Framework for Identifying Highly Consequential AI Use Cases

November 2023

Special Competitive Studies Project &
Johns Hopkins University Applied Physics Laboratory

Introduction

The Special Competitive Studies Project (SCSP) in collaboration with the John Hopkins University Applied Physics Laboratory (JHUAPL) developed the "Framework for Identifying Highly Consequential AI Use Cases."

The genesis of this collaborative effort stems from the comprehensive analysis of the United States' approach to AI governance in SCSP's 2022 *Mid-Decade Challenges to National Competitiveness* report which outlined four pivotal AI governance principles.[1] In Mid-Decade Challenges, SCSP recommends that the United States should leverage its robust sector regulatory expertise to uptake AI regulation and govern AI use cases and outcomes by sector. Regulators also need the necessary resources to adopt AI regulation. A critical AI governance principle is that U.S. regulators cannot regulate every AI use case. Rather, the U.S. regulatory approach must focus their efforts on highly consequential AI use cases, whether beneficial or harmful. This framework operationalizes these principles.

With those principles as the guiding post, this framework is a classification tool that regulators can use to determine on which AI uses and outcomes to focus their regulatory efforts. Without context and awareness of the potential harms and benefits of an AI use case, it is difficult to comprehensively classify the impacts of an AI system. This framework is intended to aid regulators in making an informed initial classification of an AI use case. It does not speak to what regulatory action should be taken, only that an AI use or class of AI use cases will significantly impact society and requires their attention.

The framework is a starting point. It is a template that regulators can modify for their sector-specific needs. However, it will need reiteration, with more input particularly from regulators, to make it more effective and implementable. It should be viewed as a living document.

The framework fits within SCSP's broader mission to strengthen America's long-term competitiveness as AI and other emerging technologies shape our national security, economy, and society. To review the breadth of SCSP's work, visit SCSP.ai.

The partnership that produced this report brought together experts and practitioners from across the national security, regulatory, and emerging technology communities through a series of roundtables. SCSP and JHUAPL would like to thank the many government experts and regulators, academics, civil society leaders, and industry experts for their time and insight. It was prepared by SCSP and JHUAPL staff and, as such, may diverge from the opinions of some expert contributors.

[1] Mid-Decade Challenges to National Competitiveness, Special Competitive Studies Project at 87 (2022).

Background

Artificial intelligence (AI) is a transformative technology with the potential to revolutionize many aspects of our lives. While AI offers benefits such as improving health, education, productivity, and solving some of the world's most pressing problems, it also has the potential to be harmful. AI use can result in the spreading of disinformation and discrimination, for example. In order to shape AI for the public good, we must both incentivize the beneficial use of AI while recognizing and mitigating the worst of the societal harms.

AI technologies must be developed and used ethically and responsibly. This requires governing AI, using both regulatory and non-regulatory mechanisms, in ways that align with established democratic values. At the same time, an American approach to AI must balance innovation with regulation.

To achieve these objectives, the Special Competitive Studies Project provided four AI governance principles in its first report, *Mid-Decade Challenges to National Competitiveness*.[2] These principles are:

- **Govern AI use cases and outcomes by sector.** The risks and opportunities presented by AI are inextricably tied to the context in which it is used. Therefore, sector-specific governance is the best approach to balancing interests to achieve optimal outcomes.

- **Empower and modernize existing regulators.** Existing regulatory bodies were created in a different technology era. The United States needs to empower and modernize its key regulators, and energize their engagement for the new AI era.

- **Focus governance on high consequence use cases.** It is impractical to govern every AI use or outcome. Regulators should focus their efforts on AI uses and outcomes that will be highly consequential to society, including potential unintended uses, whether beneficial or harmful.

- **Strengthen non-regulatory AI governance.** Strong, robust non-regulatory governance mechanisms should be used in addition to regulatory guardrails to properly shape AI development and use to ensure flexibility, adaptiveness, and relevance.

With respect to governing through regulation, sector regulators are faced with the challenge of regulating the rapidly evolving field of AI technology. AI systems are becoming increasingly sophisticated and widespread, and it is difficult to regulate every system and use case. Some AI development and use cases pose a potential for significant negative or positive impacts on society, and thus, warrant more attention than others. As a result, regulators need to be strategic about which AI use cases and outcomes they focus their regulatory efforts upon.

