Protecting People with Disabilities: A Guide for Non-Technical Committee Members in Understanding the Regulations Needed to Design Ethical AI

by Trisha Neogi

Submitted to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Inclusive Design Toronto, Ontario, Canada, 2024

ABSTRACT

Artificial Intelligence (AI) promises large-scale efficiencies that enable faster and “better” decisions. What was once a tool for researchers and technologists has now been made accessible to corporations, regulators, and individuals. Through its rate of development and increased adoption, AI systems and tools are being used to replace human decision-making at a speed that surpasses regulation and intervention. The speed of mass AI adoption and lack of regulation towards protecting communities most impacted by the technology. This is resulting in statistical discrimination and cumulative harm against the most vulnerable groups in society, people with disabilities.

To bring attention to the statistical discrimination and cumulative against people with disabilities, this design and research project contributes to the work of the Capacity Building Seed group and their efforts in standardizing and publishing equitable AI regulations as part of the Accessible Standards Canada priorities. This design and research project contributes to bridging the technical and legal gaps for non-technical committee members that require this information to make informed decisions about the proposed clauses. The outcome of this design and research project is a capacity building resource, which supports a larger working group who developed the Seed Standards, which are proposed regulatory standards for equitable AI regulations that protect people with disabilities in efforts to prevent further harm.

Keywords: Artificial Intelligence, AI Regulation, Trustworthy AI, People with Disabilities, Statistical Discrimination, Data Outlier, Cumulative Harm.

ACKNOWLEDGEMENTS

Dr. Jutta Trevianous, Principal Advisor

Jutta, I cannot thank you enough for your patience, guidance, and support. You have exemplified what it means to be an inclusive and accommodating advisory and design practitioner. I am so inspired by your work and advocacy.

Lisa Liskovoi, Capacity Building Resource Group member

Angelika Seeschaaf-Veres, Capacity Building Resource Group member

Lisa and Angelika, thank you for your support, review, and feedback on the capacity building resource. Your time and guidance have been integral to the creation of this resource.

Contributors of the Seed Standard and framework

To all the individuals that have contributed to this meaningful document- thank you. You have opened my eyes and changed my perspective on how to think about AI systems.

1. INTRODUCTION

1.1 My Background

On November 20, 2022, OpenAI launched its early demo of ChatGPT, an AI chatbot that was capable of learning the complexities of human language and emotion.[1] While artificial intelligence[2] is not a new concept, ChatGPT's viral launch is the first to make large language models[3] accessible to the larger population.[4] As a chatbot, ChatGPT's accessibility gave way to excitement and renewed energy to experiment, create, and innovate, where individuals and corporations have been utilizing the technology to increase efficiency.[5] However, the launch of ChatGPT also gave way to immense criticism, backlash, and fear. Fear of change is not unusual when something new is released, however, the accessibility of ChatGPT made critics weary of the tool's validating, reliability, and over-promised functionality.[6] Irrespective of research and opinion, ChatGPT has given the public access to a technology that was predominately used by technologists and researchers, and as someone with a non-technical background, the launch of ChatGPT and its limited transparency gave me concern.

When ChatGPT first launched, I was a first-year graduate student in the Inclusive Design program at OCAD University and working a corporate job as a human resources[7] practitioner. I originally wanted to focus my major research project[8] on corporate organizational structures and whether corporate hierarchies are oppressive towards marginalized employees. However, when ChatGPT launched, I became specifically interested in its critiques of how replacing human decision-making correlates to the decline in human-thinking and information manipulation.[9] This interest stems from my corporate work experience utilizing AI-driven recruitment tools to hire candidates. I

[1] (Marr, 2024)
[2] Artificial Intelligence - AI
[3] Large Language Models - LLMs
[4] (Edwards, 2023)
[5] (Marr, 2024a)
[6] (Chomsky et al., 2023)
[7] Human Resources - HR
[8] Major Research Project - MRP
[9] (Dans, 2023)

have utilized AI-hiring tools in the past to help screen candidate resumes. However, as a non-technical person, I did not have confidence in the results produced by the AI tool as it produced results without sufficient explainability. In addition, I was concerned of potential biases in the technology as the AI hiring tool I was experimenting with did not transparently disclose the tool's decision-making process. The gender discrimination perpetuated by Amazon's AI hiring tool[10] is a well-known cautionary tale in the HR space, which concerned me when utilizing an AI hiring tool myself. As someone who believes in building diverse workplaces, I felt it was important to mitigate biases where possible, which led to me stop using AI-based hiring tools altogether. I held similar reservations when experimenting with ChatGPT, and although I was impressed by the chatbot's capability and accessibility, my concerns grew around the tool's validity, reliability, and potential for misinformation. I decided to pivot my MRP research to focus on ethical AI and better understanding AI implications towards people experiencing the most risk and harm.

As someone without a technical background,[11] my understanding of how LLMs are trained, its decision-making process, and the formula behind its operations is limited. In addition, the majority of AI systems and tools lack data and decision transparency,[12] and without transparency into AI decisions, I question how a non-technical person, such as myself, can verify the tool's trustworthiness and ensure biases are limited. In addition to questioning the trustworthiness of AI systems and tooling, I question the impact AI systems may have towards vulnerable groups and whether these systems are inherently biased. As AI continues to evolve and be adopted, I question how regulators are considering the implications of AI towards civil populations[13] when designing AI standards and policy.

1.2 Individuals Most Impacted- People with Disabilities

[10] (Dastin, 2018)
[11] In this MRP, a non-technical background refers to a lack of digital literacy.
[12] (Dhinakaran, 2023)
[13] Civil population (civilians) are considered individuals who do not represent a business, industry, or government entity

Generative AI typically leverages historical data.[14] To train AI, the historical data set is used to help the tool learn the ideal result, or target optima. AI, therefore, is trained to pursue an optimal pattern. When it comes to utilizing datasets, AI has been trained to look at commonalities to search for the optimal patterns. Historically marginalized minorities, including people with disabilities tend to be the data outliers in a data distribution.[15] Jutta Treviranus has dubbed this as the "Human Starburst"[16] where 80 percent of the data points cover the central 20 percent of data distribution, in contrast to the marginalized minorities, whose data points are outside of the dominant starbursts or clusters. Although spread out, these marginalized data points however, cover 80 percent of the data distribution. Seeking the statistical average or statistical power provided by the central cluster, or optimal pattern, enables AI to reach a concentrated number of people, which enables the system to impact a large dataset without needing further customization. This is how AI can be applied to many users and scale, mirroring the pattern of economies of scale in other markets. However, in the pursuit of the data average, data outliers, or the 20 percent of data points outside the dominant cluster, are ignored by AI, or experience undue harm. This is known as the "outlier problem," which refers to the difficulty of representing individuals who are beyond the dominant average.[17] This algorithmic exclusion has been defined as "statistical discrimination"[18] by Dr. Treviranus within the Seed standard or proposed draft standard and affects people with disabilities as they diverge from the optimal pattern, or statistical average. As the system continues to statistically discriminate against data outliers,19 harm begins to accumulate, further exacerbating the risk and harm of generative AI systems and tools.[19]

People with disabilities are most likely to be at the extreme edges of data. On the one hand, when design includes people with disabilities, the opportunities can be life changing as technology can make things possible. On the other hand, when technology excludes people with disabilities, it can cause immense harm as the system or tool will

[14] Seed Standards Document- https://idrc.ocadu.ca/
[15] (Treviranus, 2020)
[16] (Treviranus, 2019)
[17] Seed Standards Document- https://idrc.ocadu.ca/
[18] (Treviranus, 2020)
[19] (Treviranus, 2020)

discriminate against them.[20] A common misconception[21] is that people with disabilities can be grouped together in their similarity of diverging from the average, thus creating its own data cluster. However, due to the unique characteristics of each disability and how disabilities may present differently on different people, people with disabilities are their own individual data sets and cannot be grouped into patterns or averages. This results in either being represented in a very small dataset or in a single data point within a data sample.[22] This means that unique data sets are not accurately represented in data distributions as their data falls outside the dominant cluster AI systems are trained to pursue. The statistical reasoning[23] produced by AI will either inaccurately represent outlier populations or ignore the dataset completely since it diverges from the target optima. As a result, people with disabilities have been disproportionately harmed by generative AI tools and systems.[24]

As countries race to regulate AI[25] adequate representation and advocacy is becoming increasingly important to ensure AI standards and regulations are protecting those who experience the most risk. Without adequate representation of people with disabilities or their active participation in the development of regulatory standards, the AI standards and regulations will not adequately protect these communities, thus continuing the harm towards these communities at a systemic level.

International standards seem to recognize some harm in AI systems, such as the US Equal Employment Opportunity Commission (EEOC), who offers guidance on hiring with AI and the implications towards Americans with Disabilities Act (ADA)[26] However, there are currently no standards or regulations that address statistical discrimination or cumulative harm. When looking at the number of people negatively impacted, the numbers may be small compared to the dominant average, however, when ignored, the

[20] (Silvers, A., & Francis, L. P., 2005)
[21] (*Misconceptions about disability,* 2024)
[22] (Treviranus, 2020)
[23] Statistical reasoning is a method of how AI makes decisions
[24] (Noone, 2021)
[25] (Smuha, 2021)
[26] EEOC: https://www.eeoc.gov/

outlier groups continue to be excluded, thus continuing the cycle of discrimination and exclusion towards outlier populations.

1.3 Seed Standards and Capacity Building Support Group

Jutta Treviranus and the Inclusive Design Research Centre ("IDRC")[27] have produced a framework of AI standards that is in line with Accessible Canada Act's AI regulatory priorities.[28] Known as the Seed standards, each clause within the document framework highlights the technical specifications and guidance found in reviewed standards, codes, guidelines, and academic research in effort to ensure that people with disabilities can participate fully in the design, implementation, procurement, and feedback loop of generative AI systems.[29] The document framework proposes regulatory standards meant to treat people with disabilities equitably in decisions made or guided by AI.

The Seed standards is a proposed standards document that focuses on areas where people with disabilities may face barriers in the accessible participation in the design, development, and use of, and the equitable treatment by AI systems. The clauses in proposed in the Seed Standards document will supplement and enhance general guidance and directives that support equitable practices when implementing AI at the Federal level. The standard will also provide proactive guidance on how to prevent harm towards people with disabilities in emerging applications of AI. The following is a list sections the Seed standard document addresses and provides guidance in. This MRP focuses on the Equitable AI and Organizational Process sections:[30]

1. Accessible AI
    a. People with disabilities as full participants in AI creation and deployment
    b. People with disabilities as users of AI systems
2. Equitable AI
    a. Statistical discrimination
    b. Reliability, Accuracy, and Trustworthiness
    c. Freedom from negative bias
    d. Equity of decisions and outcome

[27] https://idrc.ocadu.ca/
[28] SEED Document- https://idrc.ocadu.ca/
[29] The design, implementation, procurement, and feedback loop is also known as the artificial intelligence life-cycle
[30] SEED Document- https://idrc.ocadu.ca/

    e. Safety, Security and Protection from Data Abuse
    f. Freedom from Surveillance
    g. Freedom from Discriminatory Profiling
    h. Freedom from misinformation and manipulation
    i. Transparency, Reproducibility and Traceability
    j. Accountability
    k. Individual agency, informed consent and choice
    l. Support of human control and oversight
    m. Cumulative Harms
3. Organizational Processes to Support Accessible and Equitable AI
    a. Plan and justify the use of AI systems
    b. Design, develop, procure, and/or customize AI systems that are accessible and equitable
    c. Conduct ongoing impact assessments, ethics oversight and monitoring of potential harms
    d. Train Personnel in Accessible and Equitable AI
    e. Provide transparency, accountability, and consent mechanisms
    f. Provide access to equivalent alternative approaches
    g. Handle feedback, complaints, redress, and appeals mechanisms
    h. Review, refinement, halting and termination mechanisms
4. Accessible Education and Training
    a. Training and Education in AI
    b. Training and Education in Accessible and Equitable AI
    c. Accessibility and Equity Feedback, Complaints and Data about Harms applied to improve AI Systems

This framework has been designed to add a considerable amount of context and information to further enhance the rationale behind each standard within the framework. However, it requires a background of digital literacy, technical expertise, and legal expertise as it utilizes technical and legal language specific to AI systems design. Since the framework is designed to protect and support people with disabilities, it requires active participation, feedback, and involvement from committee members representing and advocating for people with disabilities. There is a capacity gap between the intended users being able to provide feedback and utilize the Seed Standards document as it requires intended users to have the expertise to interpret the information and actively participate in the regulatory discussion. To address this capacity gap, a Capacity Building Support Group was created to support the greater group who proposed the Seed Standards.

Leveraging the Seed Standards produced by the IDRC team, this inclusive design project enables individuals most impacted by AI risks to understand the risks, opportunities, and the decisions that need to be made with respect to the design of AI standards. This inclusive design project was developed in collaboration with the Capacity Building Support Group[31] in efforts to bridge AI knowledge and accessibility gaps and enable individuals most impacted by AI risks to advocate for necessary protections at the Federal level. This inclusive design MRP contributed to the Capacity Building Support Group through the design of inclusive resources to support the standards under the Equitable AI and Organizational Processes sections.

1.4 Capacity Resource Users- Committee Members

The intended users of this capacity building resource are individuals most at risk of being harmed by AI systems and experience technical, legal, experience and other knowledge barriers required to appropriately advocate for regulatory protections in an informed way. More specifically, the capacity building resource is meant for committee members who are invited to participate in regulatory discussions pertaining to AI standards in Canada and require information on the technical terms, current legal standards and frameworks in place, and AI processes and tooling.

