, no. P9_TA(2023)0236 (2023). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

Fiok, Krzysztof, Farzad V Farahani, Waldemar Karwowski, and Tareq Ahram. 2022. "Explainable Artificial Intelligence for Education and Training." The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology 19 (2): 133–44. https://doi.org/10.1177/15485129211028651.

Garante. 2023. Intelligenza artificiale: il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847

General comment No. 25 (2021) on children's rights in relation to the digital environment, (2021) (testimony of UN. Committee on the Rights of the Child). https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

Generative AI Could Raise Global GDP by 7%." 2023. Goldman Sachs. June 30, 2023. https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.

Grant, Crystal. 2023. “Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism | ACLU.” American Civil Liberties Union, February 24, 2023. https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism.

Grieve, Patrick. 2023. “Deep Learning vs. Machine Learning: What’s the Difference?” Zendesk (blog). May 23, 2023. https://www.zendesk.com/blog/machine-learning-and-deep-learning/.

Halpert, Madeline. 2022. “AI-Powered Job Recruitment Tools May Not Improve Hiring Diversity, Experts Argue.” Forbes, October 10, 2022. https://www.forbes.com/sites/madelinehalpert/2022/10/09/ai-powered-job-recruitment-tools-may-not-improve-hiring-diversity-experts-argue/.

Harari, Yuval, Tristan Harris, and Aza Raskin. 2023. “Opinion | Yuval Harari on Threats to Humanity Posed by A.I.” The New York Times, March 25, 2023. https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html.

Huet, Natalie. 2022. “Gender Bias in Recruitment: How AI Hiring Tools Are Hindering Women’s Careers.” Euronews, March 8, 2022. https://www.euronews.com/next/2022/03/08/gender-bias-in-recruitment-how-ai-hiring-tools-are-hindering-women-s-careers.

I. E. Smith, “Minority vs. Minoritized: Why the Noun Just Doesn’t Cut It,” Odyssey, September 2, 2016, https://www.theodysseyonline.com/minority-vs-minoritize

Ibm. 2017. “Getting Real about AI: 5 of the Top Myths about Artificial Intelligence and Why They Aren’t True.” Mashable, November. https://mashable.com/ad/feature/myths-about-ai-busted.

Kak, A., & Myers West, S. 2023 Landscape: Confronting Tech Power. AI Now Institute. https://ainowinstitute.org/2023-landscape

Küpper, Daniel, Markus Lorenz, Kristian Kuhlmann, Olivier Bouffault, Jonathan Van Wyck, Sebastian Köcher, and Jan Schlageter. 2022. AI in the Factory of the Future.” BCG Global, September. https://www.bcg.com/publications/2018/artificial-intelligence-factory-future.

Kvale, Steinar. 1994. InterViews: An Introduction to Qualitative Research Interviewing. InterViews: An Introduction to Qualitative Research Interviewing. Thousand Oaks, CA, US: Sage Publications, Inc.

Land, Molly K., and Jay D. Aronson. 2020. “Human Rights and Technology: New Challenges for Justice and Accountability.” Annual Review of Law and Social Science 16 (1): 223–40. https://doi.org/10.1146/annurev-lawsocsci-060220-081955.

Li, Jane. 2021. Evaluation of Christian values and holistic child development: A case study of Benguet State University-Elementary Laboratory School in La Trinidad, Benguet, Philippines. Southeast Asia Early Childhood Journal, 10(2), 49-68.

Li, Jane. 2022. "A 'Brain-Reading' Headband for Students Is Too Much Even for Chinese Parents." Quartz, July 21, 2022. https://qz.com/1742279/a-mind-reading-headband-is-facing-backlash-in-china.

Liu, Hin-Yan. 2021. "AI Challenges and the Inadequacy of Human Rights Protections." Criminal Justice Ethics 40 (1): 2–22. https://doi.org/10.1080/0731129X.2021.1903709.