Multiple frameworks exist for assessing and mitigating risks associated with AI. While these mechanisms are effective for their particular missions, there is no widely accepted approach for

[2] Mid-Decade Challenges to National Competitiveness, Special Competitive Studies Project at 87 (2022).

identifying whether an AI development or use case is of high consequence to society before further regulatory actions are taken.

The United States and our allies and partners need an approach to AI regulation that promotes innovation by regulating AI systems that have the most significant consequences on society. Such an approach would allow for the continued development of beneficial AI uses, while addressing potentially significant harms from development and use.

A risk-based approach to AI regulation that identifies highly consequential AI use cases aligns with democratic values. This approach respects individual rights, liberties, and freedoms and harnesses impactful benefits of AI systems for public good, while protecting individuals and society from the worst of the harms. The draft EU Artificial Intelligence Act (the "EU AI Act") also takes a risk-based approach to regulation, meaning that it regulates AI systems based on the potential risk of harm they pose to individuals or society.[3] This framework employs a risk-based approach. However, instead of relying on a static list of technologies and applications, this framework takes into account the dynamic nature of AI technologies and provides regulators with flexibility in its application.

An approach to identifying highly consequential AI systems should be both risk-based and flexible. Such an approach should provide justified confidence in AI systems (by supporting trustworthiness & responsibility) for the public, certainty to industry, and flexibility to regulators to apply sector-specific expertise and experience, as different sectors and uses have different risk thresholds. An American approach should also provide insight into potential beneficial impacts of AI development or use on society that regulators can choose to regulate (e.g., incentivize through funding or consider whether a benefit warrants equitable access by the whole of society).

This document sets forth a framework for identifying highly consequential AI use cases (ID HCAI Framework) that might have significant benefits or harms on society. The framework is a tool that can help regulators ensure that the development and use of AI systems align with democratic values. By using this template, regulators can focus their efforts on AI systems that are highly consequential to the public, standardize their approach across sectors, and adapt their approach to the specific needs of different sectors. Additionally, by documenting their processes and decision making, regulators can help to ensure accountability and transparency. This framework template should be adopted and tailored by regulators for sector-specific needs.

Some AI use cases will require regulatory focus, while others will not. This framework aims to help regulators identify AI use cases in the "gray area." An initial high-level assessment must be made as to whether the AI use case under consideration warrants the resources to conduct a more thorough assessment of whether an AI system is highly consequential. The initial judgment will help determine whether the AI development or use case under consideration has foreseeable harms that could pose a significant impact on health, safety, or fundamental rights, or substantial benefits that should overwhelmingly incentivize the AI development and use. If not, then no

[3] Tambiama Madiega, Artificial Intelligence Act, European Parliamentary Research Service (2023).

further assessment is required and the AI use case is determined to not be of high consequence. Otherwise, the complete framework should be applied to help determine whether the AI use case is highly consequential to individuals or society. A suggested best practice is to document the process and rationale used at every decision point in the ID HCAI Framework.[4] It is also recommended that regulators establish a registry of evaluated AI use cases and their classifications with exceptions (e.g., for national security or justified industry secrecy). The registry should contain a mechanism by which the public can provide input on these evaluated AI system use cases. This will inform the public of assessments and classifications, allowing them to inform regulators of any contextual changes triggered by the continued use of the AI system that may affect periodic reassessments. It will also have the added benefit of informing industry about the regulators' evaluation process.

The framework interprets AI as computational systems that do some of the predictions, recommendations, classifications, and other decision making that traditionally are in the province of humans.[5] This definition includes systems which are not possible without AI, and those that make use of AI-based components, AI-enabled functions, or AI-derived algorithms. The framework is intended for assessments of AI systems as a whole, vice components, and their concrete impacts on society that result from how they change the context or condition of society. It further proposes that assessments be performed by regulators with input from multi-disciplinary experts,[6] including the public, which is best positioned to evaluate impacts on society. In addition, societal impacts are those resulting both from the use of the AI system as well as its development (e.g., impacts on data workers and from environmental impacts).