2. DESIGN GOAL

This inclusive design research project seeks to bridge the capacity gap between the Seed Standards and its intended users, AI standards committee members[32] representing and advocating for people with disabilities. The goal of this inclusive design research project is to provide individuals most impacted by AI risks and have the most as stake with regulatory standards with clearlanguage[33] resources to understand the risks and opportunities of AI so that they are equipped to provide feedback on the decisions that need to be made with respect to the design of the clauses in the Seed Standards.

[31] (Liskovoi et al., 2024)
[32] Committee members are known as individuals representing and advocating for the needs of people with disabilities in regulatory meetings to ensure regulatory standards are informed by people with disabilities and protect against harm.
[33] Clearlanguage resources can also be seen as non-technical resources

Before regulatory standards are set, the Canadian government will likely consult key stakeholders impacted by the proposed regulation. The consulted stakeholders typically include technical experts in the subject of the standards, business entities, academics, and other government entities. These committees, however, rarely include the subject of the standards[34] and typically are the ones advocating for the least amount of regulation in order to minimize the change required on their part. As an example, to support the development of temporary AI standards and regulations, the Government of Canada hosted roundtable discussions to seek stakeholder feedback on a Canadian code of practice for generative AI.[35] These roundtable discussions gathered feedback from 92 stakeholder groups and involved representatives with expertise and experience in generative AI, including Canada's Advisory Council on Artificial Intelligence and Canada's AI research institutes and industry. Business entities, academic institutions, regulatory committees, and government representatives were named in the stakeholder participant list.[36] The summary document of the meeting outlined the themes and debates of the discussion, which mainly centered around the barriers of regulation, particularly towards businesses. Representation, however, from vulnerable groups, such as people with disabilities and other outlier communities impacted by generative AI systems, was missing from the list. This matters because the outcome of the consultative process produced Canada's temporary code of conduct for generative AI practices, which did not consider the voices of those most impacted by the risks and implications of AI.

Designing inclusive AI standards requires the feedback and input of those impacted by the system. It is not uncommon for the voices and testimonies of vulnerable groups to be missed; however, the inclusive design framework [37] requires the active participation of those who are likely to be most harmed in the design process. To support the current gap in the regulatory process of generative AI in Canada, committee members

[34] SEED Document- https://idrc.ocadu.ca/
[35] (*Consultation on the development of a Canadian code of practice for generative artificial intelligence systems* 2023)
[36] (*What We Heard – Consultation on the development of a Canadian code of practice for generative artificial intelligence systems, 2023)*
[37] (*Welcome to the Inclusive Design Guide,* 2016)

representing and advocating for people with disabilities must feel equipped to discuss the risks, implications, and opportunities of AI. ensuring the systems are designed with the subjects impacted most by the system. Unfortunately, these individuals also are the most underserved when it comes to receiving the digital, technical, financial, and legal expertise required to understand and advocate for equity in the development of AI systems. There is a capacity gap between the committee members' expertise needed to advocate for better protections, and the resources that other entities have access to. There is a need to bridge this capacity gap to level the playing field amongst committee members so that the groups most impacted can accurately advocate for their communities amongst businesses and other entities who desire the least amount of regulation.

In a recent publication, Smuha[38] highlights the current race to AI regulation to ensure the systems are trustworthy. The positive outlook is that international regulators are paying attention to the impacts of AI. However, without the accurate and robust representation of those most impacted by AI systems, the protections and standards in place will not apply or not be enough to protect the most vulnerable to risk and harm. This inclusive design research project seeks to bridge the expertise and knowledge gaps amongst committee members who need the information to contribute to more informed discussion. Through resource design, this inclusive design and research project seeks to enable committee members to be better informed of the risks, implications, and opportunities of AI towards people with disabilities.

3. ENVIRONMENT SCAN

AI risk management and regulation is a topic in flux and heightened by the advances in generative AI technology and recent advancements made in LLMs. Prior to the development of the Seed Standards document and corresponding Capacity Building Support Group, a broad literature review and environmental scan was conducted with a focus on reviewing relevant material pertaining to disability and the broader area of AI ethics, and potential harms and opportunities of AI. Given the pace of AI development

[38] (Smuha, 2021)

and adoption, gray literature, temporary standards, websites, news articles and unpublished research was reviewed and leveraged within the capacity building support resources.

3.1 Canadian Standards

As of 2023, Canada's *Artificial Intelligence and Data Act* does not regulate the development, procurement, use, or management of AI in Canada.[39] To address the broad risk profiles of advanced generative AI systems, Canada's Minister of Innovation, Science and Industry announced a voluntary code of conduct,[40] which has been signed by 23 Canadian businesses and entities. This code temporarily provides Canadian entities with common standards and enables them to demonstrate, voluntarily, that they are developing and using generative AI systems responsibly until formal regulation is in effect. The code is based on expert feedback received during a consultation process [41] on the development of a Canadian code of practice for generative AI systems. The temporary code of conduct is aimed to help strengthen Canadians' confidence in the AI systems as the government works towards formalizing its regulatory standards. However, the consultation process did not have representation from disability advocacy or civil groups, and other marginalized civil groups. In addition, the representatives within the consultative groups critiqued that regulating generative AI was too broad of a subject with opposing views. For example, some academics stressed the importance of data and development transparency, as well as the disclosure of datasets and training methods. In contrast, others expressed concerns over litigation risks if transparency is mandated.[42] The outcome of the consultative process resulted in temporary standards to guide Canadian entities in utilizing generative AI, however, the code itself does not identify who is most at risk nor does it address risk prevention measures to protect vulnerable groups.

---

[39] (*Artificial Intelligence and Data Act, 2023)*
[40] *(Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems,* 2024)
[41] (*Consultation on the development of a Canadian code of practice for generative artificial intelligence systems,* 2023)
[42] (*What We Heard – Consultation on the development of a Canadian code of practice for generative artificial intelligence systems, 2023)*

In 2017, the International Organization for Standardization (ISO) published an AI and ML framework for describing a generic AI system using ML technology.[43] The framework describes the system components and their functions in the AI ecosystem, and is applicable to all types and sizes of organizations, including public and private companies, government entities, and not-for-profit organizations, that are implementing or using AI systems.[44] Canada is an ISO member and has contributed to the development of the standard. This framework, however, has a digital literacy barrier as the document is in technical terms and requires the interpretation of a subject matter expert. In addition, there is also a financial barrier as the framework is only made available through purchase. These barriers make this framework inaccessible for individuals seeking to inform themselves of the Canadian standards and protections.

Unlike other standardized *Accessible Canada Ac*t topics, it is evident from this environmental scan that there are no existing Canadian AI accessibility standards to reference, and the existing equity and AI ethics standards or legislation do not address equitable treatment of people with disabilities other than to mention people with disabilities may be harmed and should be considered.

3.2 International Standards

In developing the Seed Standards, a broad international jurisdictional scan was conducted to identify both proposed and adopted standards, policy and legislation related to artificial intelligence (AI) and machine learning. As artificial intelligence is a form of information and communication technology (ICT), relevant guidance and legislation regarding information and communication accessibility was also included in the scan. Most AI standards are behind paywalls, such as the ISO JTC1 SC42 standards document. In referencing standards and guidance, the Seed Standards group has favoured openly available standards and guidance over standards with paywalls where possible.[45] The Information Security Management Systems Requirements (IEC) for the International Organization for Standardization (ISO) has been abided by where

[43] (ISO JTC1 SC42, 2017)
[44] (CSA Group., 2022)
[45] SEED Document- https://idrc.ocadu.ca/

applicable.[46] The ISO standards used in the Seed Standards document have been further contextualized in the capacity building support resources.

Similarly seen in the environmental scan conducted for Canadian AI standards, international equity, and ethical AI standards or legislation do not address equitable treatment of people with disabilities other than to mention people with disabilities may be harmed and should be considered.

4. INCLUSIVE DESIGN FRAMEWORK

4.1 Secondary Research

Secondary research, such as blog posts, online news articles, academic papers, legislation, the AI incident database,[47] and the Seed Standards document were leveraged to provide further insight into the real-world impacts and harm of AI systems today and what needs to be considered to prevent further harm. No human subject research or engagement was done as part of this inclusive design MRP.

The primary research conducted by the Seed Standards group included virtual co-design sessions conducted with people with disabilities and other experts as part of the development of the Seed Standards. In addition to the virtual co-design sessions, interviews with leading technical, policy and ethics experts were conducted. Meetings with national and international committees focused on the topic, including with the US Access Board, the US National Institute for Standards in Technology, the European Disability Forum, the World Economic Forum, the Global Leadership Alliance, the World Wide Web Consortium, the World International Property Organization, UNESCO,

[46] ISO/IEC 27001 is the world's best-known standard for information security management systems (ISMS). It defines requirements an ISMS must meet. The ISO/IEC 27001 standard provides companies of any size and from all sectors of activity with guidance for establishing, implementing, maintaining and continually improving an information security management system. Conformity with ISO/IEC 27001 means that an organization or business has put in place a system to manage risks related to the security of data owned or handled by the company, and that this system respects all the best practices and principles enshrined in this International Standard.

[47] AI Incident Database: https://incidentdatabase.ai/

G3ICT, the ISO/IEC JTC1SC42 committee tasked with AI standardization and others.[48] Organizations and initiatives that address AI equity were consulted. This research was leveraged within the capacity support resources to support further contextualization of the standards and harm reduction.

4.2 The Three Dimensions of Inclusive Design

The Inclusive Design Research Centre developed a guiding framework for practicing inclusive design. The framework has the following three dimensions:[49]

1. Recognize, respect, and design with human uniqueness and variability.
2. Use inclusive, open & transparent processes, and co-design with people who have a diversity of perspectives, including people that can't use or have difficulty using the current designs.
3. Realize that you are designing in a complex adaptive system.

Leveraging the inclusive design framework, the capacity support resources aligned to each of the three dimensions to address the complexity and adaptiveness of AI systems and produce a resource that can continuously be contributed to and evolved along with the advancements in AI.

The first dimension of the framework is to recognize the uniqueness of each individual.[50] As individuals, we are all complex and unique beings with our own specific characteristics and traits. This capacity resource offers a spectrum of personal choices for users to understand each clause. The resource outlines a plan language summary, then a rationale, and then impacts and case studies to further highlight risks and preventative opportunities of the clause. By offering multiple methods of building context, the user can determine how much information they need to understand the clause. This dimension addresses the importance of designing for uniqueness and complexities. Designing for the individual may be considered impossible when attempting to reach a wide audience, however, when involving impacted individuals as part of the design process and designing for uniquenes, the designs themselves

[48] SEED Document- https://idrc.ocadu.ca/
[49] (Treviranus, 2021)
[50] (Treviranus, 2018a)

account for diverse complexities that can still be shared with large groups of people. For example, this capacity resource gap is a resource for people with disabilities, however, the implications can span across other minority groups as many of the case studies and impact reports consider intersectional identities.

The second dimension of the framework is to use inclusive, open and transparent processes, and co-design with people who have a diversity of perspectives, including people that cannot use or have difficulty using the current designs.[51] When designing for an area of opportunities, traditional researchers and design methodologies take on the role of designing for the "problem." However, when designing "for," researchers and designers take on a position of power, extracting information and details for groups of people they are meant to design for. However, without designing with the participants, the researcher and designer will not be able to accurately design "for" as they will not know what was missed or what was not considered. Inclusive design takes the approach of designing "with." Rather than assuming a position of power, inclusive design practitioners co-design and co-create with groups of people to support them in designing what they need. Referring to the first design dimension, to design for uniqueness, the unique voices must be considered as part of the design process. By leveraging the co-designs of the Seed Standards group and my own lived experiences as a non-technical and non-expert individual, this capacity resource can be used by committee members to participate fully in the process of designing the standard and making informed decisions.

The third dimension of the framework is to realize that you are designing in a complex adaptive system.[52] Design cannot be stagnant as the environment around us is in constant flux and change. Stretching the responsiveness and adaptability of the designs we live with supports a diversity of human knowledge, skills, and perspectives. This means that design must live within the complex adaptive system it participates within. In other words, design cannot be developed and operated in a silo; it must be able to live on in a continuous iteration. This capacity resource has contextualized the standards

[51] (Treviranus, 2018b)
[52] (Treviranus, 2018c)

and linked it to contextual circumstances, preparing committee participants to understand the possible risks, impacts, and opportunities. This capacity gap can live on and continue to be maintained as standards evolve.

5. CAPACITY SUPPORT RESOURCES

Each clause in the Seed Standards document builds off one another, where the standard is intended to protect against AI systems causing undue harm and risk. As someone with a non-technical background, I reviewed each clause to see if I was able to understand the terms without further research. If further research was required, I simplified the technical terms and broke down complex ideas into clear sections to aid committee understanding. To further build on each clause and provide more insight into the need for each clause, I developed a rationale for each section. The rationale is meant to bridge the expertise and knowledge gap, where it outlines the cause and effect of each clause and justifies why the standard has been put in place. The rationale gives committee members insight into the inner workings of AI systems and its implications. To support the committee's understanding of the practical application of the clause, I included examples under each clause to highlight the impacts of AI systems and the further harm caused if the clause is not considered. I included case studies specific to impacts on people with disabilities where possible, however, applicable examples and case studies were sometimes limited. In the instances where I was unable to source specific examples pertaining to people with disabilities, I leveraged case studies that impacted other marginalized populations as these communities are also considered data outliers. To specify, data outliers should not be grouped into a singular category, however, when considering risk and harm, outliers experience more risk and more harm than the dominant average. So, when sourcing examples of impact, harm against marginalized populations will be more severe when compared to the dominant group. The implications of harm can be shared amongst outlier groups, where people with disabilities face the most harm as they are on the furthest edge.