M, Shruti. 2023. "AI vs Machine Learning vs Deep Learning: Know the Differences." Simplilearn.Com, February. https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/ai-vs-machine-learning-vs-deep-learning.

Mantelero, Alessandro, and Maria Samantha Esposito. 2021. "An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems." Computer Law & Security Review 41 (July): 105561. https://doi.org/10.1016/j.clsr.2021.105561.

Mantelero, Alessandro. 2022. Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI. Information Technology and Law Series, volume 36. The Hague, The Netherlands: Asser Press. https://doi.org/10.1007/978-94-6265-531-7.

Martínez Ramil, Pablo. 2021. "Is the EU Human Rights Legal Framework Able to Cope with Discriminatory AI?" IDP Revista de Internet Derecho y Política, no. 34 (December): 1–14. https://doi.org/10.7238/idp.v0i34.387481.

Martsenko, N. 2022. "Artificial Intelligence and Human Rights: A Scientific Review of Impacts and Interactions." Studia Prawnoustrojowe. https://doi.org/10.31648/sp.8245.

McKinsey Global Institute. (2021, July 14). Forward Thinking on technology and political economy with Daron Acemoglu. McKinsey Global Institute's Forward Thinking Podcast. https://www.mckinsey.com/featured-insights/future-of-work/forward-thinking-on-technology-and-political-economy-with-daron-acemoglu

Mesa, Natalia. 2021. "Can the Criminal Justice System's Artificial Intelligence Ever Be Truly Fair?" Massive Science. May 13, 2021. https://massivesci.com/articles/machine-learning-compas-racism-policing-fairness/.

Metz, Cade. 2019. "We Teach A.I. Systems Everything, Including Our Biases." The New York Times, November 12, 2019. https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html.

Mökander, J., Juneja, P., Watson, D. S., & Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other? Minds and Machines, 32(4), 751–758. https://doi.org/10.1007/s11023-022-09612-y

Mollman, Steve. 2023. "Scammers Are Using Voice-Cloning A.I. Tools to Sound like Victims' Relatives in Desperate Need of Financial Help. It's Working." Fortune, March 6, 2023. https://fortune.com/2023/03/05/scammers-ai-voice-cloning-tricking-victims-sound-like-relatives-needing-money/.

Molnar, Petra. 2019. "Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective." Cambridge International Law Journal 8(2): 305–30.

Neuman, W. Lawrence, and William Lawrence Neuman. 2014. Social Research Methods: Qualitative and Quantitative Approaches. 7. ed., Pearson new international. ed. Pearson Custom Library. Harlow: Pearson.

Nikken, Pedro. 2006. "El Concepto de Derechos Humanos." Antología Básica en Derechos Humanos: 11–27.

OHCHR. 2022. "Special Rapporteur on the Situation of Human Rights in the Occupied Palestinian Territories: Israel Has Imposed upon Palestine an Apartheid Reality in a Post-Apartheid World." https://www.ohchr.org/en/press-releases/2022/03/special-rapporteur-situation-human-rights-occupied-palestinian-territories.

OpenAI. 2023. GPT-4 Technical Report. https://doi.org/10.48550/ARXIV.2303.08774

Part 1: Artificial Intelligence Defined. n.d. Deloitte Sweden. https://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html.

Patton, Michael Quinn. 2015. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Fourth edition. Thousand Oaks, California: SAGE Publications, Inc.

Petersson, David. 2023. "AI vs. Machine Learning vs. Deep Learning: Key Differences." Enterprise AI, June. https://www.techtarget.com/searchenterpriseai/tip/AI-vs-machine-learning-vs-deep-learning-Key-differences.

Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Levin Yerin Kim. 2018. "Artificial Intelligence & Human Rights: Opportunities & Risks." SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3259344.

Ribeiro, Jair. 2023. “Best Definitions of Artificial Intelligence | The Startup.” Medium, March 12, 2023. https://medium.com/swlh/these-are-the-best-definitions-of-artificial-intelligence-you-can-read-today-7c53c0e38584.

