Taxonomy of Human Rights Risks Connected to Generative AI

Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI

About this Paper

As a supplement to the UN B-Tech Project's foundational paper on generative AI, this document explores human rights risks stemming from the development, deployment, and use of generative AI technology. Establishing such a rights-based taxonomy is crucial for understanding how the United Nations Guiding Principles on Business and Human Rights (UNGPs) should be operationalised in addressing human rights risks connected to generative AI.

This taxonomy is concerned with demonstrating how the most significant harms to people related to generative AI are in fact impacts on internationally agreed human rights.

WHY HUMAN RIGHTS?

Public discourse around generative AI has broadly acknowledged some, though not all, of the ways this technology may negatively impact people and society. However, this discussion is often not framed in terms of how these impacts may undermine entitlements guaranteed by international human rights law. There are a number of logical and practical benefits to being precise about how generative AI may adversely impact human rights, in particular that:

- International human rights are currently the only internationally agreed set of moral and legal norms collectively expressed by humanity as central to living a life of dignity and respect.
- Focusing on human rights reinforces that existing State duties to protect human rights and corporate responsibilities to respect human rights can, and should be, invoked to govern generative AI.
- Human rights provide a focus on positive and negative outcomes that reach a threshold of impacting individuals' basic dignity.
- Focusing on human rights gives companies, regulators and civil society a well-established list of impacts against which to assess and address the impacts of generative AI systems.
- Human rights are connected to at-risk stakeholders' lived experience, and are also more specific than terms often used to describe the characteristics of generative AI systems, including '"safe", "fair", "responsible" or "ethical".

- There are international, regional and national social movements and institutions already tasked with promoting and protecting human rights, many of which are also actively engaged in elaborating what certain rights mean in new contexts.

The goal of this document is thus to provide a framework for understanding how negative societal impacts connected to generative AI products and services often also rise to the level of harming internationally protected human rights. Notably, these human rights impacts are often heightened for groups or populations already at elevated risk of becoming vulnerable or marginalised, including women and girls. Moreover, in some cases the societal and geographic contexts in which these impacts occur may heighten the saliency of risks, for example where adverse rights impacts occur in Global South contexts.

HUMAN RIGHTS RISKS

This taxonomy examines human rights that may be adversely impacted by generative AI, providing real world examples for each right. These rights, listed below by order of their appearance in the Universal Declaration of Human Rights (UDHR), are:

- Freedom from Physical and Psychological Harm
- Right to Equality Before the Law and to Protection against Discrimination
- Right to Privacy
- Right to Own Property
- Freedom of Thought, Religion, Conscience and Opinion
- Freedom of Expression and Access to Information
- Right to Take Part in Public Affairs
- Right to Work and to Gain a Living
- Rights of the Child
- Rights to Culture, Art and Science

Below, the taxonomy considers these rights in turn, providing for each: a summary of why the right is at risk from the development, deployment and/or use of generative AI; a selected list of key international human rights law articles pertaining to the right; and a list of real world examples in which generative AI may threaten the right in some way, including sources.

While the taxonomy does not attempt to comprehensively list all potential risks to human rights, it does offer an examination of some of the main ways in which human rights are currently at risk from generative AI. Many of these human rights risks may have been associated with earlier forms of AI, but risks may be, or in some cases already have been, altered or exacerbated by the particularities of generative AI. In other instances, harms linked to generative AI have manifested in novel ways.

As noted in the taxonomy's conclusion, generative AI is evolving rapidly. As use cases for this technology expand, and as the technology itself becomes better understood, additional associated risks to human rights will inevitably appear. The taxonomy is largely concerned with human rights risks and impacts

that are currently being observed or may soon be. Other potential risks are still emerging and in the future may represent some of the most serious threats to human rights linked to generative AI. In all cases, the taxonomy focuses on how generative AI specifically–rather than AI more broadly–poses risks to human rights.

Finally, it should also be noted that human rights are interrelated and interdependent. A single use case of generative AI may place several fundamental rights at risk, depending on factors such as the geography in which such systems are deployed, which groups of rightsholders are affected, and in which sector generative AI systems are used.

B-Tech

Freedom from Physical and Psychological Harm

SUMMARY

The right to bodily security, guaranteed across the international human rights framework, may be placed at risk from generative AI systems in various ways. In some cases, inadvertent misinformation originating in generative AI may lead to negative impacts on these rights, e.g., by causing harm to individuals' mental health. In others, generative AI model outputs may be used to intentionally threaten individuals' physical or psychological security or personal liberty. While some generative AI developers have enacted safeguards to prevent models from outputting such information, some safeguards are still reportedly easily circumvented.[1] Further, where models rely on human data labelling to assess the illegal or harmful nature of outputs, the speed with which such outputs can be removed may lag, allowing adverse rights impacts to proliferate.[2]

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right to life, liberty, and security of person" (UDHR Art. 3)
- "No one shall be subjected to torture or to cruel, inhuman, or degrading treatment or punishment" (UDHR Art. 5)
- "Every individual shall have the right to enjoy the best attainable state of physical and mental health" (ICCPR Art. 16)
- "No one shall be subjected to arbitrary arrest or detention." (ICCPR Art. 9)

RISK EXAMPLES

- Disinformation created with generative AI may be used in ways that risk inciting targeted physical violence against specific individuals or groups,[3] or destabilising societies in ways that risk inciting widespread, sporadic, or random violence (in relation to fictional terrorist attacks, coups, or electoral fraud, for example).
- Image and video generators may be used to create non-consensual sexualised content, including synthetic sexualised depictions of real, non-consenting individuals ("deepfake pornography") and/or depictions of violent sexual imagery.[4] Such generators may also be used to create child sexual abuse material.[5] In all cases, women and girls are at heightened risk.[6]
- Generative AI systems may mistakenly hallucinate false information. Some such misinformation may risk inciting physical violence against specific individuals or groups or exposing them to arbitrary deprivation of liberty.[7] In other cases, generative AI misinformation may encourage users to take actions that put their own physical or psychological well-being at risk,[8] especially those in vulnerable groups (e.g., individuals in mental health crisis).[9]
- Generative AI may be used to facilitate or enable human trafficking by creating content that is used to groom or lure individuals into situations of exploitation. As is the case with human trafficking in general, women and children are at greatest risk.[10]