Referring to my design goal, this capacity support resource is meant to be a tool to bridge the technical and knowledge gaps of how AI systems are designed, developed, and implemented so that committee members can actively contribute to regulatory

discussions and advocate for protections. Sections 5.1 to 5.23 outline the specific design decisions made for each clause and how it bridges the current capacity gaps for committee members.

5.1 Statistical Discrimination

This clause aims to protect against statistical discrimination within AI systems by regulating the monitoring of the AI system to determine and measure the performance and impact of AI on decisions for deviations from the statistical averages. This Seed clause outlines the technical terms of the clause and its recommendations first, and then offers further context into what other international standards have considered. An individual with technical knowledge or expertise likely would have understood the clause, however, to bridge the capacity gap, I simplified language and used examples to highlight the implications. This clause was summarized into clear language by reorganizing the information to focus on what the clause is aiming to protect against and if there are protections in place currently. The following paragraphs outline the clear language summary of the clause:

> The “outlier problem” refers to the difficulty of representing individuals who are beyond the threshold of standard deviation. As outliers of data, these individuals within an AI system’s predictions are excluded or are wrongfully represented. This algorithmic exclusion is also known as statistical discrimination and affects people with disabilities as they diverge from the statistical mean.
>
> There are currently no standards or regulations that address statistical discrimination. The US Employment Equal Opportunity Commission 2022 recommended removing disability related data or characteristics that lead to statistical deviation from the target optima in AI hiring systems. However, this is unlikely to result in a data match as data profiles are multi-faceted and complex, one element will influence the other remaining elements.

To address the impact, I designed a rationale to further contextualize and clarify what the clause is aiming to prevent and protect against. The following rationale is designed to address the specific harm the clause is meant to mitigate:

> To address statistical discrimination, systems must be monitored consistently to determine and measure the performance and impact of AI on decisions for deviations of

statistical averages. Impact assessments should be used to monitor the full range of impact.

To gather data on the full range of experience, AI tooling should consider an inverse approach. For example, in hiring, AI tooling could use exploration algorithms instead of pattern matching. For social media, AI tools could use inverted metrics to highlight novel contributions rather than the most popular ones. When risk assessments are conducted, edge cases should receive priority analysis and solutioning.

Due to their complex and unique profiles, people with disabilities are often outliers in datasets as their profiles are varied and unique, with complex characteristics. Their profiles do not align with the majority or even with each other. When AI is trained to recognize patterns through statistical reasoning, it will disregard and discriminate against any outliers. As AI continues to be adopted at scale, statistical discrimination against outliers becomes a systemic issue, becoming more difficult to catch and rectify as the scale is much broader.

In effort to further highlight the need for the clause, the following case study was chosen to reflect the biases of how AI tools are trained to be optimized for the statistical average. This case was found through the AI incidents database and is meant to highlight the risks in utilizing AI to replace human decision making due to the historical data used to train AI. The conclusion offers further context into how this case connects to the clause and the impacts towards people with disabilities:

People with disabilities are considered data outliers as they tend to fall on the margins of datasets. Statistical reasoning seeks to find the data averages and overlooks data outliers in pursuit of the greater average. When statistical reasoning is used in a decision system, such as AI, the system will decide against data outliers, regardless of the quality of data used. AI systems amplify and accelerate data generalization, thus systematically discriminating against people with disabilities.

AI is trained to optimize for the future by pulling data from the past, which is then used to create the target optima, also known as the ideal persona. In pursuit of the target optima, outliers are excluded as they do not fit into a generalized pattern. As the AI tool continues to evolve and learn, it seeks to become more accurate in aligning to the target optima, which further discriminates against outliers such as people with disabilities.
For example: [53]

[53] (*Risks of bias and discrimination in AI hiring tools,* 2013*)*

- AI used to analyze facial expressions in video interview tools have been ineffective with speech patterns, facial movements, skin color, and even when human subjects wear glasses, accessories, or head coverings.
- When training the dataset, a group of researchers used images of cancerous specimens in hopes that the AI would diagnose tumors from the images. The AI had learned to identify rulers, rather than diagnose tumors. When they analyzed the data, most of the images of cancerous specimens had included rulers beside them for scale. The AI model picked up on the rulers as they were consistent across all the images whereas the tumors from the images had unique characteristics.

The examples presented above demonstrate how AI tools either produce negative results or exclude the data point entirely, when deviating from the target optima. People with disabilities deviate from the status quo, so when the tool is trained to pursue the greater average, it will discriminate against outliers.
This case provides an example of how statistical reasoning used in AI contributes to the systemic discrimination against people with disabilities, and why AI tooling requires scrutiny to mitigate risks related to bias, discrimination, privacy, and ethics.[54]

5.2 Reliability, Accuracy, and Trustworthiness

This clause aims to inform committee members of the inner workings of how AI systems statistically discriminate against people with disabilities. Building upon the previous clause, this clause discusses how AI is trained to produce discriminatory results against data outliers. Due to the technical complexities of this clause, I clarified the technical terms and offered insight into what the clause is protecting against:

Accuracy is meant to ensure the outputs of AI systems remain close to the values accepted as being true, where measurements should be paired with clearly defined and realistic test sets, and details about test methodology should be included in associated documentation. Accuracy measurements must be detailed, specific, and may include disaggregation of results for different data segments. Accuracy is required when developing results that are trustworthy and reliable, however, the pursuit of accuracy within statistical reasoning models correlates to greater specificity towards the target optima.
Robustness looks at the system's ability to maintain performance under a variety of circumstances.[55] Achieving robustness is the goal for AI systems as it is meant to test system functionality under a broad set of conditions, anticipated and unanticipated, and ensure the system performs as expected, as well as minimize harm if operating in unexpected settings.

[54] (*Civil rights principles for hiring assessment technologies,* 2023*)*
[55] (*International Standards Organization,* 2022)

> To ensure accuracy and robustness work together to provide trustworthy results, the AI systems must be designed to guard against the pursuit of accuracy of the target optima as it leads to falsely rejecting people with disabilities due to either:
>
> - over-trust in an otherwise accurate and reliable system,
> - ignoring failures as anecdotal and affecting only a small minority.
>
> Accuracy measurements should include disaggregated accuracy results for people with disabilities. They should also consider the context of use and the conditions relative to people with disabilities.

I designed the rationale section to breakdown the risks associated with AI systems pursuing accuracy towards the target optima. The goal for this rationale is to provide committee members with necessary context in understanding the harm and impact this form of validation can have on data outliers in a way that does not rely on previous technical expertise:

> Accuracy and robustness contribute to the validity and trustworthiness of AI systems; however, they also can be in tension with one another;[56] the more accurate the system is, the more specific it is towards finding the target optima. Since people with disabilities have identities that go beyond a single characteristic, they are outliers when compared to the target optima. The pursuit of greater accuracy within statistical reasoning models can lead to falsely rejecting people with disabilities as they are considered data outliers. When the outliers don't match with the general dataset, they are either flagged, or produce an inaccurate result.[57]
>
> Automation bias[58] is an outcome of the over-trust towards AI systems. Automation bias is the predilection by humans to favour decisions by an automated process in the face of contradictory evidence even when that evidence is correct. This kind of bias leads to over-trust in a system to the point that people assume its outputs are always correct and become lax when monitoring the system. Failures and incorrect predictions are missed as a result, which can negatively influence responses to complaints raised by people with disabilities as the complaints go against the data majority. As the system is perceived as reliable, the complaints get dismissed as anecdotal since they are rare and felt only by a small minority. However, if the minority group was inaccurately represented to begin with, the tool itself is not accurately representing a subset of the population (people with disabilities).

I included two case studies to further depict the risks and implications of the pursuit of

[56] (*ISO/IEC TS 5723: Trustworthiness Vocabulary*, 2022)
[57] (Walch., 2023)
[58] (*ISO/IEC TR 24028: Overview of trustworthiness in artificial intelligence, 2020*)

accuracy. The first case study was sourced from the AI incidents database and sheds light to the unintended and unexplainable fluctuations seen in the results produced by ChatGPT, which is a LLM. This case study provides an example of how LLMs change, even when they have been initially proven to be correct, thus highlighting the inaccuracies present in AI systems. The second case study highlights the negative biases of AI hiring systems towards people with disabilities and other minority groups. This case study provides a practical example of the systemic implications of pursuing the target optima and the cumulative harm towards vulnerable populations. Through this example, committee members will be able to understand the real-world impact of AI systems and the risks associated when pursuing the target optima.

Case Study 1:[59]

A recent study compared the performance of ChatGPT's abilities to complete a range of diverse tasks. The study was conducted over several months to continue measuring the chatbot's ability to solve math problems, answer sensitive questions, generate software code, and provide visual reasoning.

Researchers found dramatic fluctuations, known as drift, in the chatbot's ability to perform the above tasks. The study compared GPT 3.5 from GPT 4 and researchers found that GPT 4 was able to correctly identify a prime number with 97.6% accuracy at the start of the test, however, its accuracy fell to 2.4% three months later. In contrast, GPT 3.5 answered the same question with 7.4% accuracy at the start of the test and improved its response accuracy to 86.8% in June. This led researchers to realize the unintended consequences when tuning large language models due to its many interdependencies. Unfortunately, these side effects are poorly understood by researchers and the public as the models powering ChatGPT remain private, also known as black-boxed. In addition to populating incorrect numbers, the model failed to show how it came to its conclusions, making it more difficult to understand the "thought" process of the algorithm.

The key takeaway from this research comparison was that large language model drifts occur and it's imperative to continue monitoring the models' performance overtime. Although this example does not explicitly address disability, the changes in the accuracy of large language models over time can cause significant harm to people with disabilities. These models are being adopted by regulators, recently seen in the US[60] and integrated into social services that support vulnerable groups of people. Language models using statistical reasoning may experience drifts over time, resulting in unintended consequences, likely negatively impacting the most vulnerable members of

[59] (Confino, 2023)

[60] (Kang, 2024)

society. This example demonstrates ChatGPT experiencing a drift and producing inaccurate results on a math equation, however, if the same technology was being used to produce results for disability benefits, and it experienced a drift without human oversight, the outcome could cause undue harm towards people requiring the social benefits.

The models need to guard against producing over-trust and ignoring failures deemed as anecdotal by providing transparency into its decision-making capabilities and the datasets used.

Case Study 2:[61]
HireVue, a recruiting-technology firm, has designed a system that uses candidates' computer or cellphone cameras to analyze their facial movements, word choice and speaking voice before ranking them against other applicants based on an automatically generated employability and productivity score. HireVue's AI driven assessments have been adopted by more than 100 employers, including Hilton and Unilever, and more than a million job seekers have been analyzed. However, AI researchers have critiqued HireVue's algorithm as the statistical reasoning is not rooted in scientific fact nor does it consider a diverse range of ability. To train the tool, historical data on what a "employable" candidate looks like is used to determine the target optima. These researchers argue that in analyzing a human being like this could end up penalizing nonnative speakers, visibly nervous interviewees, or anyone else who doesn't fit the target optima for look and speech. This has direct implications on an individual's career as the system's pursuit for the target optima will discriminate against data outliers and impact the prospects of potential candidates who are different from the average. The system does not provide a transparent methodology behind its decision-making process, thus leaving candidates and employers in the dark as to why a candidate was rejected.

This case provides an example of how automated hiring practices may exasperate biases at a large scale. An example of a negative impact towards data outliers can be seen with Amazon's AI recruitment tool, which discriminated against women as the tool was trained on sourcing ideal candidates for the company based on historical candidate data, which mostly consisted of men.[62] People with disabilities are further discriminated against under automated systems like HireVue as they are the furthest data outliers, and their unique and complex characteristics cannot be generalized.

[61] (Harwell, 2019)
[62] (Oppenheim, 2018)

5.3 Freedom from negative bias

This clause is designed to protect against negative biases by focusing on the quality and diversity of the datasets utilized. To bridge the gaps of this clause, I summarized the clause to focus on how human bias leads to biased AI systems. Since this clause builds on the previous, we know AI systems produce biased results due to its pursuit of the target optima; this clause aims to highlight how an AI system can be biased in its development. I then offered definitions to describe each bias in more detail to help build technical literacy and further context into the harms and risks:

> The discriminatory biases of human designers and developers, digital disparities, or discriminatory content in training data can all lead to negatively biased AI systems affecting people from gender, race, age, ability and other marginalized minorities. AI systems will amplify, accelerate, and automate these biases in its pursuit of the target optima.

Data bias

Non-representative sampling is a form of data bias that occurs when the underlying population is not fully represented. Polluted data refers to discriminatory data or stereotypes in the training data. This is likely to occur in large data sets of text or web content.Confounding variables are extraneous factors that influence one or more of the variables used in training data and that result in spurious correlations.[63]

Data proxy bias

Bias due to proxy variables can be present in training data but techniques have been developed that detect and mitigate against its effects using external validity testing and making adjustments to the data itself, or altering the weights within the machine learning model, or modifying the outputs of the system.

Label bias

Bias in labeling can result from cognitive biases[64] as human developers are deciding the appropriate labels. Without engaging a diverse group of people with disabilities, the labeling criteria will naturally miss or contain stereotypical assumptions.

Algorithmic bias

[63] (*ISO/IEC TR 24027: Bias in AI systems and AI aided decision making,* 2021).
[64] (*ISO/IEC TR 24027: 6.3.3 Bias in AI systems and AI aided decision making,* 2021)

> Algorithmic bias can come about from engineering decisions[65], choice of algorithm,[66] feature bias[67] and hyperparameter tuning.[68] In some cases, the bias is due to sparse data for groups that have a smaller population.