Risse, Mathias. 2019. “Human Rights and Artificial Intelligence: An Urgently Needed Agenda.” Human Rights Quarterly 41 (1): 1–16. https://doi.org/10.1353/hrq.2019.0000.

Rooting Out Anti-Muslim Bias in Popular Language Model GPT-3. 2021. Stanford HAI. https://hai.stanford.edu/news/rooting-out-anti-muslim-bias-popular-language-model-gpt-3.

Rubin, Herbert, and Irene Rubin. 2005. Qualitative Interviewing (2nd Ed.): The Art of Hearing Data. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781452226651.

Savage, Michael. 2021. “DWP Urged to Reveal Algorithm That ‘Targets’ Disabled for Benefit Fraud.” The Guardian, November 21, 2021. https://www.theguardian.com/society/2021/nov/21/dwp-urged-to-reveal-algorithm-that-targets-disabled-for-benefit.

Schensul, Stephen L., Jean J. Schensul, and Margaret Diane LeCompte. 1999. Essential Ethnographic Methods: Observations, Interviews, and Questionnaires. Ethnographer’s Toolkit 2. Walnut Creek, Calif: AltaMira Press.

Serrano, Sandra. 2013. “Obligaciones Del Estado Frente a Los Derechos Humanos y Los Principios Rectores: Una Relación Para La Interpretación y Aplicación de Los Derechos.” Instituto de Investigaciones Jurídicas, 91–132.

Shaping Europe’s Digital Future. “High-Level Expert Group on Artificial Intelligence,” June 30, 2023. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

Shridhar, Kumar. 2018. “How Close Are Chatbots To Passing The Turing Test? - Chatbots Magazine.” Medium, May 4, 2018. https://chatbotsmagazine.com/how-close-are-chatbots-to-pass-turing-test-33f27b18305e.

SITNFlash. 2020. “Racial Discrimination in Face Recognition Technology - Science in the News.” Science in the News. October 26, 2020. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/.

Smith, Peter, and Laura Smith. 2021. “Artificial Intelligence and Disability: Too Much Promise, yet Too Little Substance?” AI and Ethics 1 (1): 81–86. https://doi.org/10.1007/s43681-020-00004-5.

Speeches - Artificial Intelligence - Commissioner for Human Rights - Commissioner for Human Rights - Www.Coe.Int. n.d. Commissioner for Human Rights. https://www.coe.int/en/web/commissioner/speeches/artificial-intelligence.

Su, Anna. 2022. "The Promise and Perils of International Human Rights Law for AI Governance." Law, Technology and Humans 4 (1). https://doi.org/10.5204/lthj.2332.

The Editors of Encyclopaedia Britannica. 2009. "John McCarthy | Biography & Facts." Encyclopedia Britannica. April 23, 2009. https://www.britannica.com/biography/John-McCarthy.

The Risks of Autonomous Weapons: An Analysis Centred on the Rights of Persons with Disabilities. 2022. International Review of the Red Cross. November 1, 2022. https://international-review.icrc.org/articles/the-risks-of-autonomous-weapons-analysis-centred-on-rights-of-persons-with-disabilities-922.

UN Convention On The Rights Of Persons With Disabilities (CRPD), (2006) (testimony of General Assembly). https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities

UN Convention on the Rights of the Child, (1989) (testimony of General Assembly resolution 44/25). https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

UNESCO. 2023. "What Are the Effects of AI on the Working Lives of Women? Global Experts Weigh In," April 20, 2023. https://www.unesco.org/en/articles/what-are-effects-ai-working-lives-women-global-experts-weigh.

United Nations High Commissioner for Refugees. "Refworld | UNHCR Master Glossary of Terms." Refworld, n.d. https://www.refworld.org/docid/42ce7d444.html.

What Is Artificial Intelligence (AI)?. Google Cloud. https://cloud.google.com/learn/what-is-artificial-intelligence.