B-Tech

Right to Equality Before the Law and to Protection against Discrimination

SUMMARY

The international human rights framework grants all people the right to equal protection against discrimination. The outputs of generative AI models are known to reflect cultural biases that are present in training datasets and on the internet at large.[11] This may result in the dissemination of harmful stereotypes based on race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, amplifying discrimination and existing socio-economic inequalities.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." (UDHR Art. 2)
- "All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against discrimination in violation of this Declaration and against any incitement to such discrimination" (UDHR Art. 7)

RISK EXAMPLES

- Generative AI models may produce derogatory or otherwise harmful outputs pertaining to people with marginalised identities, amplifying false and harmful stereotypes and facilitating various forms of discrimination throughout society.[12]
- Outputs of generative AI models often overrepresent culturally hegemonic groups (white, Western, male, heterosexual, cisgender, those who benefit from legacies of colonialism, etc.), which may lead to misrepresentation or underrepresentation of other groups at scale. This can entrench detrimental stereotypes, exacerbate biases and discrimination, and limit the ability of marginalised groups or individuals to exercise control over the representation of their identities in media and across the internet.[13]
- "Low resource languages" are often underrepresented in generative AI training datasets,[14] leading to underperformance of generative AI systems for speakers of these languages.[15] Underperformance of generative AI for users from certain linguistic, geographic, and cultural backgrounds may in itself constitute a form of discrimination, and threatens to widen the growing digital divide between high-resource and low-resource countries.
- Like other AI technologies used for automated decision-making (e.g., predictive policing or predictive recidivism technologies) generative AI models that are used for decision-making purposes may facilitate discrimination.[16]
- The concentration of generative AI development in the Global North serves to "[accelerate] exponentially the generation and processing of data" in these countries, exacerbating existing "data poverty" issues elsewhere in the world.[17] A lack of access to data has negative implications for economic development and thus for a variety of fundamental rights.

Right to Privacy

SUMMARY

Generative AI raises a number of concerns related to the right to privacy. Earlier iterations of AI have also been associated with these concerns, but some of the technical characteristics of generative AI models suggest a heightened risk of facilitating adverse impacts to these rights. These include the vast quantities of training data scraped from the internet by some large language models; large language models' reliance on ingesting data from individual users in the form of text prompts; and generative AI systems' capacity to create harmful, false and convincing content that may be used to directly attack an individual's privacy, honour or reputation.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "No one shall be subjected to arbitrary interference with their privacy, family, home or correspondence, nor to attacks upon their honour and reputation" (UDHR Art. 12)

RISK EXAMPLES

- Training data ingested by generative AI models may contain personally identifying information and other types of sensitive or private information in ways associated with novel privacy concerns.[18] Users' ability to provide informed consent to the collection, use and storage of their data for training of generative AI models may be compromised by the use of web-scraped datasets.[19]
- Users may input private or sensitive information into generative AI model prompts without fully understanding how their data will be collected, stored, and used. This data is often used to re-train models,[20] and it is unclear to what extent such sensitive information could reappear in subsequent model outputs to other users.
- The large scale collection, storage, and processing of data (including sensitive personal data) associated with generative AI models may increase vulnerabilities and user exposure to data breaches, hacks, and other security breaches. Some generative AI models can reportedly be hacked to extract copies of the data on which they were trained via "model inversion."[21]
- Data collected by generative AI models from users may be aggregated and sold without users' informed consent.
- Generative AI tools greatly reduce the difficulty of analysing and summarising massive corpuses of text data, including social media content. In some contexts, this may supercharge existing forms of State surveillance that risk privacy violations on a large scale.
- The capacity of generative AI to create individually targeted advertisements at scale may incentivize businesses to collect ever more personal information from users, with negative effects on the right to privacy.[22]
- Broadly, the generation of false, defamatory information pertaining to specific individuals constitutes an attack on a person's honour and reputation.[23] This may result from the intentional use of generative AI models to create and disseminate defamatory disinformation[24] or the unintentional hallucinations of generative AI models.[25]

Right to Own Property

SUMMARY

Generative AI models' ingestion of large quantities of data may entail adverse impacts to individuals' right to own moral and intellectual property. Training processes for some generative AI models may involve the unauthorised use of protected works, adversely impacting those works' original authors' right to own property. The capacity of generative AI systems to create content that mimics existing works by human creators also threatens original authors' property rights.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right to own property alone as well as in association with others [...] No one shall be arbitrarily deprived of his property." (UDHR Art. 17)
- "Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary, or artistic production of which they are the author." (UDHR Art. 27 (2))

RISK EXAMPLES

- Some generative AI models are trained on large quantities of text scraped from the internet, which may include sources that are intellectually protected.[26] This in itself can constitute an adverse impact to individuals' right to own property and protection of material interests where original authors have not given consent for their works to be used for training purposes.[27]
- Users may be able to use generative AI models trained on copyrighted data to generate content in the likeness of the original works of others in ways that negatively impact original authors' right to property.[28]
- In some cases, generative AI models may directly reproduce the original works of others, further adversely impacting original authors' right to property.[29]
- Generative AI models may be used to produce more effective content for cybersecurity attacks such as spam, phishing, or data breaches, resulting in loss or damage of property.[30]

B-Tech

Freedom of Thought, Religion, Conscience, and Opinion

SUMMARY

The autonomy to freely form and hold opinions is a core element of the international human rights framework. Human rights experts have previously warned that AI technology risks "invisibly [supplanting, manipulating or interfering] with the ability of individuals to form and hold their opinions..."[31] Generative AI may pose a heightened risk in this area. Many users will find it difficult to discern what internet content is synthetic and what is genuine–a recent study suggests that humans may actually be more likely to believe false information when it is created with generative AI tools.[32] Moreover, many generative AI systems that involve direct interaction with users are designed to mimic forms of human communication. When enacted without proper consent, these characteristics indicate a heightened risk that individuals' thoughts and opinions may be unduly and invisibly influenced by generative AI.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone shall have the right to freedom of thought, conscience and religion. ... No one shall be subject to coercion which would impair his freedom to have or to adopt a religion or belief of his choice." (UDHR Art. 18)
- "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (UDHR Art. 19)
- "Article 18 [ICCPR] does not permit any limitations whatsoever on the freedom of thought and conscience." (Human Rights Committee General Comment 22, para. 3)

RISK EXAMPLES

- The online dissemination of false information created with generative AI may threaten the right to free thought and opinion of internet users who encounter that information without knowing that it is false or of synthetic origin.[33] This could include, for example, manipulation of individuals' beliefs about politics or science through the spread of targeted disinformation intentionally created with generative AI tools or misinformation hallucinated by generative AI models.[34]
- Users' reliance on generative AI systems to formulate their ideas may also pose a risk to freedom of thought and opinion. Where people using generative AI tools to develop and express their thoughts incorporate incorrect, biased or incomplete information outputted by generative AI systems into their ideas, users' thinking may be manipulated in unforeseen ways.
- Some generative AI systems that involve direct user interaction are designed to mimic the communication patterns of human beings more than others.[35] Where such systems fail to make users sufficiently aware that they are not communicating with a human, users' ability to form opinions free of manipulation may be infringed. Even where users know intellectually that they are interacting with a machine, a generative AI system that is intentionally anthropomorphised to mimic a human may still infringe on these rights to free thought and opinion.