To further highlight the impact and build awareness and context, I focused the rationale on the complex and unique characteristics of people with disabilities and why it is difficult for AI systems to register multiple data points within a single input. This rationale is meant to demonstrate the limiting capabilities of AI as it is designed today and why AI needs to be designed differently to better protect against harm:

> When it comes to the data, "persons with a disability" is not a single measurable variable. It is a category that aggregates people with disparate capabilities such as poor vision or who are blind, persons with hearing loss, mobility impairments, cognitive differences, and so on. It is likely that "persons with a disability" itself is a confounding data variable since having a disability can influence many other variables in the dataset. Avoiding this type of bias is complicated given "persons with a disability" is not a simple single dimensional variable. Proxy variables are used by human trainers to fill in the data gaps, however, detecting data proxy bias can be a complex problem even in the case of a simple variable such as age, as described in *How discrimination occurs in data analytics and machine learning: Proxy variables*.[69] Many seemingly neutral variables are, in fact, associated with age, including:
>
> - Given name: the popularity of first names changes from generation to generation and tend to clump together at different points in time.
> - Address: Where a person lives depends on their age. Younger individuals tend to rent apartments, somewhat older individuals have their own house, and still older people tend to live in retirement homes.
> - Smoking: smoking was popular in the past, and older people are more likely to smoke than younger people.
> - Vegetarianism: a relatively more recent diet choice.
>
> Removing age from the data is not enough to remove age bias because of its relationship with these proxies. More importantly, it is not clear how to find and adjust for all these associations. The paper cited above suggests finding all the correlated variables and making small adjustments to the dataset so that it is impossible to infer a given characteristic (age) from the rest of the data. The proxy variables are no longer proxies after the adjustments have been made.

[65] (*ISO/IEC TR 24027: 6.4 Bias in AI systems and AI aided decision making,* 2021)
[66] (*ISO/IEC TR 24027: 6.43 Bias in AI systems and AI aided decision making,* 2021)
[67] (*ISO/IEC TR 24027: 6.4.2 Bias in AI systems and AI aided decision making,* 2021)
[68] (*ISO/IEC TR 24027: 6.4.4 Bias in AI systems and AI aided decision making,* 2021)
[69] (Cevora, 2020)

Age is considered a singular measurable variable, however, “persons with disabilities” contains an assortment and can often be unrelated to one another. There are numerous proxy variables that are obscure and difficult to determine *a priori*. Proxy variables can often hide the impossibility or inappropriateness of certain measurements and replace them with others that are associated with stereotypes or negative biases against people with disabilities (e.g. using BMI or body-normative fitness measures as a proxy for measuring health and wellbeing). External validity testing might help where the prediction or output of the AI/ML can be compared against what should occur with respect to persons with disabilities, but it is unclear if all bias can be removed.
The most difficult algorithmic issue to address with respect to persons with a disability is that AI algorithms seek the optimum, or target optima, given the data they are trained on. Due to statistical discrimination, persons with disabilities are far from an optimum and the algorithm marginalizes them even more. They are thereby excluded from AI predictions, or the algorithms then designed to discriminate against unrecognized individuals or individuals that deviate from the optimum.

A technique to compensate for sparse data and avoid using proxies is to use transfer learning.[70] First, an AI/ML system is developed with a large dataset. Then the model of this system is used with the smaller set of data to adjust the larger model’s predictions towards the members of the smaller set. However, due to the complex nature of “person with a disability” noted in the previous section presents barriers as it is difficult to generalize the characteristics of unique data points.

Techniques have been developed that attempt to mitigate bias or create AI systems that are not discriminatory,[71] however, it is unclear that these techniques work in addressing bias against people with disabilities.

The following case study was found through the AI incidents database and was included to help committee members understand the implications of inherently biased AI systems and how it can contribute to cumulative harm:[72]

Mighty Well, an adaptive clothing company that makes fashionable gear for people with disabilities placed a Facebook advertisement for one of its most popular times. The zip-up hoodie with the slogan “I am immunocompromised — Please give me space” was wrongfully flagged by Facebook’s automated advertising center and rejected for violating policy. Specifically, Facebook’s policy around promotion of “medical and health care products and services including medical devices,” although no such promotion was made in the advertisement. In this instance, Facebook’s algorithm may have been biased against disability-related terms, which resulted in the platform flagging the anomaly. Mighty Well appealed the decision and got the ad published, however, it raised a bigger question around the biases in Facebook’s advertising algorithm and how the algorithm qualifies the target optima for advertisements. Unfortunately, other adaptive

[70] (*ISO/IEC TR 24027:* 8.3.3.2 *Bias in AI systems and AI aided decision making,* 2021)
[71] (Trewin, 2018)
[72] (Friedman, 2021)

clothing brands experience the same as Mighty Well and need to appeal each item individually. For some stores, this means appealing hundreds of items individually, placing undue strain on these small businesses.
Businesses often need to utilize Facebook advertisements to grow and reach their audiences. However, Facebook's advertising algorithm is unable to recognize complex characterizations, or make connections based on context. In this example, the algorithm flagged the medical term and was unable to contextualize the term "immunocompromised" beyond the characteristic of "medical service or device," This disadvantages people with disabilities on a global scale as Facebook's algorithm wrongfully flags a medical term used in the disability community. In the case of Mighty Well, the online storefront ended up resolving the flag only after the company appealed each wrongful flag made by the algorithm, and as the case indicates, this is not an isolated incident.

This case presents an example of how negative biases within the algorithm can impact accessibility of products geared towards people with disabilities, causing further discrimination against and isolation of people with disabilities.

5.4 Equity of decisions and outcome

Building on the previous, this Seed clause aims to protect against biased AI systems. In an ideal state, AI systems are designed with all users in mind; the reality, however, is that not all AI systems will follow the same standards. In addition, even the most equitable systems are subject to change over time, which we saw in case study [173] of section 5.2 of this MRP. This clause acts as a check and balance of AI decisions and relies on committee understanding of the previous clauses. To summarize the clause, I focused on mirroring the language from the previous clauses to maintain consistency:

> Regulated entities utilizing AI systems will monitor the outcomes of AI systems with respect to people with disabilities and use the data to further tune and refine the AI systems. As part of the review, regulated entities will develop systems to explain and monitor the decisions made by AI systems to ensure results are equitable to people with disabilities and reasonable with respect to the decisions made by the tool.

Without proper context or guidance, regulatory standards often lack the information needed to understand how to put a clause into action. To help bridge this gap, I focused the rationale on leveraging AI explainability (XAI) as a method to explain and monitor the decisions made by AI. Since the term is technical, I included an international

[73] (Kang, 2024)

standard of AI explainability and the negative impacts if explainability is not integrated into AI:

> The European GDPR introduced the requirement of AI explainability,[74] where outputs must have a corresponding rationale for the decisions produced by deep learning AI systems. Initially, this requirement was widely critiqued as it did not include techniques to determine rationale, however, the requirement resulted in innovative approaches to explain AI decisions. Despite the growth of explainable artificial intelligence (XAI) techniques, the explanations produced are obscure and often unrelated to meaningful factors. This negatively impacts people with disabilities as AI decisions are often misaligned with the desired outcomes of the tooling. To be useful, the outcomes must consider the complex systems and barriers people with disabilities face, and the information about the outcomes must be used to tune the decision system over time.

The implications of AI decisions can cause immense harm, however, the harm caused by AI is often not apparent until it is too late. Without AI explainability, the AI tool does not offer transparency into decisions, which leads to over-trusting the decisions of the system without actually being able to verify results. This is problematic as harm may be caused unintentionally. AI explainability is fairly technical, so to bridge the gap, I included a case study[75] that demonstrates the harm caused when AI decisions cannot be explained or is not monitored. This case study was found through the AI incidents database:

> In 2015, the American Civil Liberties Union (ACLU) filed a class action lawsuit against the Idaho Department of Health and Welfare. The lawsuit contended that the Idaho Department of Health and Welfare cut Medicaid assistance for adults with developmental disabilities without adequate notification or procedural protections. In 2016, the federal court agreed that the Idaho Department of Health and Welfare "arbitrarily deprived participants of their property rights and hence violated due process." The judge ruled that on Medicaid Act grounds— the act requires explanation for any Medicaid coverage reductions. As part of the ruling, ACLU received the program's data formula and hired experts to review the assessment itself and the data used to create the assessment process. This analysis proved that the data processing used by the state program was not accurate as the system was using a small subset of flawed historical data, the testing process did not produce reliable results, and the structure of the formula used had statistical flaws. The non-representative sampling and algorithmic bias present in the creation of this assessment left over 4000 Idahoans with developmental and intellectual disabilities with inadequate coverage without transparent rationale or adequate notice.
>
> This case offers an example of why data integrity and consistent data monitoring over time is critical to mitigating further discrimination against people with disabilities. This

[74] (*The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence*, 2020)
[75] (Stanley, 2017)

> case highlights how over-trusting the technology and not having transparency into the explainability of the AI model discriminates against its most vulnerable population.

5.5 Safety, Security and Protection from Data Abuse

Technology that leverages personal data is at risk of a data breach. From the previous clauses, we have learned that AI leverages historical data to support its decision making. This means that the data is tied to individual identity. When the data represents a large population, individual identity is anonymized; however, due to the unique traits of people with disabilities, their data points are often singled out, making them vulnerable to data abuse. This clause is aimed to protect against data abuse through prevention. To bridge the capacity gap, the summary section was kept relatively high-level to introduce the topic:

> In the case of data breaches or malicious attacks of AI systems, regulated entities will develop plans to identify risks associated with disabilities and clear and swift actions to protect people with disabilities. Where applicable, regulated entities will monitor to ensure that people with disabilities are not disproportionately flagged on AI systems used to investigate tax fraud, security risks, and other forms of investigation.

To add context, the rationale focused on how unique identities are easily targeted and focusing on harm prevention rather than reacting to data breaches when they happen:

> People with disabilities, especially people with intellectual disabilities and people relying on auditory information or other alternative access systems, are frequently targeted with fraudulent claims and scams. Privacy and confidentiality protections do not work for highly unique populations as they can easily be re-identified due to their unique traits and re-targeted. Differential privacy is used to remove data characteristics to anonymize users and protect privacy, however, these data characteristics are critical in servicing the needs of people with disabilities as they need to be identified to monitor and tune the AI system. In addition, AI systems used to flag security risks, anomalies that require financial audits, tax fraud, insurance risks, or security threats, disproportionately flag people with disabilities because they deviate from the target optima.
>
> People with disabilities are the most vulnerable population to data abuse and misuse as they need to give up their privacy to gain essential services. Rather than removing data

> characteristics, regulated entities can mediate harm by anticipating a data breach and creating plans that encompass prevention and preparedness.

Sourcing a case study that was specific to AI data abuse towards people with disabilities was challenging as the AI incidents and web search did not produce specific results. This is a common occurrence when seeking specific examples of AI impacts on people with disabilities since the impact is not deemed large enough when compared to the general population. However, this is an issue as these outlier groups are left on the edges of society and continue to experience harm that often goes unaddressed. To support the need for this clause, I included a case study[76] from the Netherlands to demonstrate the harm caused by AI systems designed to remove data characteristics rather than having preventative data breach and abuse plans in place. This case study was source from the AI incidents database:

> Since 2013, 26,000 families in the Netherlands were wrongly accused of social benefits fraud partially due to a discriminatory algorithm. After investigation, it was found that the algorithms and automated systems used left little room for accountability or basic human compassion. The automated system seemingly discriminated based on nationality, flagging people with dual nationalities as likely fraudsters.
> This is an example of how black box algorithms can discriminate against a population's most vulnerable. This problem is not unique to the Netherlands; The Australian government faced its own "robodebt" scandal when its automated system wrongfully flagged benefits fraud, clawing back a total of $2.5 billion from welfare recipients.[77] This case, too, came down to a poorly designed algorithm without human oversight.
> Before the increased use of automated systems, the decision to cut off a family from benefits payments would have to go through extensive review. Now, such choices have increasingly been left to algorithms, or algorithms themselves have acted as their own form of review.
>
> For those classified by the automated system as a fraudster, limited follow-up investigations were conducted and victims were unable to gain transparency into the AI's decision-making. Further investigation revealed that the tax office had applied the mathematical Pareto principle to their punishments, assuming without evidence that 80 percent of the parents investigated for fraud were guilty and 20 percent were innocent. While some efforts to increase algorithmic transparency have been made recently, many of the automated systems in use in society remain opaque, even for researchers. Beyond transparency, safeguards and accountability are especially important when

[76] (Geiger, 2021)
[77] (Pearson, 2020)

algorithms are given enormous power over people's livelihoods.

Although this example does not explicitly touch on the impact towards people with disabilities, it offers insight into why regulators need to consider preventative measures for data abuse and data breach instead of assuming what fraud looks like and designing it into the AI system to flag.

5.6 Freedom from Surveillance

This Seed clause was drafted in clear language, so I kept the clause as is:

> Regulated entities will refrain from using AI tools to surveil people with disabilities.

To bridge the gap, I focused the rationale on the harm caused by surveillance measures towards people with disabilities. As mentioned, the clauses relate and refer to one another; this clause may feel similar to the previous one, however, it specifically addresses AI surveillance meant to monitor whereas the previous clause addresses data breaches and data abuse meant to flag:

> AI productivity tools and AI tools that surveil people use metrics that unfairly measure people with disabilities who do things differently. These tools are not only an invasion of privacy but also unfairly judge or assess people with disabilities.