Three AI lifecycle points at which the framework can be applied:

- Regulators foresee a new application for AI,[7]
- A new application for AI is under development or proposed to a regulatory body, and
- An existing AI system has created a highly consequential impact that triggers an *ex-post facto* regulatory review.

The high-level steps to the framework are:

A. **Preliminary analysis:** Determine whether the AI application has foreseeable harms or benefits that could impact, for example, health, safety, or fundamental rights, and consequently may need to be regulated. This is intended to be an initial filter to determine whether a fuller assessment is needed.

---

[4] Documentation should include details such as the "moment" in time that an assessment is performed, which will help inform future reassessments and address potential contextual changes (e.g., the societal relationship with the AI system), rationale for making the judgment that there are no foreseeable harms or benefits that necessitate use of the framework, and the decisions made throughout the full assessment.

[5] Artificial Intelligence Risk Management Framework, U.S. National Institute of Standards and Technology at 1 (2023) ("The AI RMF refers to an AI system as an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022).").

[6] This includes domain experts that can provide research insights (e.g., from universities and Non Governmental Organizations) who understand the evolving nature of relevant/emerging impacts to consider.

[7] In this scenario, regulators should assess AI systems as they would be deployed and used.

B. **Parallel analysis of harms and benefits:** If there is foreseeable harm or benefit, conduct a more comprehensive harms/benefits analysis, which involves performing parallel harm and benefit assessments.

    a. Enumerate and evaluate the magnitude of foreseeable and actual harms from the AI system development and use.
    b. Enumerate and evaluate the magnitude of foreseeable and actual benefits from the AI system development and use.

C. **Final decision on high consequence:** Using the magnitude assessment results, determine if the AI use case is of high consequence.

    a. If yes, a sector-specific regulator must determine how best to take next steps to regulate the AI development and/or use (e.g., whether to create incentives, mitigate harms, or establish bans).

D. **Periodic reassessment:** Periodically monitor sectoral AI use to determine if the list of AI systems identified as highly consequential remains appropriate for that sector given contextual changes and whether revisions to classifications are necessary.

The potential specific harms and benefits are grouped into ten corresponding categories (see **Table 1** and **Appendix 1**). The framework provides specific harms and benefits for each category, as examples, with the recognition that specific harms/benefits will be unique to sectors. Harms and benefits are further characterized by magnitude (e.g., the scope of a harm or benefit). The framework provides factors to calculate the magnitude of an identified harm or benefit. Specifically, harms are characterized by four severity factors and four likelihood factors. Benefits are characterized by four impact factors and two likelihood factors. Lastly, the document offers ways to make high consequence determinations based on the quantification of a system analysis.

Categories of Harms and Benefits

Table 1. A list of ten harm categories (left) and benefit categories (right). Corresponding categories of harm and benefit are identified on the same row.

Table 1. ID HCAI Framework Corresponding Categories of Harms and Benefits

Harm Categories	Benefit Categories
Physical Injury	Physical Health
Emotional or Psychological Injury	Emotional or Psychological Health
Opportunity Loss	Opportunity Access
Economic Loss	Economic Access
Liberty Loss	Liberty Protection
Privacy Loss	Privacy Protection
Negative Environmental Impact	Positive Environmental Impact
Manipulation	Incentivization
Social Detriment	Social Improvement
Operational Degradation	Operational Improvement

Appendix 1 provides tables of specific harms and benefits for each category, with corresponding descriptions. These tables are intended to guide the framework user (e.g., a sector-specific regulator) through consideration of examples of the types of specific harms or benefits associated with each category. Note that potential violation of fundamental rights is incorporated into the specific harms lists. The tables are meant to be illustrative and not exhaustive lists.

As noted in the high-level steps above, this process begins with an AI development or use case for review. While no application will be completely free of any potential harm and all presumably have some potential benefit, the framework assumes that this process has been employed because the possibility of some significant AI related harm or benefit has been identified as a reasonable outcome. To determine whether the AI development or use should be regulated, a framework user should explore the extent of those potential harms and benefits.