Continued on next page

Freedom of Thought, Religion, Conscience, and Opinion, *continued*

- Where generative AI systems fail to communicate to users about limitations on a system's performance or training data, this may also negatively impact the right to freedom of opinion. For example, where users are unaware that a system was only trained on information up until a certain date, or trained to prioritise certain viewpoints, they may unknowingly form opinions on the basis of incorrect and outdated model outputs.[36]
- The capacity of generative AI to create individually targeted advertisements at scale may grant businesses an unprecedented level of influence over internet users, in ways that may threaten the right to freedom of thought and opinion.[37]

B-Tech

Freedom of Expression and Access to Information

SUMMARY

The right to free expression is both a fundamental human right unto itself and core to a variety of other rights. Moreover, the right to access reliable information is a key element of the right to free expression.[38] Where individuals are prevented from accessing factual information or from determining what information is factual and what is not, the right to free expression is infringed. Generative AI–including the capacity to rapidly produce false content that appears human-generated and authoritative at scale–may pose risks to the right to free expression in various ways.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (UDHR Art. 19)
- "Freedom of information requires as an indispensable element of willingness and capacity to employ its privileges without abuse. It requires as a basic discipline the moral obligation to seek the facts without prejudice and to spread knowledge without malicious intent" (UNGA Resolution 59 of 1949)

RISK EXAMPLES

- Generative AI may be leveraged by malicious actors to create false but convincing content that is weaponised in targeted ways to threaten free expression–for example, the use of generative AI-created disinformation to harass journalists or political opponents into self-censorship.[39] Female public figures are especially at risk of targeted online harassment.[40]
- The proliferation of inaccurate internet content created with generative AI tools–whether disinformation or misinformation–may drown out or obscure evidence-based and fact-checked information online, broadly threatening individuals' and communities' right to access information. The consequences of this, including an erosion of public trust in news media and political processes[41], may negatively impact a variety of other rights as well as democratic governance more broadly.
- The underperformance of generative AI systems for speakers of low resource languages threatens these communities' right to freely express themselves using generative AI,[42] including by limiting their ability to access information through generative AI systems.
- The tendency of generative AI model outputs to overrepresent culturally hegemonic groups may result in a failure to produce a diversity of opinions and information by and about minority or historically oppressed groups. The resulting dearth of this information in online spaces represents a risk to the right to access information broadly.
- Overbroad moderation of generative AI output may result in limitations to freedom of expression (e.g., prohibiting users from generating graphic outputs which may be intended for purposes of artistic expression or satire, or limiting the potential of generative AI to aid individuals with disabilities in communicating their thoughts or emotions).

B-Tech

Right to Take Part in Public Affairs

SUMMARY

Recent electoral processes have been negatively impacted by an erosion of public trust in political processes and democratic institutions, in part fuelled by targeted campaigns using disinformation and misinformation on multiple media channels.[43] The organized spread of disinformation, in conjunction with other factors, can negatively impact the rights of individuals and communities to freely participate in public affairs in various ways.[44] Worryingly, generative AI-created content, often resting on statements made by political actors, expands this risk. False but convincing content resulting from generative AI platforms, including in the form of deepfakes, is now easier to produce at scale and could undermine the right to participate in public affairs around the world.[45] While some generative AI developers have announced plans to introduce safeguards, such as efforts to prevent AI chatbots from impersonating real candidates or government officials or watermarking AI images,[46] this is an area of concern in view of the scale and speed with which such content can spread, and the difficulty of fact-checking this information.[47]

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right to take part in the government of his country, directly or through freely chosen representatives.
- Everyone has the right of equal access to public service in his country.
- The will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures." (UDHR Art. 21)
- "Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2 and without unreasonable restrictions:
  a. To take part in the conduct of public affairs, directly or through freely chosen representatives;
  b. To vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors;
  c. To have access, on general terms of equality, to public service in his country." (ICCPR Art. 25)

RISK EXAMPLES

- Generative AI audio and/or video deepfakes featuring false depictions of political figures or depictions of fictional events with political significance may negatively impact individuals' right to vote freely for candidates or causes of their choosing. For example, individuals engaging with sophisticated deepfake depictions of politicians making fictional statements may unknowingly have their political beliefs altered based on falsehoods, hindering their ability to exercise their right to participate in public affairs, and potentially affecting the guarantee to an election which expresses the free expression of the will of the voters.[48] Even where deepfakes are not expressly presented as authentic, convincing depictions of real political figures may confuse voters and negatively impact their rights.[49]

Right to Take Part in Public Affairs, *continued*

- Deepfakes may also be used to suppress voter turnout. For example, fake depictions of election officials may disseminate false but convincing information about voting hours, location, eligibility, etc. This in turn can serve to disenfranchise voters. Alternatively, deepfakes of political candidates may falsely suggest that candidates are withdrawing from elections.[50]
- Generative AI may also pose risks to election processes and security. For example, deepfakes impersonating election officials may allow malicious actors to gain access to sensitive election security or administration information.
- Text content created with generative AI, for example in the form of fraudulent but authoritative-appearing statements or press releases from government agencies or political campaigns produced at scale, may also impact the right to participate in public affairs.[51]
- Where individuals form political opinions based on conversations with chatbots, generative AI's tendency to hallucinate false information may result in voters' political beliefs being manipulated through incorrect information from what they believe is an authoritative source.[52]
- Chatbots and other generative AI platforms may be used to automate the process of influencing public political opinion online through huge numbers of internet comments and social media posts related to politics. While this may not represent a risk to human rights in and of itself, in the hands of malicious actors these tools may greatly ease the launching of targeted campaigns to undermine free political processes in ways that do threaten the right to participate in public affairs.[53]