The following case study[78] was sourced from the web and highlights the negative impacts of AI-based surveillance and productivity tools:

> Eight of the 10 largest private U.S. employers use AI to track the productivity metrics of individual workers, many in real time. However, when tracking productivity, the software often flags various disabilities as ‘not-standard” as the results deviate from the target optima.
>
> Due to the lack of transparency in how productivity is measured and assessed, employees with disabilities are at a disadvantage. For example, the tool will flag an employee who may be slower to type, however the tool does not account for an employee with one hand or offer the option to disclose that information. The UN Convention on Rights of Persons with Disabilities,[79] and most regulators state that the obligation to make accommodations is triggered when the person makes the request.

[78] (Doyle, 2023)
[79] The UN Convention on Rights of Persons with Disabilities: https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-crpd

> However, automated requirement processes (including those that deploy AI) do not permit the employee to request adjustments or accommodations. The burden remains on individuals to prove they have been treated badly by the algorithm.
>
> This case offers an example of how surveillance tools can disproportionately target and negatively impact people with disabilities, causing further discrimination. When AI surveillance is designed to pursue the target optima, in this case, surveilling target productivity, people who do things differently are then flagged and penalized in their workplace. In addition, if the accommodations process is also automated, refuting misclassifications becomes a black boxed process with limited follow-up. This harms vulnerable groups as the system systemically limits accessibility.

5.7 Freedom from Discriminatory Profiling

This clause is meant to address targeted discriminatory profiling that occurs when specific traits and unique characteristics are flagged by AI systems. The Seed clause included technical terms such as biometrics and predictive policing; to bridge the gap, I defined the technical terms to make the clause more accessible:

> Regulated entities will refrain from using AI tools for:
> - biometric categorization, which is categorization based on body measurements and human characteristics, such as fingerprints or facial recognition;
> - emotion analysis, which is a data analysis process based on extracting human emotion or sentiment through in large datasets;
> - predictive policing, which is the prediction of data in order to prevent specific future outcomes.

To bridge the gap further, I focused the rationale on outlining how discriminatory profiling occurs within AI systems. Similar to statistical discrimination, this clause aims to protect against the pursuit of the target optima, where the clause highlights how AI systems assume or correlate unrelated data points in its attempt to find data patterns. It is important for committee members to see the many ways AI tools target against data outliers and be able to address specific examples of how AI excludes so that their experiences are not dismissed as one offs:

> People with disabilities are disproportionately vulnerable to discriminatory profiling. Disability is often medicalized, unfounded assumptions are made when attempting to correlate data points to find common themes or patterns. When the system is designed

to pursue a target optimum, discriminatory profiling can be exasperated as outliers can be negatively flagged.

To represent the necessity of this clause, I sourced two case studies from the AI incidents database that were specific to people with disabilities:

Case study 1:[80]
In 2018, Austria's Public Employment Service developed a system to predict a job seeker's employment prospects and allocate appropriate forms of support to them. This 'AMS algorithm' works by automatically classifying job seekers and calculating individual 'IC' scores based on their gender, age, citizenship, education, health, care obligation and work experience, amongst other factors, to determine their relative employability. It then assigns an individual to one of three prospective employability groups. Academics and civil rights groups found that the algorithm discriminates against women over 30, women with childcare obligations, migrants, and people with disabilities by giving these groups lower scores by placing them in lower categories, even if they had the same qualifications as men or non-disabled people. By contrast, men with children were not negatively weighted by the algorithm.
The algorithm also seemed to discriminate against people living in areas of the country where unemployment rates tend to be high, thus furthering negative stigmas of people with disabilities.

By August 21, 2020, the Austrian data protection authority declared the system illegal; its deployment should be suspended.[81]

This case provides an example of how AI systems can profile against data outliers and systemically discriminate against people with disabilities. In addition, discriminatory profiling also limits any appeals to the decision due to the lack of transparency around how decisions are made by the system. The dangers of this lies in targeting profiling, where if the system goes unregulated, the system regulators can target specific groups and further discriminate against them.[82]

Case study 2:[83]
RentGrow, a tenant screening firm that leverages algorithm-based software to review credit, criminal records and eviction checks of potential tenants. Housing law advocates say thousands of people are mistakenly flagged by tenant screening software that culls criminal records data from many sources made by CoreLogic, RentGrow, RealPage, AppFolio and other companies. The rental industry has accelerated over the last two

[80] (AIAAIC, 2023)
[81] (Kayser-Bril, 2020)
[82] (Geiger, 2021)
[83] (Farivar, 2021)

decades[84] as the rental market has increased, and the digitization and real estate analytics market has boomed. Nearly all landlords now use some sort of tenant screening software to find who they consider to be the highest-quality tenants, which often discriminates against vulnerable populations. The tenant-screening industry is largely unregulated and can further discriminate against people with disabilities as the system is designed to flag and automatically reject potential tenants that deviate from target tenant optima. Tenant screening companies are currently being evaluated and scrutinized by government agencies and law makers, but the technology is already being utilized by landlords, resulting in rejecting potential tenants based on discriminatory inputs.

This case provides an example of how biometric screening tools can discriminate against people with disabilities and how these tools operate and are adopted faster than government regulation and consideration. This results in undue harm to vulnerable populations.

5.8 Freedom from misinformation and manipulation

This clause aims to protect against the spread of misinformation or data manipulation. To summarize, I organized the points into bullet form to better clarify the flow of information. I also defined toxicity monitors, which is a technical term that may not be well-known:

Regulated entities will ensure that AI systems:

- do not repeat or distribute stereotypes or misinformation about people with disabilities;
- are not used to manipulate people with disabilities;
- engage people with disabilities to determine data moderation criteria used in AI systems.

Toxicity monitors, which are censorship flags, employed by regulated entities will not censor or exclude people with disabilities or prevent the discussion of social justice issues based on censored words.

This rationale offers an excerpt from a recent The Organization for Economic Cooperation and Development (OECD) report to shed light on the significance of the clause and the manipulation that occurs through AI:

[84] (Rich, 2001)

A recent OECD report outlines a variety of AI manipulation methods. The following report excerpt highlights how large scale manipulation can emerge from small scale influence:[85]

> The EU AI act aims to ban AI systems targeting vulnerable individuals based on age and physical or mental disabilities. The widespread use of AI systems that rely on extensive user data can exploit people's cognitive differences, making them vulnerable to manipulation. To be vulnerable in this context means deviating from others on a psychometric trait, i.e. a psychological characteristic measured on a scale, in an exploitable way.
>
> Exploiting psychometric differences can lead to the creation of predictive models that show how certain groups of people will respond to a given stimulus, making them vulnerable to exploitation. *Nudge-ability* has been used to describe individuals' susceptibility to the influence of different choice architectures. Exploiting minor differences in psychological constructs can be effective on a group level, especially with proxy measures that use online data to determine people's psychometric profiles. Digital footprints on social media measure individual differences to a point where they have higher accuracy than those made by people's close acquaintances. Differences in these traits have been exploited as a vulnerability in influencing behaviour.

The following case study[86] was sourced from the AI incidents database and highlights the risks and implications of AI systems used to replace government entities and decision-making. This case study offers a concerning look into the harms caused by the UK government in adopting an AI tool that did not consider the needs of people with disabilities, thus systematically discriminating against the group of people most in need of social services. This case study was selected to show committee members the real and harmful outcomes of AI, and how this clause aims to protect and prevent the following from occurring in Canada:

> The Department for Work and Pensions (DWP), which is responsible for the UK's social security system and benefits support for some of the country's most vulnerable, actively surveils its claimants in an attempt to crack down on benefits fraud in the country. The DWP deploys surveillance methods, such as social media monitoring, covert physical surveillance, and data gathering from online accounts, to build a case against claimants in order to withdraw benefits coverage and potentially press criminal charges. The DWP is legally able to surveil these "open-source" channels by utilizing algorithms that mine the web. Due to varying regulations regarding privacy and access to personal data, these algorithms can compile the web activity of the claimant. The data gathered on claimants often doesn't match preconceived notions of people with disabilities, which

[85] (Armstrong et al., 2023)
[86] (Gabert-Doyon, 2021)

results in their activity getting flagged as suspicious. Activities such as booking or upgrading travel, attending fundraisers or other functions, and getting a gym membership may be flagged as indicators of fraud, even if they are necessary or justified. Upgrading travel is often necessary to provide extra leg room, and physical activity is advised by medical professionals, however this information is flagged by the algorithm, which then leads to questioning the actions of the claimant.

The DWP is partially authorized to use investigation and data-gathering tactics, however, recent reports indicate that the surveillance methods used by the DWP are “aggressive,” “intimidating” and “invasive,” causing fear in the lives of both the claimants and the claimant’s families. When movement-based activity is being tracked and flagged as suspicious, the claimant is now in the precarious position of providing the rationale behind their online activity. Not only does this impact the overall health and wellbeing of the claimant, but it furthers the stigma around people with disabilities faking claims.

This case offers an example of how differences flagged by AI systems can be used to manipulate how the information is used (in this case, against providing social services) against people with disabilities.

5.9 Transparency, Reproducibility and Traceability

To further AI explainability, this clause is designed to protect against opaque decision-making. Section 5.4 of this MRP recommends the use of AI explainability when designing for more equitable decision making. To build on that, regulating AI transparency will support system monitoring and iterations when disputes arise. As demonstrated through the clauses and examples thus far, we can see that AI decisions change over time, and without transparency, outcomes may change without being able to pinpoint the cause of issue, thus preventing corrections from being made. It is possible that AI system may not always be able to transparently disclose its decisions. In those cases, the system needs to provide supplementary information to support with transparency. To bridge the gap, I summarized the clause to reflect these points:

AI systems designed, developed, produced, and/or deployed by regulated entities will enable transparency of AI decisions and their impact.

Where possible, decisions made by AI systems will disclose the decision-making process to support any disputes.

Where decision-making process is unable to be disclosed, clear documentation of goals, definitions, design choices, and assumptions regarding the development of an AI system shall be made publicly available.

The Seed Standards document provided commentary on additional legislation and frameworks to consider. To provide further context and support the technical and legal capacity gap, I structured the rationale to focus on defining transparency and describing its significance:

> Transparency means making complex technology easier to understand by the general public. A reason for making AI transparent is to engage people so that they understand how the system affects them, allowing them to raise meaningful commentary or recommend meaningful changes to the system. This is especially the case where the AI system discriminates against them.[87]
>
> Transparency is relevant to all stages of AI. With respect to training data and labels, there should be a summary of where the data came from, the specific characteristics of the data that were used for training, and the rationale for any labels. Similarly, the algorithm used should document where it came from and why it was chosen for the given AI system.[88]

I included this case study[89] to demonstrate the critical need for transparency in AI systems and to support committee members in further understanding the significance of this clause:

> The U.S. Justice Department is investigating the county's child welfare system to determine whether its use of the influential algorithm discriminates against people with disabilities and other protected groups. Over the past six years, Allegheny County has served as a laboratory for testing AI-driven child welfare tools that crunch large amounts of data about local families to try to predict which children are likely to face danger in their homes. Today, child welfare agencies in at least 26 states and Washington, D.C., have considered using algorithmic tools, and jurisdictions in at least 11 have deployed them, according to the American Civil Liberties Union.
>
> AI-driven tools enable county hospitals to enter patient and family information, which provides an automated risk score and can alert child services as part of its output. As part of a yearlong investigation, the AP obtained the data points underpinning several algorithms deployed by child welfare agencies, including some marked "CONFIDENTIAL." Among the factors they have used to calculate a family's risk, whether outright or by proxy: race, poverty rates, disability status and family size. They include whether a mother smoked before she was pregnant and whether a family had previous child abuse or neglect complaints.

[87] (Center for Democracy & Technology, 2023)
[88] (Sinders, 2022)
[89] Ho, 2023

> The developers behind the algorithm claim transparency as they make their models public, however, families are unable to question or gain rationale behind the outputs of the tool due to confidentiality claims made by officials. Officials equally do not have access to the rationales behind decisions made by the screening assessments. The adoption of these screening assessments has enabled cash-strapped agencies to focus on children needing protection, however, the tool seems to produce biased results against parents with disabilities. Without transparency behind the decision making, parents are unable to refute wrongful assessments or receive a rationale behind their scores. The impact has resulted in separated families and potential lifelong developmental consequences for the impacted children.
>
> This case study demonstrates the detrimental effects of opaque AI systems towards people with disabilities, and their surrounding community. As seen in this example, public models dare not enough in creating meaningful transparency; and meaningful transparency is necessary in creating ethical AI systems as it includes feedback from the people being impacted by the tool when designing the system, and it enable for future iterations as the process is transparently documented and publicly available. This also ensure all stakeholders involved in the creation, procurement, and launch of the AI systems can contribute towards the iterations.

5.10 Accountability

Without transparency, there cannot be accountability. To bridge the gap in this clause, I focused on addressing the impact of the lack of accountability and the harm it causes. I then connected the impact to why it is necessary for regulated entities to establish accountability processes:

> Without a link to human responsibility, AI systems producing harmful results cannot be held accountable for causing harm. As AI continues to expand, it becomes increasingly difficult to dispute results or iterate on changes as it is unclear who is responsible for what. When developing AI systems, a traceable chain of human responsibility, accountable to accessibility expertise, must make it clear who (person / institution) is liable for decisions made by an AI system.
>
> Regulated entities will establish accountability of the AI training process, which involves assessing the breadth and diversity of training sources, tracking provenance/source of training data, verifying lack of stereotypical or discriminatory data sources, and efforts to ensure the training and fine-tuning processes do not produce harmful results for people with disabilities.