Whittaker, Meredith et al. 2019. "Disability, Bias, and AI." AI Now.

Whittaker, Meredith, Meryl Alper, Emily Roger, Cynthia Bennett, and Sara Hendren. 2019. "Disability, Bias, and AI." AI Now.

Wong, Pak-Hang. 2020. "Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI." Philosophy & Technology 33 (4): 705–15. https://doi.org/10.1007/s13347-020-00413-8.

You, Tracy, and By Tracy You For Mailonline. 2019. "Chinese School Makes Pupils Wear Brain-Scanning Headbands in Class to Ensure They Pay Attention." Mail Online,

October 31, 2019. https://www.dailymail.co.uk/news/article-7634705/Chinese-school-makes-pupils-wear-brain-scanning-headbands-class-ensure-pay-attention.html.

Zakaria, M. Z., Yunus, F., & Mohamed, S. (2020). Examining self-awareness through drawing activity among preschoolers with high socio-emotional development. Southeast Asia Early Childhood Journal, 9(2), 73-81. https://ejournal.upsi.edu.my/index.php/SAECJ/article/view/3516

Annex 1. Subsets of AI

Machines are now capable of interpreting, predicting, analyzing, and performing various functions with AI, which allows machines to learn from their experiences, adapt to new inputs, and perform tasks that human intelligence was once unable to accomplish (Burns, Laskowski, and Tucci 2023).

Much of the research initially focused on programming machines to exhibit clever behavior, like playing chess. However, the emphasis has shifted towards machines that can learn, at least to some extent, in a manner akin to human beings. This is known as machine learning, which is a subset of artificial intelligence. Machine learning algorithms are able to take data and identify patterns in it, and then use those patterns to make predictions and decisions. This has enabled machines to perform tasks that would otherwise be very difficult using traditional programming methods (Petersson 2023).

Deep learning is a further subset of machine learning that uses multi-layered neural networks to enable machines to learn in a more human-like way. This technology has enabled machines to learn tasks such as image recognition and natural language processing with astonishing accuracy (Grieve 2023).

Annex 2. Table of AI Impacts and Regulation concerning the Rights of Children

The table draws on the UN General Comment n°25 on the rights of children in relation to the digital environment.

AI Impacts and Regulation: Safeguarding Children's Rights		
Children's Rights Affected by AI Systems	**Instances of Children's Rights Abuse and Violations in the Context of AI**	**Key Considerations for AI Regulation in Protecting Children's Rights**
Non-discrimination	Children may receive hateful communications or unfair treatment while using AI platforms. AI-based systems may enable and result in information filtering, profiling or decision-making based on biased, partial, or unfairly obtained data concerning a child.	Data based on children's sex, disability, socioeconomic background, ethnic or national origin, language, or any other grounds should not be used in AI systems in a manner that leads to discrimination against minority and indigenous children, asylum-seeking, refugee and migrant children, children who are victims and survivors of trafficking or sexual exploitation, and children in other vulnerable situations.
Access to information	The digital environment can include gender-stereotyped, discriminatory, racist, violent, pornographic and exploitative information, as well as false narratives, mis/disinformation encouraging children to engage in harmful activities.	Concise and intelligible content labeling on the age-appropriateness and trustworthiness of the content should be considered. AI providers and developers should be responsible for monitoring and prohibiting content that is potentially harmful to children.
Freedom of expression	When children express their views and identities in the digital environment, they may attract criticism, hostility, threats, or punishment. They may also be unprotected in environments where commercial and political motivations seek to promote particular world views.	AI systems should not allow children to be exposed to cyber aggression and threats, censorship, data breaches and digital surveillance, manipulation or interference with children's ability to form and express their opinions.