Right to Work and to Gain a Living

SUMMARY

Generative AI carries the potential to drastically alter economic and labour markets, as well as daily work practices. While some changes to the way humans work are inevitable as a result of technological advancement, generative AI is likely to have substantial economic and labour impacts that affect individuals' right to work and gain a living in a number of ways. Moreover, the deployment of generative AI in labour settings will disparately affect different groups of rightsholders and categories of jobs.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment." (UDHR Art. 23 (1))
- "The States Parties to the present Covenant recognise the right to work, which includes the right of everyone to the opportunity to gain his living by work which he freely chooses or accepts, and will take appropriate steps to safeguard this right" (ICESCR Art. 6 (1))
- The States Parties to the present Covenant recognise the right of everyone to the enjoyment of just and favourable conditions of work" (ICESCR Art. 7)
- "Each Member which ratifies this Convention shall promote continuous improvement of occupational safety and health to prevent occupational injuries, diseases and deaths, by the development, in consultation with the most representative organisations of employers and workers, of a national policy, national system and national programme." (ILO Promotional Framework for Occupational Safety and Health Convention Art. 2)
- "Workers and employers, without distinction whatsoever, shall have the right to establish and, subject only to the rules of the organisation concerned, to join organisations of their own choosing without previous authorisation." (ILO Freedom of Association and Protection of the Right to Organise Convention Art. 2)
- "Each Member for which this Convention is in force undertakes to declare and pursue a national policy designed to promote, by methods appropriate to national conditions and practice, equality of opportunity and treatment in respect of employment and occupation, with a view to eliminating any discrimination in respect thereof." (ILO Discrimination (Employment and Occupation) Convention Art. 2)

RISK EXAMPLES

- Companies may replace workers with generative AI tools[54] or pause hiring for roles that may be performed by generative AI in the future.[55] This displacement, especially where social safety nets are weak or non-existent, may threaten individuals' right to protection against unemployment. The categories of workers most likely to be displaced, such as clerical workers,[56] may mean disproportionate job losses among specific groups who are heavily represented in these professions, including women.[57]
- Human creatives are at elevated risk of being displaced by generative AI. Human artists in various fields have already been replaced by generative AI-created content[58] and some artists have engaged in collective bargaining in response.[59]

Continued on next page

Right to Work and to Gain a Living, *continued*

- Workers engaged in labour disputes with employers may be at heightened risk of being replaced with generative AI tools, with implications for workers' rights to association and protection against unemployment.[60]
- Generative AI models may be used by companies to monitor employee performance, raising concerns about the accuracy of such tools. Chatbots used to create employee performance reviews, for example, could introduce biases based on gender, race, or age.[61]
- The development of some generative AI systems requires data labelling to be performed by humans. Some generative AI developers have reportedly outsourced this work to workers in low-income countries who have described exploitative working conditions.[62]
- Chatbots and other generative AI-assisted search tools may direct users away from source websites where information is hosted. Resulting reductions in website traffic may cause small-medium sized online businesses to lose advertising revenue.[63]

B-Tech

Rights of the Child

SUMMARY

Children are especially susceptible to human rights harms linked to generative AI. Children are less capable of discerning synthetic content from genuine content, identifying inaccurate information, and understanding that they are interacting with a machine rather than a human being. These dynamics place children at heightened risk of adverse human rights impacts.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Motherhood and childhood are entitled to special care and assistance. All children, whether born in or out of wedlock, shall enjoy the same social protection" (UDHR Art. 25)
- "State Parties shall ensure to the maximum extent possible the survival and development of the child" (CRC Art. 6 (2))
- "State Parties recognise the important function performed by the mass media and shall ensure that the child has access to information and material from a diversity of national and international sources, especially those aimed at the promotion of his or her social, spiritual and moral well-being and physical and mental health." (CRC Art. 17)
- "State Parties shall take all appropriate legislative, administrative, social and educational measures to protect the child from all forms of physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation, including sexual abuse" (CRC Art. 19 (1))

RISK EXAMPLES

- Generative AI models or social media applications that utilise generative AI may lack age-appropriate restrictions, allowing underage users to share private information[64] or be exposed to content that is inappropriate or damaging for their age (e.g., online content pertaining to self harm or eating disorders,[65] which has proven especially damaging to girls historically[66]).
- Underage users may be more likely to turn to generative AI chatbots for advice on topics such as sexual health; the advice dispensed may contain misinformation that puts children at risk.[67]
- Generative AI models may affect or limit children's cognitive or behavioural development where there is over-reliance on these models' outputs, for example when children use these tools as a substitute for learning in educational settings. These use cases may also cause children to unknowingly adopt incorrect or biased understandings of historical events, societal trends, etc.[68]
- Generative AI models may create new avenues for children to be exposed to harmful content. Children may be more susceptible to misinformation, scams, or phishing attacks facilitated by generative AI models.[69]

B-Tech

Rights to Culture, Art and Science

SUMMARY

All people have equal rights to participate in culture, enjoy art, and benefit from scientific progress. These rights may be placed at risk by generative AI in various ways. The tendency of generative AI models to overrepresent certain cultures to the disadvantage of others in both their development and their outputs make these systems less accessible to and useful for diverse populations and speakers of low resource languages. Moreover, the capacity for generative AI outputs to supplant human-created art suggests further negative implications for the right cultural and artistic enjoyment.

RELEVANT HUMAN RIGHTS INSTRUMENTS

- "Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits" (UDHR Art. 27)
- "Everyone has the right to enjoy the benefits of scientific progress and its applications" (ICESCR Art. 15)

RISK EXAMPLES

- Where synthetic artistic content created by generative AI crowds out human-created art in on and offline spaces, this has broad negative implications for the rights to enjoyment of art and culture.[70]
- The under-performance of generative AI systems for speakers of low resource languages threatens these individuals' right to share in the scientific advancement that generative AI represents.
- The concentration of generative AI development in the U.S. and Europe[71]–especially in closed source development settings– may make generative AI systems less capable of producing outputs that accurately represent the cultural values, beliefs and practices of users from diverse geographies.[72] This may dissuade individuals from these cultures from using and benefiting from generative AI technology, imperilling their right to participate in culture.

Conclusion

Generative AI is a technology with wide-ranging impacts. To be sure, some of these impacts will be beneficial both for economic productivity and for the enhancement of human rights themselves. For example, generative AI, if developed, deployed and used responsibly, has the capacity to improve access to information, enable free expression, and potentially enhance the rights to health, education, due process, and access to public services, among others.