To provide further context into the harms and risks of the lack of accountability in AI, I focused the rationale to outline a possible chain of events that could take place in any AI system:

> There is a chain of actors when fully developing an AI system to use for tasks. To illustrate:
>
> - One organization's business is to create and polish training data that has certain properties.
> - Another organization or company uses that training data to build an AI model where the training data's properties are relevant to the purpose of the model.
> - A third group incorporates that model into a production ready system and markets that system to clients for their use.
>
> Then those clients use the systems for making decisions that, sometimes, have a harmful impact on individuals. Who is responsible for the negative outcome? To quote FAT/ML's Principles for Accountable Algorithms, "the algorithm did it" is not an acceptable excuse.[90]
>
> Who or what organization is responsible depends on the stage at which harmful impacts were introduced. Transparency is a necessary enabler here since it provides guidance for determining what went wrong with the AI system and, more importantly, at what stage of its development. Different people or organizations are responsible for the stage(s) that they contributed to. This implies that there needs to be a traceable chain of human responsibility.
>
> Regarding people with disabilities, accessibility expertise should be a factor in determining where the negative impact was introduced, why it is harmful to them, and thereby who is responsible for the harm.

It was difficult sourcing a case study[91] that was specific to people with disabilities due to the very reason the clauses advocating for transparency and accountability; due to the lack of both transparency and accountability, it is difficult to pinpoint specific system failures, which results in making generalized claims about AI not working. I decided to use a news article that impacted marginalized communities. As noted, the implication of harm towards racialized communities can extend to people with disabilities:

> Dr. Anil Kapoor was diagnosed with stage four colon cancer. The commonly prescribed cancer drug Fluorouracil (5-FU) killed him within three weeks of taking the drug that was meant to prolong his life. As the Canadian healthcare system continues to stretch in capacity and resources, some provinces now pre-screen cancer patients for genetic variants — differences in people's DNA — that can lead to serious illness and even death when taking the medication. Dr. Kapoor passed his initial prescreen, however, tests later revealed a genetic variant that wasn't included in the pre-screening, which led to his passing. After further investigation, it was revealed that current pre-screening guidelines are based on studies that largely leave out populations that aren't white, a known problem

[90] (Diakopoulos, et al., 2022)
[91] (Marchitelli & Blair, 2023)

based on medical studies they found from North America and other parts of the world. Of the provinces that pre-screen for potential toxic reactions, many check for what are considered the four most common genetic variants instead of conducting full genome test due to cost and infrastructure. After investigation, it was found that the four most common variants in pre-screening mostly involve patients who are white, leaving other populations more vulnerable as patients are not informed that generic test may not be applicable to them. Patients can conduct a full genome test, which would be an out-of-pocket cost, however, it is up to the doctor's discretion on whether to inform the patient on moving forward with a full test. Since the decision was determined by a. screening assessment, the family was unable to hold a singular person or institution accountable. Instead, they took it upon themselves to bring this information to Go Public (CBC story submission link) to warn other patients and help to educate.

This example highlights how the lack of human accountability in AI systems links to a lack of system monitoring, causing negative harm towards vulnerable populations. The implications of this example can extend to people with disabilities. The algorithm was meant to automate the pre-screening of genetic variants and qualify patients for drug use, however, the data used to train the algorithm did not factor in unique profiles, thus wrongfully representing populations who diverge from the statistical mean. However, without the system being linked to human decision makers and without system transparency, the impacted families of the victim are unable hold the institution accountable as it isn't clear who is responsible for the data or corresponding decisions.

This pre-screening tool has been adopted by many provinces across Canada and continues to be used to pre-screen patients to save on time and cost. The harmful implications of continuing to use a tool that doesn't transparently disclose its innerworkings or disclose its decision makers makes wrongful predictions increasingly difficult to catch and track. In this case, Dr. Anil Kapoor was a medical professional, part of a family of other medical professionals. The family members were able to advocate for further testing for themselves due to their knowledge in the space, which resulted in some clarity, however, this is often not the case for many other patients. Dr. Kapoor's death was avoidable if he was informed of his genetic makeup and the risks behind the drug.

This case study offers an example of why accountability is necessary in AI systems, especially when they are used to replace human intervention. In this example, Dr. Kapoor's death was preventable, however, the lack of diverse data was not disclosed as part of test, making it difficult to understand who/what entity is responsible for its iteration. In this case, Dr. Anil Kapoor's family has taken this matter to the public to bring more awareness, however, the onus now falls on victims instead of regulators and developers. Dr. Kapoor's family was in a position to help bring awareness, but people who are future outliers may not be able to advocate in the same way, thus resulting in more harm being caused.

5.11 Individual agency, informed consent and choice

To highlight the legislative requirement of this clause, I included an Accessible Canada Act Principle to present the current protections people with disabilities have in place:

> Accessible Canada Act Principle 6d states that "all persons must have meaningful options and be free to make their own choices, with support if they desire, regardless of their disabilities."[92]
>
> To abide by ACA, AI providers will offer users to request an equivalently full-featured and timely alternative decision-making process that is, at the user's choice, either:
>
> a. performed without the use of AI, or
> b. made using AI with human oversight and verification of the decision.
>
> Algorithmic service providers will establish reasonable service level standards (including equivalent quality, detail, currency, availability, and timeliness of response) for all levels of optionality, ensuring that the human service modes meet the same standards as those of the fully automated default. Providers will avoid creating disincentives that reduce individual freedom of choice to select human-supported or AI-free service modes. This requires alternative services to be sufficiently well-resourced, equivalent in functionality, and comparable to the unsupervised default.

The following rationale helps to further support why this clause is necessary:

> The risk of unsupervised automated decision-making is that impacted individuals will have no means to avoid or opt out of the harms of statistical discrimination unless human-supervised and AI-free alternatives are provided. These alternative services need to provide an equivalent level of service, currency, detail and timeliness.

The following example[93] differs from the other examples presented in this MRP as it offers a framework into how developers can consider implementing choice and informed consent into AI systems. The purpose of this framework example is meant to give committee members insight into alterative ideas that they can recommend during committee discussion:

> This blog post articulates the friction between people and algorithmic systems, where AI systems are generally adapted based on user feedback, however, there's a lack of transparency in how that feedback is gathered and what feedback is used. The researcher in this post advocates for all people to question the good intentions of AI and instead, leverage an actional framework that helps put ethical AI into practice. This framework is known as Terms-we-Serve-with.[94] This framework sets out to enable people to recognize, acknowledge, challenge and transform existing power asymmetries in AI. It is meant to enable practitioners, builders, and policymakers to foster transparency, accountability, and engagement in AI, empowering individuals and communities navigating cases of algorithmic harms and injustice to transform them by aligning AI tools with a psychology of care and service.

[92] (*Accessible Canada Act,* 2024*)*
[93] (Rakova, 2023)
[94] (Rakova, 2023)

> The blog post goes on to say that friction is important in developing and progressing the tool forward, but only when decisions and iterations are made transparent to the user, and that users should have the autonomy to give and take away consent as this process will give way to creating more unique experiences with AI instead of a generalized experience.

5.12 Support of human control and oversight

As seen throughout the case studies presented in this MRP thus far, AI is being utilized to replace human decision-making. When AI is left on its own with no accountability or transparency, the results can and will cause immense harm against vulnerable populations. This clause is proposing AI to support human decision-making rather than replace. To support committee members in feeling equipped to advocate for this protection, I summarized the clause to focus on actionable recommendations:

> AI systems will provide a mechanism for reporting, responding to, and remediating or redressing harms resulting from statistical discrimination by an AI system, which is managed by a well-resourced, skilled, integrated, and responsive human oversight team.
>
> AI system providers will collect and disclose metrics about harm reports received and contestations of decisions. These reports shall adhere to privacy and consent guidelines.
>
> Reports of harm, challenges, and corrections to the decisions of an AI system will be integrated into a privacy-protected, continuous feedback loop used to improve the model results. This feedback loop will involve human oversight, including consultation with disability community members to ensure that harms are sufficiently remediated with the AI model.

To support with capacity building committee context and knowledge I included the following principles and legislation:

> OECD AI Principle #3[95] outlines the need for those adversely affected by an AI system to be able to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.
>
> The NIST AI Risk Framework[96] defines a taxonomy of harms that include harms to people (including individuals, communities, and broader societal impacts), and to ecosystems (including supply chains, the environment, and natural resources.

[95] (OECD AI Principles, 2019)
[96] (The NIST AI Risk Framework, 2023)

I leveraged the same case study[97] used in the previous clause as the framework directly addresses the need for human oversight rather than using AI to replace human decision-making:

> This example was used in the previous clause (section 4.2.12) and is applicable here as well as it encompasses the need for human oversight and control over AI decisions to ensure it's not negatively harming or impacting groups of people; and if there is harm, there is accountability in the feedback process.

5.13 Cumulative Harms

Statistical discrimination and cumulation harm are the most harmful risks against people with disabilities, however neither of the terms are brought up in legislation today. This clause outlines the protections necessary to mitigate cumulative harm. To prevent cumulative harm from occurring, regulators must first implement all the clauses mentioned in the Seed Standards document in order as each clause builds on the another. To support with bridging the gap, I have summarized the direction committee members can advocate for in to help prevent cumulative harm from occurring:

> There shall be a process to assess impact from the perspective of the individuals with disabilities, and to prevent statistical discrimination that is caused by the aggregate effect of many cumulative harms that intersect or build up over time as the result of AI decisions that are otherwise classified as low and medium risk.

To support this Seed clause further, the following rationale is meant to highlight the impacts of cumulative harm overtime to give committee members specific language and framing to use when discussing the harm and impact of AI against people with disabilities.

> Cumulative discrimination can be defined as the inequities built up in small increments by individuals (but also across generations) through risks deemed insignificant to the majority, producing a vicious cycle of discriminatory risk assessment is that continually re-created.[98] This form of discrimination by AI tooling causes cumulative harm from many low impact decisions skewed against someone who is an outlier or small minority.

[97] (Rakova, 2023)
[98] (Wilson, 2011)

The following case study[99] is not specific to people with disabilities; however, it highlights the effects of cumulative harm over time towards racialized groups and how AI can perpetuate systemic bias and racism. This case study is meant to articulate the impacts of cumulative harm and provide an applicable example of the harm in action:

> This example highlights the racial bias in risk assessments used by the criminal justice system. Due to the historical data used to train these learning models, the assessments often connect race with one's likeness to repeat crime. What the learning model doesn't factor in is the history of targeted oppression towards racialized community members, causing racialized people to be more likely profiled and incarcerated than non-racialized community members. Since the learning model leverages historical data to predict the likelihood of future crimes, the learning model is inherently biased and discriminatory against racialized people.
>
> This example goes on to compare the actual states behind the risk assessment versus the actual cases of repeat offenders, which produced inaccurate results between projected repeat offenders and actual repeat offenders. The concerning aspect of these assessments is that they are being used in court rooms today to support judges in their decision making, which the tool still leverages biased historical data. Without transparency into the data itself and without clear accountability or human intervention, the results produced by the tool are increasingly difficult to refute. This adds to the cumulative harm experienced by outliers as the system designed is biased against them. This extends to all outliers in any community.
>
> This example highlights how aggregate effects of cumulative harm negatively impact data outliers and the importance of reviewing all AI decisions.

5.14 Organizational Processes to Support Accessible and Equitable AI

This is a new section of the Seed Standards document. The previous Seed Standards section focused on designing equitable AI systems as a whole whereas this section specifically addresses the organizations surrounding the AI lifecycle (design, develop, procure, customize, regulate) and what they should consider when participating in the AI lifecycle. This includes the designers, the developers, the implementers, the regulators, and the organizations protecting disability rights. As the introductory clause, the summary highlights what the section will be addressing. Many of the clauses in this section will build on the clauses from the previous section. The design goal for this section was to maintain clear language to continue building committee legal and

[99] (Angwin et al., 2016)

technical knowledge, as well as make connections with previous clauses to showcase the connectivity of how the AI standards can work together to protect against harm:

> To make AI systems accessible and equitable to persons with disabilities, regulated entities will ensure its organizational processes include and engage people with disabilities in decision-making throughout the AI lifecycle. These processes will be accessible to people with disabilities as staff members, contractors, clients, disability organization members, or members of the public.
>
> The organizational processes to which this clause applies include but are not limited to processes used to:
>
> a. Plan and justify the need for AI systems,
> b. Design, develop, procure, and/or customize AI systems,
> c. Conduct continuous impact assessments and ethics oversight,
> d. Train users and operators,
> e. Provide transparency, accountability, and consent mechanisms,
> f. Provide access to alternative approaches,
> g. Handle feedback, complaints, redress, and appeals mechanisms,
> h. Provide review, refinement, and termination mechanisms

To address the specific regulations needed in the workplace, the rationale focused on how AI systems cause harm within the workplace through continuing statistical discrimination and cumulative harm:

> AI systems are designed to favour the average and decide with the average. Within workplaces, AI systems will favour the general employee base, leaving behind outliers. Given the speed of change within AI systems and the comparative lack of methods to address harm to people with disabilities, equity and accessibility must depend largely on the processes of the AI system rather than its testable design criteria.

This case study[100] specifically highlights how AI systems systemically discriminate against outliers due to its pursuit of the target optima. This example is specific to gender; however, the harmful implications expand to people with disabilities as the system is not designed to target or assess unique characteristics. This case study was a web-based search and was chosen to highlight the importance of protecting candidates and employees against statistical discrimination and cumulative harm caused by AI:

> In 2014, Amazon automated its hiring process to efficiently hire highly skilled talent. However, in 2015, the team realized that the tool did not evaluate candidates in a gender-

[100] (Dastin, 2018)

neutral manner. That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry (Dastin, 2018). In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain," and downgraded graduates of all-women's colleges.