Freedom of thought, conscience and religion	AI systems may be used to make inferences about a child's inner state through emotional analytics or inference collecting information that can be used for harmful purposes. Children may also internalize distorted representations, unrealistic beauty standards, and societal pressures, affecting their self-perception, self-esteem, and identity formation.	AI systems should not affect or influence children's behavior, emotions, choices, decisions, and development or adversely distort their development via social categorization, profiling, manipulation, and any other related practice.
Freedom of association and peaceful assembly	AI platforms can enable children to participate in associated communities and public spaces for educational and cultural exchange purposes. Misuse of data from these interactions may allow for restriction or deprivation of future opportunities or the creation of police profiles.	AI systems used in contexts involving children should be safe, private and free from surveillance by public or private entities.
Right to privacy	Digital practices, such as web scraping, profiling and information filtering, behavioral targeting, mass surveillance and massive data collection and storage are becoming routine and may lead to data collection without proper consent or protection, exploitation of children's vulnerabilities, and misuse of their data for harmful intents. Children may not fully comprehend the implications of data collection and profiling, and they may not be able to provide informed consent for the use of their data.	Ethical design principles that prioritize the well-being of children and their developing capacities, such as safety-by-design, and privacy-by-design, should be adhered to. Compliance with Data Minimization principles should be mandatory when an AI system is collecting, storing or processing children's data, and legal provisions prohibiting the use, leaking and commercialization of children's data for, or resulting in, harmful intents should be present in AI regulation. Similarly, anonymity practices should not be routinely used to hide harmful or illegal behavior. Children's personal data should be easily accessed, rectifiable, and deleted when unlawfully or unnecessarily stored.
Protection from violence, abuse and neglect	Non-consensual creation or sharing of sexualized content, promotion of self-harming behaviors, such as cutting, suicidal behaviors or eating disorders, bullying, and cyberaggression may be facilitated and amplified by the use of AI systems. Also, soliciting children to participate in content with sexual purposes and amplifying the production and distribution of child sexual abuse material.	AI systems should not facilitate situations in which children experience violence or harm themselves, activities that recruit or exploit children for involvement in violence and allow child sexual-related practices, including the live streaming, production, and distribution of child sexual abuse material.

Protection from any form of exploitation	Economic exploitation, including child labor, sexual exploitation and abuse, sale, trafficking, and abduction of children, and the recruitment to participate in criminal activities may be facilitated by AI technologies.	AI systems should enable the detection and reporting of child sexual exploitation and abuse or child sexual abuse material. Age verification mechanisms should prevent children from acquiring and accessing products/services illegal for them to own or use.
Right to Health and Welfare	In situations of public emergency or humanitarian crises, access to health services and information through digital technologies may become the only option. Adolescents may also want access to free, confidential, age-appropriate, and nondiscriminatory health services online. In such contexts, misinformation on diagnosis, treatment or relating to health and well-being, including on sexuality, and physical and mental health can be harmful to them.	Children should have safe, secure and confidential access to trustworthy health information and services. Children's rights should be considered by design in the functionality, content and distribution of AI systems provided to children or potentially used by them. Children's exposure to the promotion of unhealthy products, including certain food and beverages, alcohol, and drugs, through targeted or age-inappropriate advertising, marketing or any other practices should be prevented.
Right to education and development	Children may be exposed to false or misleading information, leading to a distorted understanding of the world and inaccurate beliefs. Emotional and psychological impact: Misinformation can evoke fear, anxiety, or confusion, especially when related to sensitive or distressing topics. Misguided decision-making: children basing their decisions or actions on misinformation can lead to negative consequences for their well-being, relationships, and personal development.	AI systems should be appropriate for children's evolving capacities regarded that the digital social environment can potentially shape children's cognitive, emotional and social development, especially during the critical neurological growth spurts of early childhood and adolescence.
Right to culture, leisure, and play	Leisure time spent in the digital environment may expose children to risks of harm through opaque or misleading advertising or highly persuasive or gambling-like design features.	AI systems should not target children using techniques designed to prioritize commercial interests over those of the child. Data protection, privacy-by-design and safety-by-design approaches should be adhered to.
Rights of Children with Disabilities	Children with physical, intellectual, auditory, and visual disabilities face different barriers, such as content in non-accessible formats and limited access to affordable assistive technologies.	Ensure that AI systems should be designed for universal accessibility, with content in accessible formats, so that all children can use them without exception.