However, as detailed in this taxonomy, generative AI is also giving rise to risks to human rights. In some cases, generative AI is altering the scope of human rights risks already associated with some digital technologies. For example, online misinformation and disinformation are known to pose risks to free expression, and generative AI's capacity to create convincing, false content at scale is exacerbating those risks. Likewise, excessive collection of user data by businesses is a well known risk to the right to privacy. As the growth of generative AI shifts modalities of digital content creation away from word processors and other programs whose data is held locally on individual devices and toward chatbots and other platforms that grant businesses access to data inputted via user text prompts, the resulting quantum leap in the amount of user data held by companies is heightening existing privacy concerns.

In other cases, characteristics unique to generative AI are leading to human rights risks that differ not only in scope, but also in kind. For example, the intertwining of human agency with generative AI as people increasingly rely on these tools to develop and express their thoughts poses a risk to freedom of opinion and thought in novel ways that are only beginning to be observed. Elsewhere, generative AI is increasingly making powerful computer programming capacities that used to reside primarily in large organisations available to smaller organisations and individuals. Malicious actors' newfound access to these capabilities is opening up new avenues of risks to human rights, including privacy and property rights.[73]

Finally, generative AI is likely to pose additional risks to human rights that will emerge over the medium term. As this transformative technology evolves, and as corporate and government use cases evolve with it, we can expect that human rights will be threatened in new ways. For example, experts consulted by B-Tech raised concerns about the potential fusion of multiple generative AI models into single larger systems that autonomously generate inputs and outputs;[74] the implications of such systems for the autonomous dissemination of huge quantities of disinformation are troubling.[75] The application of generative AI to armed conflict is also still in its nascency but is sure to accelerate, bringing with it serious human rights concerns.[76]

All stakeholders should have an interest in ensuring that generative AI can deliver benefits to humanity without endangering human rights, both in the near term and in the future. Effectively identifying human rights risks linked to generative AI and acting to prevent, mitigate and remedy current and future human rights harms is essential to realising this goal. The UNGPs provide the authoritative framework for this process.

B-Tech

Acknowledgements

The UN B-Tech team expresses thanks to all the experts and stakeholders that provided input into this Generative AI Project supplement in particular representatives from the OECD, the Global Network Initiative, BSR and Shift.

Endnotes

[1] Cade Metz, 'Researchers Poke Holes in Safety Controls of ChatGPT and Other Chatbots', The New York Times, 27 July 2023, sec. Business, https://www.nytimes.com/2023/07/27/business/ai-chatgpt-safety-research.html.

[2] Garrett Morrow et al., 'The Emerging Science of Content Labeling: Contextualizing Social Media Content Moderation', SSRN Scholarly Paper (Rochester, NY, 3 December 2020), https://doi.org/10.2139/ssrn.3742120. p. 23

[3] Renate Mattar, 'Germany's Far Right Extremists Are Using AI Images To Incite Hatred', Worldcrunch, 7 April 2023, https://worldcrunch.com/tech-science/ai-images-extremists-germany.

[4] Tatum Hunter, 'AI Porn Is Easy to Make Now. For Women, That's a Nightmare.', Washington Post, 14 February 2023, https://www.washingtonpost.com/technology/2023/02/13/ai-porn-deepfakes-women-consent/.; Karen Hao, 'Deepfake Porn Is Ruining Women's Lives. Now the Law May Finally Ban It.', MIT Technology Review (blog), 12 February 2021, https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/.; PR Newswire, 'Generative AI Is the New Attack Vector for Platforms, According to ActiveFence Threat Intelligence', Yahoo Finance (blog), 23 May 2023, https://finance.yahoo.com/news/generative-ai-attack-vector-platforms-140000029.html.

[5] Guy Hedgecoe, 'AI-Generated Naked Child Images Shock Spanish Town of Almendralejo', BBC News, 23 September 2023, sec. Europe, https://www.bbc.com/news/world-europe-66877718.

[6] Laura Barron-Lopez, 'Women face new sexual harassment with deepfake pornography.', PBS Newshour, 27 June 2023, https://www.pbs.org/newshour/show/women-face-new-sexual-harassment-with-deepfake-pornography.

[7] Tiffany Hsu, 'What Can You Do When A.I. Lies About You?', The New York Times, 3 August 2023, sec. Business, https://www.nytimes.com/2023/08/03/business/media/ai-defamation-lies-accuracy.html.; Will Bedingfield, 'A Chatbot Encouraged Him to Kill the Queen: It's Just the Beginning', Wired, 10 October 2023, https://www.wired.co.uk/article/chatbot-kill-the-queen-eliza-effect.

[8] Chloe Xiang, '"He Would Still Be Here": Man Dies by Suicide After Talking with AI Chatbot, Widow Says', Vice (blog), 30 March 2023, https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says.

[9] Staff Writer, 'NEDA Suspends AI Chatbot for Giving Harmful Eating Disorder Advice | Psychiatrist.Com', Psychiatrist.com, 5 June 2023, https://www.psychiatrist.com/news/neda-suspends-ai-chatbot-for-giving-harmful-eating-disorder-advice/.

[10] Sarah Ferguson, 'Fleeing Boko Haram, and Tricked into Sexual Exploitation', UNICEF USA (blog), 3 January 2020, https://www.unicefusa.org/stories/fleeing-boko-haram-and-tricked-sexual-exploitation. See also Committee on the Elimination of Discrimination against Women, 'General Recommendation No. 38 (2020) on Trafficking in Women and Girls in the Context of Global Migration' (UN Convention on the Elimination of All Forms of Discrimination against Women, 20 November 2020), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CEDAW/C/GC/38&Lang=en.

[11] Roberto Navigli, Simone Conia, and Björn Ross, 'Biases in Large Language Models: Origins, Inventory, and Discussion', Journal of Data and Information Quality 15, no. 2 (22 June 2023): 10:1-10:21, https://doi.org/10.1145/3597307.

[12] Marie Lamensch, 'Generative AI Tools Are Perpetuating Harmful Gender Stereotypes', Centre for International Governance Innovation, 14 June 2023, https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/; Center for Countering Digital Hate, 'AI Image Tool Midjourney Is Being Used to Generate Conspiratorial and Racist Images', Research + Policy (CCDH, 11 August 2023), https://counterhate.com/research/ai-image-tool-midjourney-generate-racist-and-conspiratorial-images/#about ; Johana Bhuiyan, 'WhatsApp's AI Shows Gun-Wielding Children When Prompted with "Palestine"', The Guardian, 3 November 2023, sec. Technology, https://www.theguardian.com/technology/2023/nov/02/whatsapps-ai-palestine-kids-gun-gaza-bias-israel.