Amazon edited the programs to make them neutral to these terms, but that was no guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory.

Automated hiring process is used by many large organizations (such as Goldman Sachs and Hilton Hotels), however, the validity, fairness, and explainability of AI hiring tools is still very far off from where it needs to be to make equitable decisions (Dastin, 2018).

Amazon has now shut down this program and uses a "water-downed" version of the tool to eliminate duplicate applications.

This case study highlights the importance of listing the organizational process and to help determine checkpoints to measure and explain AI decisions at every stage of the AI tool. Without including these checkpoints to assess explainability and validity, the tool will likely produce discriminatory results and cause harm. In this example, Amazon sought out to hire for the "best fit," which leveraged historical data of successful candidates. The issue with this process was that the historical data was gender biased to begin with, and with the tool being trained to hire the best "fit," it discriminated against anyone who did not align with the specific profile. By only looking at the results of the tool versus the organizational process, Amazon realized the tool was gender-biased one year after it launched. If checkpoints were built throughout the process, Amazon would have likely caught the biases earlier on. Although this case study addresses gender, it provides an example of how the tool can be biased against any outlier groups, including people with disabilities.

5.16 Plan and justify the use of AI systems

Section 5.11 of this MRP highlighted the need for informed consent when faced with AI decision-making. This clause looks like the rationale for utilizing AI in the first place. This clause is meant to address the intention and rationale behind the need and use of AI systems and regulate the involvement of people most impacted. This clause specifically highlights the term "impact assessment," which is a technical term. To bridge the gap, I defined impact assessment in the summary section and clarified how impact assessments will be used to help make better informed decisions:

> To adequately consider the direct or indirect impact on people with disabilities, regulated entities will review the positive and negative impacts of an AI system through conducting an impact assessment. The results on the impact assessment will inform the degree of

> harm and opportunity towards people with disabilities. In addition, people with disabilities will be an active participant in planning and justification of AI systems.
>
> Where the AI system is intended to replace or augment an existing function, people with disabilities that face the greatest barriers in accessing or benefiting from the existing function shall be consulted in the decision-making.
>
> This clause applies to all AI systems whether or not it is determined that they directly affect people with disabilities.

The rationale continues to bridge the technical capacity gap as it further explains the negative impacts behind using risk-based assessments and the need for involving people with disabilities as part of the justification process:

> Regulatory frameworks differ in their approach to evaluating risk, including categorizing what the risk is, its impact, and/or harm. Identifying harm and developing risk frameworks depends upon reports of harm after it has occurred. Due to risk evaluations utilizing historical data, risk frameworks can manifest statistical discrimination in determining the balance between risk and benefit to the detriment of the statistical minorities.
>
> Documenting risks require the engagement of individuals and organizations that are most impacted by the risks. To advise and inform on risk, people with disabilities need to be active participants in impact assessments as part of the evaluation and justification of AI systems.

I leveraged the same case study[101] from section 5.13 of this MRP as it highlights the discriminatory nature of risk assessments against outliers:

> Risk assessments can negatively impact racialized individuals due to the discriminatory nature of defining “risk” and “harm” (Angwin, et al., 2016); however, it is applicable towards people with disabilities as risk assessments only look at risk to a significant number of people rather than the quality of the risk or range of impact. Assessing risk should be part of the greater impact analysis, however, due to the discriminatory nature of assessing risk in the past, assessing risk alone is not enough in understanding the full impact of a system.

5.17 Design, develop, procure, and/or customize AI systems that are accessible and equitable

[101] (Angwin et al., 2016)

This clause recognizes the need for people with disabilities to be at the center of the design, development, procurement, and customization of equitable AI systems; without the active involvement and participation of those most impacted, the AI systems cannot be considered equitable. To support with the capacity gap, I used clear language to communicate the need for involving those most impacted at the early stages of decision making:

> Design, procurement, machine re-training, and customization criteria for equitable AI systems must include the requirements of clauses 5.1 and 5.2 of this MRP.
>
> Input from people with disabilities and disability organizations need to be sought in all decisions relating to designing, developing, procuring, and/or customizing AI systems.
>
> Before implementing an AI system, people with disabilities and disability organizations must be engaged to test the direct and indirect impacts of the AI system. This engagement shall be compensated.
>
> To justify the use of AI systems, accountability and equity criteria will need be verified by a third party with expertise in accessibility and disability equity before a procurement decision of an AI system is finalized.

The rationale aims to highlight the critical need for involving people with disabilities early in the development and design of AI processes rather than waiting until the system is built:

> Under the Accessible Canada Act[102], regulated entities are obligated to consider accessibility and consult with people with disabilities when procuring systems. Due to its speed of change, AI systems require more careful and informed engagement, requiring input from a broad range of people impacted by the choices.

Taking a different approach, this case study[103] offers a positive outlook of how AI systems can be equitably designed when those most impacted are brought into the design process early on and are educated on the technical and legal capacity gaps as part of the design process. To help bridge the technical gap, including a success story to show committee members what the positive outcomes could look like to further understand what we are working towards building:

[102](ACA, 2024)
[103] (Azzo, 2023)

Recent years have seen growing adoption of AI-based decision-support systems (ADS) in homeless services, yet we know little about stakeholder desires and concerns surrounding their use. This research paper seeks to understand impacted stakeholders' perspectives on a deployed ADS that prioritizes scarce housing resources. The researchers employed AI lifecycle comic boarding, an adapted version of the comic boarding method, to elicit stakeholder feedback and design ideas across various components of an AI system's design. Feedback was gathered from county workers who operate the ADS daily, service providers whose work is directly impacted by the ADS, and unhoused individuals in the region. Research findings demonstrate that stakeholders, even without AI knowledge, can provide specific and critical feedback on an AI system's design and deployment, if empowered to do so.

This case study offers an example of the positive impacts of including diverse stakeholders, including those being impacted by the system, to provide feedback throughout the process. As part of the research process, the researchers informed and educated participants and found that the insights and perspectives provided by non-technical individuals helped iterate the system to cause less harm. This process can apply to involving people with disabilities through the design of an AI system as long as participants are informed and educated.

5.18 Conduct ongoing impact assessments, ethics oversight and monitoring of potential harms

Designing equitable AI systems requires the collective efforts of all parties involved in the AI life cycle. Building off sections 5.10 and 5.11 of this MRP, accountability can be addressed when transparency is in place; and to maintain transparency and accountability, all parties involved must contribute to the monitoring and betterment of AI systems:

> Regulated entities shall maintain a public registry of harms, contested decisions, reported barriers to access, and reports of inequitable treatment of people with disabilities related to AI systems.
>
> A publicly accessible monitoring system encompassing all federally regulated entities that employ AI systems shall be established and maintained to track the cumulative impact of low, medium and high impact decisions on people with disabilities.
>
> Thresholds for unacceptable levels of risk and harm shall be established with national disability organizations and organizations with expertise in accessibility and disability equity.

To continue address harm mitigation, this rationale focuses on harm reduction by looking at all levels of harm in impact and risk assessments instead of only focusing on high-risk:

> Existing regulations and proposed regulations, including the EU Act[104], only focus on high risk or high impact decisions. However, when placing focus on high-risk, other classifications of risk can still cause harm towards people with disabilities. To help prevent cumulative harm experienced by people with disabilities, monitoring systems should also consider harms from low and medium impact AI decision systems.

I chose to use the *Machine Bias* case study in this section as well as it highlights the negative impacts (statistical discrimination) of cumulative harm and why it is necessary to review all impacts of AI decision systems, not just high risk:

> Outliers already make up a reduced dataset compared to the general dataset. Low and medium impacts of AI systems may not present issues to generalized datasets, however, when low or medium impacts are flagged within an already reduced population with diverse and complex characteristics, the cumulative harm is often not considered. One of the examples listed in this case study looked at the risk assessments used by the US

[104] (EU Act, 2023)

> criminal justice system and the corresponding results based on an individual's race. In this example, a White man and a Black woman were evaluated, where the White man was deemed a non-repeat offender, and the Black woman was flagged as most likely to repeat. After a few years when the profiles were revisited, the White man had multiple repeat offenses and the Black woman did not, however, was still flagged in the system. This assessment did not negatively impact the White man, however, did negative impact the Black woman as her profile was flagged, which contributes to further harm caused by the system. This specific risk assessment was used to determine how likely an individual would commit another offense and it produced discriminatory results.
>
> This example demonstrates the importance of looking at the whole picture rather than just high-risk AI impacts high-risk may be prevented when addressing low-medium risk flags.

5.19 Train Personnel in Accessible and Equitable AI

To distribute the collective responsibility of designing, developing, and supporting equitable AI systems, individuals as part of the AI life cycle must receive training so that they are aware of what to protect against:

> All personnel responsible for any aspects of the AI life-cycle will receive training in accessible and disability equitable AI. This training shall be regularly updated by regulators and include harm and risk detection strategies.

The rationale further articulates why this is necessary, which contributes to bridging the technical capacity gap:

> Due to the speed of change seen in AI systems, harm and risk prevention or detection requires awareness and vigilance by all personnel, not just regulators. AI deployment is often used to replace human labour, reducing the number of humans monitoring and detecting issues. As less humans are involved in the process, ensuring AI systems are not causing harm against vulnerable populations is critical in their safety and security. This means that all personnel involved in AI systems (designing, developing, procuring, customizing, regulating) should be trained to understand the impact of AI systems on people with disabilities and ways to prevent harm.

To demonstrate the collective responsibility of designing, developing, and maintaining equitable AI systems, this article[105] provides an example of how to train non-technical individuals on contributing to improving existing AI systems:

> This article provides an example of researchers and collaborators at Purposeful AI engaging with, co-designing, and training non-technical individuals on understanding AI systems and contributing towards AI iterations through feedback. This example is meant to show that training non-technical people on ethical AI is possible if they are empowered

[105] (Azzo, 2023)

through training and education aligned to their level of understanding. Doing so enables more awareness of how the system works, and enables more perspective and feedback, which can lead to AI systems causing less harm.

5.20 Provide transparency, accountability, and consent mechanisms

The clause offers organizations a list of what's needed to gather informed consent from users. Committee members could share this list when speaking with organizations who are unsure how to gather informed consent or feel blocked on taking the first step:

i. what data was used to pre-train an AI, customize, or dynamically train an AI system,
j. data labels and proxy data used in training,
k. the decision to be made by the AI and the determinants of the decisions,
l. the names and contact information of individuals within the regulated entity accountable for the AI systems and resultant decisions.

To ensure the information is accessible and understood, it needs to be provided in non-technical and plain language so that the potential impact of the decision is clear.

It must be possible to withdraw consent at any time without negative consequence.

The following case study[106] is another positive example of regulation and collaborative practices to support equitable AI:

This article offers a positive example of government regulation contributing to ethical AI. In summary, OpenAI launched "incognito mode," which allows users to opt out of the tool saving and using personal data. This feature was added on because of the pressures placed by the GDPR and European data regulators, and other AI companies are following suit if they aim to operate within the EU. The result is that people are now taking control over their own data, which is a positive step caused by government regulation. As seen in the examples presented in section 4.2, informed consent is critical to building ethical AI as historically, data has been used against vulnerable populations, such as people with disabilities.

5.21 Provide access to equivalent alternative approaches

As we have seen in previous clauses, informed consent and choice is critical in designing equitable AI systems. This clause highlights the need for choice and outlines what organizations can do to ensure choice is available for users that opt out:

[106] (Heikkil, 2023)

> The organization shall retain individuals that have the necessary expertise to make equitable human decisions regarding people with disabilities when AI systems are deployed to replace decisions previously made by humans.

I used the same case study[107] from section 5.2 of this MRP as it highlights the negative impacts of AI taking over human decision making without human intervention or giving participants the option to request alternative approaches. Due to the biases in AI tools, people with disabilities must be offered alternative avenues for the system to be accessible and equitable.

5.22 Handle feedback, complaints, redress, and appeals mechanisms

This clause offers guidance to organizations on how to handle feedback and appeals. Committee members can leverage this list and advocate for the following information to be gathered during regulatory meetings:

- are easy to find, accessible, and actionable,
- acknowledge receipt and provide response to feedback and incidents in no more than 24 hours,
- provide a timeline for addressing feedback and incidents,
- offer a procedure for people with disabilities or their representatives to provide feedback on decisions anonymously,
- communicate the status of addressing feedback to people with disabilities or their representatives and offer opportunities to appeal or contest the proposed remediation.

I leveraged the case study[108] used in section 5.9 of this MRP as it highlights the harm against people with disabilities when AI systems do not have a mechanism for feedback, complaints, or appeals. In this case, the parents who were inaccurately flagged by the AI system were unable to appeal the decision as there was a lack of explainability and accountability to who owns the decisions made by the tool.

5.23 Review, refinement, halting and termination mechanisms

In the situation where the system malfunctions or equity criteria for people with disabilities degrade or are no longer met, the AI system should be halted until the

[107] (Geiger, 2021)
[108] (Ho & Burke, 2023)

malfunction or inequitable treatment is addressed, or the system is terminated. The capacity gap includes two case studies:

1. Case study[109] used in section 5.4 as it depicts the changes in AI systems overtime.
2. Case study[110] used in section 5.13 as it demonstrates the pausing and eventual termination of an inequitable AI system.

6. LIMITATIONS

6.1 Proof of Concept

Due to time and project scope limitations, this MRP was unable to test the effectiveness of the capacity support resource with its intended users. This capacity support gap is proof-of-concept. The next step for work is to test it with real committee members requiring this information.