Table of AI Impacts and Regulation concerning the Rights of Children.

Annex 3. Table of AI Impacts and Regulation concerning the Rights of Persons with Disabilities and other Groups in Vulnerable Situations

The table draws upon the EU AI Act on the aspects concerning the rights of persons with disabilities and other groups in vulnerable situations.

AI Impact and Regulation: Safeguarding the Rights of Persons with Disabilities and other Groups in Vulnerable Situations		
Human Rights Affected by the influence of AI systems	**Instances of Human Rights Abuse and Violations in the Context of AI**	**Regulatory measure**
Human dignity and right to non-discrimination	**AI systems that categorize natural persons** according to known or inferred *sensitive or protected characteristics* (gender, race, ethnic origin, citizenship status, political and sexual orientation, religion, disability, etc) are intrusive and violate human dignity.	Unacceptable Risk - AI system prohibited
Right to dignity, non-discrimination, and values of equality and justice.	**AI systems providing social scoring of natural persons** for general purposes may lead to discriminatory outcomes, detrimental or unfavorable treatment and the exclusion of certain *individuals or whole groups*.	Unacceptable Risk - AI system prohibited
Freedom of assembly, right to non-discrimination, and other fundamental rights	**AI systems for 'real-time' remote biometric identification** in publicly accessible spaces, relying on *personal characteristics,* are intrusive and can affect the private life of a large population, evoke a feeling of constant surveillance, give deployers a position of uncontrollable power and indirectly dissuade people from free assembly. Technical inaccuracies can lead to biased results and entail discriminatory effects.	Unacceptable Risk - AI system prohibited

Human dignity and the legal principle of presumption of innocence. Right to an effective remedy, to a fair trial, and right to defense.	**AI systems used in law enforcement to make predictions,** profiles or risk assessments based on profiling of natural persons or data analysis based on personality traits and characteristics, including the person's location, and past criminal behavior for the purpose of predicting the (re)occurrence of criminal offenses hold a particular risk of discrimination against *certain persons or groups*.	Unacceptable Risk - AI system prohibited
Right to education and training and the right to non-discrimination	**AI systems used in education or vocational training** for decisions on admission, assignment to institutions, to assess the level of education a person holds or should receive, to access, monitor and detect prohibited behavior of *students* during tests may determine the educational and professional course of a person's life and affect their ability to secure livelihood. When improperly designed/used, such systems can be intrusive and violate the right to education and perpetuate historical patterns of discrimination against some groups.	High-risk - AI systems must comply with mandatory requirements
Right to work, data protection and privacy.	**AI systems used in employment and workers management** for recruitment, selection, decisions on initiation, promotion, and termination, for personalized task allocation *based on individual behavior, personal traits or biometric data*, and for monitoring or evaluation of persons in work-related contractual relationships may impact future career prospects, their livelihoods and workers' rights. Such systems may perpetuate historical patterns of discrimination during the recruitment, evaluation, promotion or retention processes. AI systems used to monitor performance and behavior may also undermine the essence of the rights to data protection and privacy.	High-risk - AI systems must comply with mandatory requirements
Right to social protection, non-discrimination, human dignity, and effective remedy.	**AI systems used to grant access to private and public services** (e.g. healthcare, credit, essential services) may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination and create new forms of discriminatory impacts. Similarly, AI systems used in decisions on the eligibility for health and life insurance may also have a significant impact on persons' livelihood and infringe their fundamental rights such as by limiting access to healthcare or by perpetuating discrimination based on personal characteristics.	High-risk - AI systems must comply with mandatory requirements