[13] Zachary Small, 'Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History', The New York Times, 4 July 2023, sec. Arts, https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html.

[14] Alexandre Magueresse, Vincent Carles, and Evan Heetderks, 'Low-Resource Languages: A Review of Past Work and Future Challenges' (arXiv, 12 June 2020), https://doi.org/10.48550/arXiv.2006.07264.

[15] Gabriel Nicholas and Aliya Bhatia, 'Lost in Translation: Large Language Models in Non-English Content Analysis', Center for Democracy and Technology (blog), 23 May 2023, https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/.

[16] Leonardo Nicoletti and Dina Bass Technology + Equality, 'Humans Are Biased. Generative AI Is Even Worse', Bloomberg. Com, 24 August 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias/; Chloe Xiang, ' Developers Created AI to Generate Police Sketches. Experts Are Horrified', Vice, 7 February 2023, https://www.vice.com/en/article/qjk745/ai-police-sketches.

[17] UNESCO, 'Guidance for Generative AI in Education and Research', July 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693, p. 14.

[18] Lauren Leffer, 'Your Personal Information Is Probably Being Used to Train Generative AI Models', Scientific American, 19 October 2023, https://www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/.; Benj Edwards, 'Artist Finds Private Medical Record Photos in Popular AI Training Data Set', Ars Technica, 21 September 2022, https://arstechnica.com/information-technology/2022/09/artist-finds-private-medical-record-photos-in-popular-ai-training-data-set/.

[19] Tim Biggs, 'Stolen information? The mystery over how tech giants train their AI chatbots', Sydney Morning Herald, 1 August 2023, https://www.smh.com.au/technology/lawsuits-highlight-opaque-methods-big-tech-firms-use-to-train-ai-chatbots-20230719-p5dpks.html.

[20] Nicholas Carlini et al., 'Extracting Training Data from Diffusion Models' (arXiv, 30 January 2023), https://doi.org/10.48550/arXiv.2301.13188.

[21] HM Government, 'Safety and Security Risks of Generative Artificial Intelligence to 2025' (UK: GOV.UK, 25 October 2023), https://assets.publishing.service.gov.uk/media/653932db80884d0013f71b15/generative-ai-safety-security-risks-2025-annex-b.pdf.

[22] Ben Jacobs, 'POV: How Generative AI Is Changing Surveillance Capitalism', Fast Company, 28 March 2023, https://www.fastcompany.com/90871955/how-generative-ai-is-changing-surveillance-capitalism.

[23] Emanuel Maiberg, 'It Takes 2 Clicks to Get From "Deep Tom Cruise" to Vile Deepfake Porn', Vice (blog), 17 May 2022, https://www.vice.com/en/article/qjb7b7/ethical-deepfakes-deep-tom-cruise-ai-generated-porn.

[24] Ashish Jaiman, 'The Danger of Deepfakes', The Hindu, 1 January 2023, sec. Technology, https://www.thehindu.com/sci-tech/technology/the-danger-of-deepfakes/article66327991.ece; Joseph Cox, 'AI-Generated Voice Firm Clamps Down After 4chan Makes Celebrity Voices for Abuse', Vice (blog), 30 January 2023, https://www.vice.com/en/article/dy7mww/ai-voice-firm-4chan-celebrity-voices-emma-watson-joe-rogan-elevenlabs.

[25] Vishwam Sankaran, 'ChatGPT Cooks up Fake Sexual Harassment Scandal and Names Real Law Professor as Accused', Independent, 6 April 2023, https://www.independent.co.uk/tech/chatgpt-sexual-harassment-law-professor-b2315160.html.

[26] Kai Nicol-Schwarz and Tim Smith, 'Why Harry Potter Is the Copyright Timebomb under Generative AI Models', Sifted, 18 May 2023, https://sifted.eu/articles/generative-ai-copyright/.

[27] Gil Appel, Juliana Neelbauer, and David A. Schweidel, 'Generative AI Has an Intellectual Property Problem', Harvard Business Review, 7 April 2023, https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem.

[28] Gil Appel, Juliana Neelbauer, and David A. Schweidel, 'Generative AI Has an Intellectual Property Problem', Harvard Business Review, 7 April 2023, https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem; https://news.artnet.com/art-world/class-action-lawsuit-ai-generators-deviantart-midjourney-stable-diffusion-2246770

[29] Kali Hays, 'OpenAI Now Tries to Hide That ChatGPT Was Trained on Copyrighted Books, Including J.K. Rowling's Harry Potter Series', Business Insider, 15 August 2023, https://www.businessinsider.com/openais-latest-chatgpt-version-hides-training-on-copyrighted-material-2023-8.

[30] Emily Flitter and Stacy Cowley, 'Voice Deepfakes Are Coming for Your Bank Balance', The New York Times, 30 August 2023, sec. Business, https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html.

[31] U.N. Doc. A/73/348; see also Evelyn Aswad, 'Losing the Freedom to Be Human', SSRN Scholarly Paper (Rochester, NY, 29 February 2020), https://papers.ssrn.com/abstract=3635701.

[32] Giovanni Spitale, Nikola Biller-Andorno, and Federico Germani, 'AI Model GPT-3 (Dis)Informs Us Better than Humans', Science Advances 9, no. 26 (28 June 2023): eadh1850, https://doi.org/10.1126/sciadv.adh1850.

[33] Christopher Mims, 'Help! My Political Beliefs Were Altered by a Chatbot!', Wall Street Journal, 13 May 2023, sec. Tech, https://www.wsj.com/articles/chatgpt-bard-bing-ai-political-beliefs-151a0fe4.

[34] See, for example: Pranshu Verma and Will Oremus, 'AI Voice Clones Mimic Politicians and Celebrities, Reshaping Reality', Washington Post, 16 October 2023, https://www.washingtonpost.com/technology/2023/10/13/ai-voice-cloning-deepfakes/; Microsoft Threat Intelligence, 'Sophistication, Scope, and Scale: Digital Threats from East Asia Increase in Breadth and Effectiveness', September 2023), https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1aFyW ; Alexandra Ulmer et al., 'Deepfaking It: America's 2024 Election Collides with AI Boom', Reuters, 31 May 2023, sec. United States, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/.

[35] Evan Selinger, 'We Don't Want Chatbots to Come off as People', Boston Globe, 8 May 2023, https://www.bostonglobe.com/2023/05/08/opinion/google-bard-chatgpt-dishonest-anthropomorphism-evan-selinger/.