6.2 Access to the Capacity Building Resource

This capacity-building resource is only made available to committee members that have been invited to participate in regulatory feedback and consultative sessions. This impacts access to the information as the capacity resource is meant to be shared with a specific group of individuals. It would be a positive move to share this capacity building resource with the greater population as a general educational resource to inform the public on what is necessary for equitable AI systems and why it is important.

7. CONCLUSION

7.1 Contributions to the field

Through this inclusive design and research project I have designed a resource that addresses the legal, technical, and digital literacy gaps between committee members and the information they need to know to advocate for the right and necessary protections against AI harm and risk. The goal of this MRP is to support committee members by bridging an expertise and knowledge gap. This resource can be used by

[109] (Confino, 2023)
[110] (Dastin, 2018)

any non-technical individual looking to gain more awareness and understanding of the implications of AI systems.

7.2 Next steps or future work

This capacity building resource can continue to be developed for the rest of the Seed Standards document. Beyond that, this capacity resource should be tested with the intended users and iterated upon. As AI is an quickly evolving set of technologies, the capacity building resource should be continuously updated as new information emerges.

BIBLIOGRAPHY

AI Ethics Impact Group (2020). “From Principles to Practice – An interdisciplinary framework to operationalize AI ethics.” https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf

Angwin, J., Larson, J., Kirchner, L., & Mattu, S. (2016, May 23). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

AIAAIC. (2023, February). Austria Ams Job Seeker algorithm. AIAAIC. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/austria-ams-job-seeker-algorithm

Armstrong, S., Franklin, M., Ashton, H., & Gorman, R. (2023). *The EU’s AI Act needs to address critical manipulation methods*. OECD.AI. https://oecd.ai/en/wonk/ai-act-manipulation-methods

Azzo, A. (2023, October). *Designing AI tools for underserved populations from the ground up*. Designing AI Tools for Underserved Populations from the Ground Up: Center for Advancing Safety of Machine Intelligence - Northwestern University. https://casmi.northwestern.edu/news/articles/2023/designing-ai-tools-for-underserved-populations-from-the-ground-up.html

CDT’s 2022 Annual report – a path forward for democracy. Center for Democracy and Technology. (2023, June 23). https://cdt.org/2022-annual-report/

Cevora, G. (2020). *How discrimination occurs in data analytics and machine learning: Proxy variables*.Medium. https://towardsdatascience.com/how-discrimination-occurs-in-data-analytics-and-machine-learning-proxy-variables-7c22ff20792

Confino, Paolo. “Over Just a Few Months, CHATGPT Went from Accurately Answering a Simple Math Problem 98% of the Time to Just 2%, Study Finds.” *Fortune*, Fortune, 20 July 2023, fortune.com/2023/07/19/chatgpt-accuracy-stanford-study/.

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). *Noam Chomsky: The false promise of chatgpt*. The New York Times. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Dans, E. (2023, January 27). *Chatgpt and the decline of critical thinking*. IE Insights. https://www.ie.edu/insights/articles/chatgpt-and-the-decline-of-critical-thinking/

Dastin, J. (2018, October 10). *Amazon ditches AI recruiting tool that didn’t like women - national*. Global News. https://globalnews.ca/news/4532172/amazon-jobs-ai-bias/

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H. V., Unsworth, K., Sahuguet, A., Venkatasubramanian Suresh Venkatasubramanian, S., Wilson, C., Yu, C., & Zevenbergen, B. (2022, April). *Principles for accountable algorithms*

and a social impact statement for algorithms. Fairness, Accountability, and Transparency in Machine Learning. https://www.fatml.org/resources/principles-for-accountable-algorithms#principles

Dhinakaran, A. (2023, October 5). *Overcoming AI's transparency paradox*. Forbes. https://www.forbes.com/sites/aparnadhinakaran/2021/09/10/overcoming-ais-transparency-paradox/?sh=5c07dda94b77

Doyle, N. (2023, October 5). *Artificial Intelligence is dangerous for disabled people at work: 4 takeaways for developers and buyers*. Forbes. https://www.forbes.com/sites/drnancydoyle/2022/10/11/artificial-intelligence-is-dangerous-for-disabled-people-at-work-4-takeaways-for-developers-and-buyers/?sh=7eda615935d3

Edwards, B. (2023, November 30). *Chatgpt is one year old. here's how it changed the Tech World.* Ars Technica. https://arstechnica.com/information-technology/2023/11/chatgpt-was-the-spark-that-lit-the-fire-under-generative-ai-one-year-ago-today/

Farivar, C. (2021, March 14). Tenant Screening Software Faces National reckoning. NBCNews.com. https://www.nbcnews.com/tech/tech-news/tenant-screening-software-faces-national-reckoning-n1260975

Friedman, V. (2021, February 11). Why is Facebook rejecting these fashion ads?. The New York Times. https://www.nytimes.com/2021/02/11/style/disabled-fashion-facebook-discrimination.html

Gabert-Doyon, J. (2021, March 2). How the government spies on welfare claimants. VICE. https://www.vice.com/en/article/y3g9n5/how-the-government-spies-on-welfare-claimants

Geiger, G. (2021, March 1). *How a discriminatory algorithm wrongly accused thousands of families of fraud*. VICE. https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud

Government of Canada / Gouvernement du Canada. (2023, September 27). *Artificial Intelligence and Data Act*. Innovation, Science and Economic Development Canada Main Site. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act

Government of Canada / Gouvernement du Canada. (2023b, December 7). *Consultation on the development of a Canadian code of practice for generative artificial intelligence systems*. Innovation, Science and Economic Development Canada Main Site. https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems

Government of Canada / Gouvernement du Canada. (2024, April 8). *Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI*

Systems. Innovation, Science and Economic Development Canada Main Site . https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems

Government of Canada / Gouvernement du Canada. (2023, December 7). *What We Heard – Consultation on the development of a Canadian code of practice for generative artificial intelligence systems*. Innovation, Science and Economic Development Canada . https://ised-isde.canada.ca/site/ised/en/what-we-heard-consultation-development-canadian-code-practice-generative-artificial-intelligence#a4

Harwell, D. (2019, November 9). Hirevue's AI face-scanning algorithm increasingly decides whether ... The Washington Post. https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/

Heikkil, M. (2023, May 2). *We need to bring consent to AI.* MIT Technology Review. https://www.technologyreview.com/2023/05/02/1072556/we-need-to-bring-consent-to-ai/

Ho, S., & Burke, G. (2023, March 15). *Not magic: Opaque AI tool may flag parents with disabilities*. AP News. https://apnews.com/article/child-protective-services-algorithms-artificial-intelligence-disability-f5af28001b20a15c4213e36144742f11

Inclusive Design Research Centre. (1994). Inclusive Design Research Centre. https://idrc.ocadu.ca/

Inclusive Design Research Centre. (2016). *Welcome to the Inclusive Design Guide*. Welcome to The Inclusive Design Guide | The Inclusive Design Guide The Inclusive Design Guide. https://guide.inclusivedesign.ca/

International Organization for Standardization. (2021, November 5). *ISO/IEC TR 24027: Bias in AI systems and AI aided decision making*. ISO. https://www.iso.org/standard/77607.html

International Organization for Standardization. (2020, May 28). *ISO/IEC TR 24028: Overview of trustworthiness in artificial intelligence*. ISO. https://www.iso.org/standard/77608.html

International Organization for Standardization. (2022). *ISO/IEC TS 5723: Trustworthiness Vocabulary*. ISO. https://www.iso.org/standard/81608.html

Kang, C. (2024, March 18). *The Department of Homeland Security is embracing A.I.* The New York Times. https://www.nytimes.com/2024/03/18/business/homeland-security-artificial-intelligence.html?campaign_id=9&emc=edit_nn_20240318&instance_id=117867&nl=the-morning&regi_id=97502909&segment_id=161074&te=1&user_id=a968ed997d11bd639bbed361470c36a3

Kayser-Bril, N. (2020). *Austria's employment agency rolls out discriminatory algorithm, sees no problem*. AlgorithmWatch. https://algorithmwatch.org/en/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/

Kostiainen, A. (Ed.). (2024, January 8). *Ethical principles for web machine learning*. W3C. https://www.w3.org/TR/webmachinelearning-ethics/#transparency

Marchitelli, R., & Blair, J. (2023, November 27). *This commonly prescribed cancer drug was supposed to help save this doctor's life. instead, it killed him | CBC news*. CBCnews. https://www.cbc.ca/news/canada/toronto/cancer-drug-5fu-genetic-variant-testing-1.7039145

Marr, B. (2024, February 20). *A short history of chatgpt: How we got to where we are Today*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=5c471b6e674f

Marr, B. (2024a, February 20). *10 amazing real-world examples of how companies are using CHATGPT in 2023*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/30/10-amazing-real-world-examples-of-how-companies-are-using-chatgpt-in-2023/?sh=424542761441

Misconceptions about disability. (2024). Accessible UNH. https://www.unh.edu/diversity-inclusion/accessible/disability-101/misconceptions-about-disability

Liskovoi, L., Neogi , T., Seeschaaf-Veres, A., & Treviranus, J. (2024). *Capacity Building Support Site*. Canvas. https://canvas.instructure.com/courses/8302976/modules

Noone, C. (2021, April 19). *Flawed data is putting people with disabilities at risk*. TechCrunch. https://techcrunch.com/2021/04/19/flawed-data-is-putting-people-with-disabilities-at-risk/#:~:text=Disabilities%20are%20diverse%2C%20nuanced%20and,are%20excluded%20from%20its%20conclusions.

Oppenheim, M. (2018, October 11). *Amazon scraps "sexist AI" Recruitment Tool*. The Independent. https://www.independent.co.uk/tech/amazon-ai-sexist-recruitment-tool-algorithm-a8579161.html

PEAT. (2023a, April 24). *Civil rights principles for hiring assessment technologies*. Peatworks. https://www.peatworks.org/ai-disability-inclusion-toolkit/civil-rights-principles-for-hiring-assessment-technologies/

PEAT (2023, April 24). *Risks of bias and discrimination in AI hiring tools*. Peatworks. https://www.peatworks.org/ai-disability-inclusion-toolkit/risks-of-bias-and-discrimination-in-ai-hiring-tools/

Pearson, J. (2020, August 24). *The story of how the Australian Government screwed its most vulnerable people*. VICE. https://www.vice.com/en/article/y3zkgb/the-story-of-how-the-australian-government-screwed-its-most-vulnerable-people-v27n3

Rakova, B. (2023, May 16). *Reimagining consent and contestability in AI*. Medium. https://bobi-rakova.medium.com/reimagining-consent-and-contestability-in-ai-56979a88a7fb

Rich, M. (2001). *SafeRent's math speeds up tenant evaluation process* . The Wall Street Journal . https://www.wsj.com/articles/SB996702441926667410

Silvers, A., & Francis, L. P. (2005). Justice through Trust: Disability and the "Outlier Problem" in Social Contract Theory. *Ethics*, *116*(1), 40–76. https://doi.org/10.1086/454368

Sinders, C. (2022, June). *When can we call machine learning "transparent"?: New_ public magazine*. New_ Public. https://newpublic.org/article/1950/when-can-we-call-machine-learning-transparent

Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), 57-84.

Stanley, J. (2017, June 2). *Pitfalls of artificial intelligence decision making highlighted in Idaho ACLU case: ACLU*. American Civil Liberties Union. https://www.aclu.org/news/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case

Summary of the Accessible Canada Act. (n.d.). Canada.Ca. https://www.canada.ca/en/employment-social-development/programs/accessible-canada/act-summary.html

The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job applicants and Employees. (2022). US EEOC. https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence

The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. (2020). European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530

Treviranus, J. (2019). Inclusive design: The bell curve, the starburst and the virtuous tornado. *Medium*. https://medium.com/@jutta.trevira/inclusive-design-the-bell-curve-the-starburst-and-the-virtuous-tornado-6094f797b1bf

Treviranus, J. (2018a). The three dimensions of inclusive design: Part one **. *Fwd50*. https://medium.com/fwd50/the-three-dimensions-of-inclusive-design-part-one-103cad1ffdc2

Treviranus, J. (2018b). The three dimensions of inclusive design: Part one **. *Fwd50*. https://medium.com/fwd50/the-three-dimensions-of-inclusive-design-part-one-103cad1ffdc2

Treviranus, J. (2018c). The Three Dimensions of Inclusive Design, part three**. *Medium*. https://medium.com/@jutta.trevira/the-three-dimensions-of-inclusive-design-part-three-b6585c737f40

Treviranus, J. (2020). *We count: Fair treatment, disability and machine learning, by Jutta Treviranus (OCAD University)*. W3C Workshop on Web and Machine Learning. https://www.w3.org/2020/06/machine-learning-workshop/talks/we_count_fair_treatment_disability_and_machine_learning.html

Trewin, S. (2018). *AI Fairness for People with Disabilities: Point of View*. IBM Accessibility Research. https://arxiv.org/ftp/arxiv/papers/1811/1811.10670.pdf

United Nations. (n.d.). *Convention on the rights of persons with disabilities (CRPD)* | *division for inclusive social development (DISD)*. United Nations. https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-crpd

US Department of Commerce. (2024, January 5). *AI Risk Management Framework*. NIST. https://www.nist.gov/itl/ai-risk-management-framework

Walch, K. (2023, October 5). *How AI is finding patterns and anomalies in your data*. Forbes. https://www.forbes.com/sites/cognitiveworld/2020/05/10/finding-patterns-and-anomalies-in-your-data/?sh=7d182f04158e

Welcome to the artificial intelligence incident database. (n.d.). https://incidentdatabase.ai/

Wilson, G. (2011). Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage. Contemporary Sociology, 40(3), 306-307. https://doi.org/10.1177/0094306110404515m