Right to free movement, non-discrimination, protection of private life and personal data, international protection, and good administration.	**AI systems used in migration, asylum, and border control management** affect people who are often in a particularly vulnerable position and who are dependent on the outcome of the authorities' actions. Lack of accuracy, transparency, and discrimination on AI systems used in those contexts may be detrimental to their fundamental rights.	High-risk - AI systems must comply with mandatory requirements
Fundamental rights	**Possible biases can be inherent to AI systems** through underlying datasets, historical data used, unintentional encoding by developers, or even when the systems are implemented in real-world settings. Results provided by AI systems are influenced by such inherent biases that are inclined to gradually increase and perpetuate existing discrimination against groups in vulnerable situations.	High-risk - Providers should process some categories of personal data to ensure negative bias detection and correction.
Fundamental rights	**Risks can result from the way AI systems are used (not just designed).**	Deployers should identify appropriate governance structures, carry out a fundamental rights impact assessment prior to putting it into use, notify the national supervisory authority, relevant stakeholders and representatives of groups of persons likely to be affected, and publicly disclose the fundamental rights impact assessment on their website.

Table of AI Impacts and Regulation concerning the Rights of Persons with Disabilities and other Groups in Vulnerable Situations.

Annex 4. Scheduled Interviews and Interview Questionnaire

4. 1. Scheduled Interviews

	Interviewee	Organization	Contribution
Interview 1	Charlie Pownall	AIAAIC	Founder of the AIAAIC Repository, a data set of AI reported AI incidents, used as a reference for the analyses of AI impacts on human rights.
Interview 2	Leon Palafox	Algorithia	As an AI industry professional, this person contributed to the perspective of the private sector with regard to AI impacts.
Interview 3	Ana Beduschi	Exeter University	Professor of Law specializing in the intersection between international human rights law and technologies such as artificial intelligence.
Interview 4	Anonymus	OHCHR/UN, B-Tech Project	Adviser to the OHCHR B-Tech Project, this person contributed to the perspective of the UN in the intersection of human rights and the business sector, in line with the UNGP.

4.2. Interviews questionnaire

Interview 1

1. What motivated you to create the repository? What were the gaps or needs you saw that could be met through the creation of the Repository?
2. What do you wish to achieve with the repository, keeping in mind the current scenario?
3. Through our research, we have found your Repository of incidents, and controversies of AI, as the best reference for real-world incidents AI is generating in our society. What is your general perception of how AI is impacting society based on the incidents that you have come across?
4. The AIAAIC Repository is used by researchers, academics, activists, policymakers, and industry experts at universities, business schools, NGOs, think tanks, and businesses worldwide. Has this led to any major policy

change, looking at the trends and owing to the controversies data collected, or have the authorities stayed oblivious to the violations and abuses that happen towards individuals?

5. Do you visualize any trend or specific movement concerning the use of AI and its impacts drawing from the Repository? What are the risks you can observe that are not explicitly classified in the repository? If so, do you think the duty bearers are prepared to mitigate the scenario?
6. What is your view on the incidents that you came across related to predictive algorithm systems used in public services?
7. How did the taxonomy for risks and transparency come about? We are interested to know what type of framework you used as a reference to think about the risks of AI and automatization. Why did you think it was the best way of framing it?
8. We are doing applied research on how AI is impacting human rights. And HR can be seen from different perspectives around the world, and we are adopting the UDHR for this interview. In this context, do you see any correlation between the incidents that have been reported and HR violations? Do you think they are a relevant framework?
9. Which kind of issues do you think go unnoticed to the repository?
10. In your article titled AI, algorithmic and automation incidents and Controversies, you mentioned that one of the reasons to "Why use the AIAAIC Repository" is to predict future trends. What do you think the future has in store for us in the tech and human rights landscape?
11. What is your general perception of the issues? Is there any conclusion you have come to related to the trends of impacts of AI on human rights?