[36] Gray Drenik, 'AI Tools Like ChatGPT Will Fail Without Data Integrity - Here's Why', Forbes, 30 May 2023, https://www.forbes.com/sites/garydrenik/2023/05/30/ai-tools-like-chatgpt-will-fail-without-data-integrity–heres-why/.

[37] Ben Jacobs, 'POV: How Generative AI Is Changing Surveillance Capitalism', Fast Company, 28 March 2023, https://www.fastcompany.com/90871955/how-generative-ai-is-changing-surveillance-capitalism ; Marcello Ienca, 'On Artificial Intelligence and Manipulation', Topoi 42 (20 June 2023): 1–10, https://doi.org/10.1007/s11245-023-09940-3.

[38] Irene Khan, The Special Rapporteur of the Promotion and Protection of the Rights to Freedom of Opinion and Expression, 'Disinformation and Freedom of Opinion and Expression' (Geneva: Human Rights Council, 13 April 2021), https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement.

[39] Summer Lopez, Nadine Farid Johnson, and Liz Woolery, 'PEN America Report on Artificial Intelligence and Free Expression', PEN America (blog), 31 July 2023, https://pen.org/report/speech-in-the-machine/.

[40] Lucina Di Meco and Saskia Brechenmacher, 'Tackling Online Abuse and Disinformation Targeting Women in Politics', Carnegie Endowment for International Peace, 20 November 2020, https://carnegieendowment.org/2020/11/30/tackling-online-abuse-and-disinformation-targeting-women-in-politics-pub-83331.

[41] Christine Hauser, 'Right-Wing Operatives Plead Guilty in Voter-Suppression Scheme', The New York Times, 25 October 2022, sec. U.S., https://www.nytimes.com/2022/10/25/us/politics/ohio-robocalls-wohl-burkman-guilty.html.; Philip Marcelo, 'FACT FOCUS: Fake Image of Pentagon Explosion Briefly Sends Jitters through Stock Market', APNews, 23 May 2023, AP Fact Check edition, https://apnews.com/article/pentagon-explosion-misinformation-stock-market-ai-96f534c790872fde67012ee81b5ed6a4; Lubna Jerar Naqvi, 'Beware of AI-Generated Propaganda amid Political Unrest', Geo News, 11 May 2023, Opinion edition, https://www.geo.tv/latest/486758-beware-of-ai-generated-propaganda-amid-political-unrest.

[42] Alena Gorbacheva, 'No Language Left Behind: How to Bridge the Rapidly Evolving AI Language Gap', UNDP, 6 October 2023, https://www.undp.org/kazakhstan/blog/no-language-left-behind-how-bridge-rapidly-evolving-ai-language-gap. ; Kira Schacht, 'Bridging the AI Language Gap in Africa and Beyond', DW, 29 July 2023, https://www.dw.com/en/bridging-the-ai-language-gap-in-africa-and-beyond/a-66331763.

[43] Gabriel R. Sanchez and Keesha Middlemass, 'Misinformation Is Eroding the Public's Confidence in Democracy', Brookings, 26 July 2022, https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in-democracy/; Lilian Olivia, 'Disinformation was rife in Kenya's 2022 election', Africa at LSE, 5 January 2023, https://blogs.lse.ac.uk/africaatlse/2023/01/05/disinformation-was-rife-in-kenyas-2022-election/; Camille Elemia, 'In the Philippines, a Flourishing Ecosystem for Political Lies', New York Times, 6 May 2022, https://www.nytimes.com/2022/05/06/business/philippines-election-disinformation.html

[44] Nita Bhalla, 'Online Disinformation Stokes Tensions as Kenya Elections Near', Reuters, 27 June 2022, https://www.reuters.com/article/idUSL4N2Y22HF/.

[45] Tiffany Hsu, Stuart A. Thompson and Steven Lee Myers, 'Elections and Disinformation Are Colliding Like Never Before in 2024', New York Times, 9 January 2024, https://www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html

[46] Ali Swenson, 'Here's how ChatGPT Maker OpenAI Plans to Deter Election Misinformation in 2024', Associated Press, 17 January 2024, https://apnews.com/article/ai-election-misinformation-voting-chatgpt-altman-openai-0e6b22568e90733ae1f89a0d54d64139

[47] Tiffany Hsu, Stuart A. Thompson and Steven Lee Myers, 'Elections and Disinformation Are Colliding Like Never Before in 2024', New York Times, 9 January 2024, https://www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html

[48] Daniel Zuidijk, 'Deepfakes in Slovakia Preview How AI Will Change the Face of Elections', Bloomberg, 4 October 2023, https://www.bloomberg.com/news/newsletters/2023-10-04/deepfakes-in-slovakia-preview-how-ai-will-change-the-face-of-elections; 'Deepfakes deceive voters from India to Indonesia before elections', Economic Times, 3 January 2024, https://economictimes.indiatimes.com/news/elections/lok-sabha/india/deepfakes-deceive-voters-from-india-to-indonesia-before-elections/articleshow/106504149.cms.

[49] Alex Thompson, 'First look: RNC Slams Biden in AI-Generated Ad', Axios, 25 April 2023, https://www.axios.com/2023/04/25/rnc-slams-biden-re-election-bid-ai-generated-ad

[50] Archis Chowdhury, 'Deepfake Videos Ahead Of Bangladesh Polls Should Have Us All Concerned', Decode, 11 January 2024, https://www.boomlive.in/decode/deepfake-elections-disinformation-bangladesh-india-us-uk-indonesia-24087.

[51] Marina E. Franco, 'AI and Weakened Electoral Body Threaten Mexican Presidential Elections', Axios, 26 October 2023, https://www.axios.com/2023/10/26/mexico-elections-amlo-artificial-intelligence.

[52] 'ChatGPT and Co: Are AI-driven search engines a threat to democratic elections?', Algorithm Watch, 5 October 2023, https://algorithmwatch.org/en/bing-chat-election-2023/; David Gilbert, 'Microsoft's AI Chatbot Replies to Election Questions With Conspiracies, Fake Scandals, and Lies | WIRED', Wired, 15 December 2023, https://www.wired.com/story/microsoft-ai-copilot-chatbot-election-conspiracy/.

[53] Nathan E. Sanders and Bruce Schneier, 'How ChatGPT Hijacks Democracy', New York Times, 15 January 2023, https://www.nytimes.com/2023/01/15/opinion/ai-chatgpt-lobbying-democracy.html

[54] Pranshu Verma and Gerrit De Vynck, 'ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners.', Washington Post, 5 June 2023, https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/.

[55] Brody Ford, 'IBM to Pause Hiring for Jobs That AI Could Do', Bloomberg.Com, 1 May 2023, https://www.bloomberg.com/news/articles/2023-05-01/ibm-to-pause-hiring-for-back-office-jobs-that-ai-could-kill.

[56] Pawel Gmyrek, Janine Berg, and David Bescond, 'Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality', Working Paper (ILO, 21 August 2023), https://www.ilo.org/wcmsp5/groups/public/—dgreports/—inst/documents/publication/wcms_890761.pdf.

[57] Mark McNeilly, 'Will Generative AI Disproportionately Affect the Jobs of Women?', Kenaninstitute.Unc.Edu (blog), 18 April 2023, https://kenaninstitute.unc.edu/kenan-insight/will-generative-ai-disproportionately-affect-the-jobs-of-women/.

[58] Viola Zhou, 'AI is already taking video game illustrators' jobs in China', Rest of World, 11 April 2023, https://restofworld.org/2023/ai-image-china-video-game-layoffs/.

[59] Jake Coyle, 'In Hollywood writers' battle against AI, humans win (for now)', AP News, 27 September 2023, https://apnews.com/article/hollywood-ai-strike-wga-artificial-intelligence-39ab72582c3a15f77510c9c30a45ffc8.

[60] Kate Wells, 'Can a Chatbot Help People with Eating Disorders as Well as Another Human?', NPR, 24 May 2023, sec. National, https://www.npr.org/2023/05/24/1177847298/can-a-chatbot-help-people-with-eating-disorders-as-well-as-another-human.

[61] Kieran Snyder, 'We Asked ChatGPT to Write Performance Reviews and They Are Wildly Sexist (and Racist)', Fast Company, 2 March 2023, https://www.fastcompany.com/90844066/chatgpt-write-performance-reviews-sexist-and-racist.

[62] Karen Hao and Deepa Seetharaman, 'Cleaning Up ChatGPT Takes Heavy Toll on Human Workers', Wall Street Journal, 24 July 2023, sec. Tech, https://www.wsj.com/articles/chatgpt-openai-content-abusive-sexually-explicit-harassment-kenya-workers-on-human-workers-cf191483; Partnership on AI, 'Responsible Sourcing of Data Enrichment Services', 17 June 2021, https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Responsible-Sourcing-of-Data-Enrichment-Services.pdf#page=27.

[63] Ravi Sen and The Conversation, 'After 25 years of growth for the $68 billion SEO industry, here's how Google and other tech firms could render it extinct with AI', Fortune, 21 October 2023, https://fortune.com/2023/10/21/how-generative-ai-could-change-google-search-68-billion-seo-industry/.

[64] Kashmir Hill, 'Can You Hide a Child's Face From A.I.?', The New York Times, 14 October 2023, sec. Technology, https://www.nytimes.com/2023/10/14/technology/artifical-intelligence-children-privacy-internet.html.

[65] Geoffrey A. Fowler, 'AI Is Acting "pro-Anorexia" and Tech Companies Aren't Stopping It', The Washington Post, 7 August 2023, https://www.washingtonpost.com/technology/2023/08/07/ai-eating-disorders-thinspo-anorexia-bulimia/.

[66] Tech Transparency Project, '"Thinstagram": Instagram's Algorithm Fuels Eating Disorder Epidemic', Dangerous by Design (blog), 8 December 2021, https://www.techtransparencyproject.org/articles/thinstagram-instagrams-algorithm-fuels-eating-disorder-epidemic.

[67] Connie Dimsdale, 'Teenagers Turning to AI Chatbots for Sexual Health Advice as Sex Education Put under Review', Inews.Co.Uk, 11 August 2023, https://inews.co.uk/news/teenagers-ai-chatbots-sexual-health-advice-inaccurate-information-2534555.

[68] Christopher Mims, 'Help! My Political Beliefs Were Altered by a Chatbot!', Wall Street Journal, 13 May 2023, sec. Tech, https://www.wsj.com/articles/chatgpt-bard-bing-ai-political-beliefs-151a0fe4.

[69] Roseann O'Reilly Runte, 'Young People Tell Us They Need Help Identifying Misinformation', Scientific American, 16 August 2023, https://www.scientificamerican.com/article/young-people-tell-us-they-need-help-identifying-misinformation/.

[70] David De Cremer, Nicola Morini Bianzino, Ben Falk, 'How Generative AI Could Disrupt Creative Work', Harvard Business Review, 13 April 2023, https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work.

[71] Irene Solaiman, 'The Gradient of Generative AI Release: Methods and Considerations' (arXiv, 5 February 2023), http://arxiv.org/abs/2302.04844. p. 3.

[72] Irene Solaiman et al., Evaluating the Social Impact of Generative AI Systems in Systems and Society, 2023. p. 6. https://www.researchgate.net/publication/371490141_Evaluating_the_Social_Impact_of_Generative_AI_Systems_in_Systems_and_Society

[73] Yuen Pin Yeap, 'Generative AI is the Next Tactical Cyber Weapon for Threat Actors', The Guardian, 16 October 2023, https://www.forbes.com/sites/forbestechcouncil/2023/10/16/generative-ai-is-the-next-tactical-cyber-weapon-for-threat-actors/

[74] Yohei Nakajima, 'Task-Driven Autonomous Agent Utilizing GPT-4, Pinecone, and LangChain for Diverse Applications – Yohei Nakajima', 28 March 2023, https://yoheinakajima.com/task-driven-autonomous-agent-utilizing-gpt-4-pinecone-and-langchain-for-diverse-applications/.

[75] Will Knight, 'It Costs Just $400 to Build an AI Disinformation Machine', Wired, 29 August 2023, https://www.wired.com/story/400-dollars-to-build-an-ai-disinformation-machine/.

[76] BHRRC, 'The Application of Generative AI to Warfare Raises Human Rights Concerns', Business & Human Rights Resource Centre, 2 May 2023, https://www.business-humanrights.org/en/latest-news/the-application-of-generative-ai-in-warfare-raises-human-rights-concerns/. ; Eric Schewe, 'As Militaries Adopt AI, Hype Becomes a Weapon', Jstor, 4 August 2023, https://daily.jstor.org/as-militaries-adopt-ai-hype-becomes-a-weapon/.; Katrina Manson, 'The US Military Is Taking Generative AI Out for a Spin', Bloomberg, 5 July 2023, https://www.bloomberg.com/news/newsletters/2023-07-05/the-us-military-is-taking-generative-ai-out-for-a-spin.