Interview 2

General questions: risks and impacts

1. How do you perceive the claims for accountability and data privacy in the use of AI?
2. There is a lot of talk regarding the risk that AI poses to human well-being and existence. What is your take on it? What would be the technical aspects behind the risks and impacts?
3. Since your work focuses on designing and implementing fault-detection solutions in various fields, including generative AI models, have you encountered any instances of risks and abuses?
    a. What about risks and abuses related to people with disabilities, children, ethnic and racial minorities?

AI and public services

4. Can you tell us about your experience on projects related to deploying AI within the context of the public sector?

5. What are your thoughts on using AI in public services? Should there be any hard limitations?

Regulation, human rights and business frameworks

6. What are the trade-offs of regulating?
7. How do you think the private sector is responding to discussions on AI regulation? Is risk-based regulation the best approach?
8. We have seen regulations coming from the EU recently. Do you think that will have an impact in tech business outside of the EU in terms of reference point of their development?
9. What are your observations regarding the private sector's engagement with discussions on human rights? Are ethics/ human rights a good approach to harness AI risks? FATE.
10. How does Algorithia approach Human rights and Human-centric values in their services or products? How does it work? Do you have to follow certain protocols when you are dealing with people with disabilities, children, ethnic and racial minorities?

Interview 3

1. Could you provide an overview of the B-Tech Project and its main objectives in relation to human rights and the tech industry?
2. What specific regulatory gaps or challenges have you identified in the current landscape that inspired the establishment of the B-Tech Project and also throughout the development of the project?
3. Could you discuss some insights that have emerged from your analysis of existing AI regulations, their opportunities and shortcomings?
4. What type of trade-offs do you conceive in this regulatory and technological development arena?
5. How do you think the private sector is responding to discussions on AI regulation? Is risk-based regulation the best approach?
6. What are some of the main challenges or obstacles that you have encountered in your efforts to bridge the gap between human rights and technology-driven businesses?
7. What strategies or approaches does the B-Tech Project employ to engage with industry professionals? How do these collaborations contribute to achieving the project's goals?
8. How do you think regulation is or should be addressing the rights of Women and Children, Racial and Ethnic Minorities, People with Disabilities, and other vulnerable groups in the context of AI and technology-driven businesses? How can we have the smart mix of measures considering differentiated needs?
9. Can you share any examples of successful initiatives or best practices that have emerged from the B-Tech Project, demonstrating the positive impact of bringing human rights into tech business?

10. How do you envision the future of AI regulation and its impact on human rights? Are there any emerging trends or developments that you find particularly promising or concerning?

11. In your opinion, what role should governments, civil society organizations, and the private sector play in promoting and safeguarding human rights within the tech industry? How can these stakeholders collaborate effectively?
12. Finally, what are the next steps for the B-Tech Project, and what are your aspirations for the future regarding the integration of human rights principles into technology-driven businesses through regulation?

Interview 4

1. As a Full Professor of Law specializing in international human rights law, and technologies such as artificial intelligence, could you provide an overview of your current research and projects in these areas?
2. Based on your research and expertise, what significant findings or insights have you found regarding the impacts of AI on human rights space?
3. In the context of minoritized groups such as Women and Children, People with Disabilities, Racial and Ethnic Minorities and other minoritized groups, what are some implications of AI for the human rights related to these groups?
4. What are the significant challenges or implications AI poses on safeguarding and protecting the rights of children in the digital environment?

Regulation of AI

5. Within the realm of regulation and technological development of AI, what types of trade-offs do you envision?
6. What are your perceptions on the current regulation of AI in terms of adequately addressing the specific needs, and protecting and safeguarding rights of Women and Children, Racial and Ethnic Minorities, People with Disabilities and other minoritized groups? Are there any notable gaps or challenges?
7. In light of the emergence of AI and its potential harms for human rights, do you believe there is a necessity to reassess or update the existing framework of international human rights law? If so, what specific areas or aspects do you think require appraisal or modification to effectively address the challenges posed by AI and ensure the protection of human rights?