Defining Factors for Characterizing Harm Categories

The framework employs ten categories of harms.[8] Each category includes a non-exhaustive list of specific harms, located in **Appendix 1**.

The framework user steps through each category of harms to identify and describe relevant specific harms that could result from the development and deployment of the AI application being considered.

For each specific harm, the framework user would evaluate the harm's potential severity and likelihood, and document any baselines used for comparison that led to the score determinations.[9] The severity of harm factors are scale, scope, disproportionality, and duration. The likelihood of harm factors include probability, frequency, lack of detectability, and lack of optionality. Descriptions for each of these factors are presented in **Table 2** and **Table 3**, respectively.

Table 2. Harm Factors of Severity

Factor	Description
Scale	How acutely the harm could impact a population or group throughout the AI lifecycle
Scope	How broadly the harmful impact could be experienced across populations or groups
Disproportionality	Whether an individual, group, or population is disproportionately affected by the harm over that of other individuals, groups, or populations
Duration[10]	How long the harmful impact would be experienced by a population or group

[8] To explore the harms in detail, the framework adapts the Microsoft Types of Harm List to frame AI-related harms that could emerge. See Foundations of Assessing Harm, Microsoft (2022); Types of Harm, Microsoft (2023).

[9] Baselines can include any statutory boundaries for a given sector (e.g., EEOC compliance), or corresponding harms for a non-AI counterpart. Documenting baselines may help regulators develop a library of use cases for comparison.

[10] Example considerations for this factor include the relative difficulty of an individual or group to appeal the outcome from the use of the technology (finality of outcome), difficulty in mitigating a certain risk imparted by the technology (lack of mitigation; refer to NIST RMF), or the minimal ability or speed at which the technology or an affected individual or group can recover or return to normal after a consequential event (lack of resilience). In addition, this factor should assess whether a harm occurs at a low-level continuously over time (e.g., distortion of reality due to prolonged interaction with the AI system) or whether a harm is instantaneous and acute (e.g., malfunction of threat identification system, which triggers a threat neutralization procedure resulting in bodily injury).

Table 3. Harm Factors of Likelihood

Factor	Description
Probability	The likelihood the harm could impact a population or group and whether this particular harm has occurred before (e.g., through a similar use case)
Frequency	How often a population or group would experience the harm
Lack of Detectability	Likelihood of not discovering and correcting a hazard or failure mode while it remains possible to prevent or mitigate the harm
Lack of Optionality	Limited individual choice as to whether to be subject to the effects of the technology (e.g., ability to opt-out), such as from minimal human oversight to consider and remedy problems that may be encountered from the AI system

To analyze magnitudes of harms, based on severity and likelihood factors, a framework user has "scoring" options to assess magnitude at the specific harm level or at the harm category level.

One approach is to employ different rank order categories (e.g., very low to very high, with as many categories in between as desired) with stated descriptions and explanations of the categorization of impacts.[11] The deliberative process should highlight the dimensions and factors with which each harm and harm category was evaluated.

In the alternative, a framework user can employ the Likert scale.[12] For example, the Likert scale could range from 1-5 (with 1 representing "low," 3 representing "medium," and 5 representing "high"). Note that the numbers are meant to signal categories that are relative to each other, just as using descriptive categories (e.g., very low to very high). The numbers are not meant to indicate preciseness such that a magnitude "4" is exactly twice as much as a magnitude "2."

The former approach assumes that a numerical representation is neither possible nor useful, while the latter approach assigns numerical values with some "room for flexibility." Either approach for scoring the factors allows necessary flexibility because the scales and magnitudes of the different factors will vary for different harms, and the relative judgment of what is a "very high" or "5" score will differ by sector/harm. The factors are named in such a way that a "very high" or "5" score represents a negative aspect of the harm (e.g., "long duration"). The framework user has flexibility in determining the specific weightings for each harm category given the importance of specific categories may vary depending on the sector and context. We instead provide guidance on potential dependencies for each category (see Appendix 1). Some harms may be essentially instantaneous (e.g., a physical harm), while others may extend over time (e.g., psychological harm). Thus, some magnitude factors (e.g., scope) may raise different considerations for those harm categories.

[11] Guide for Conducting Risk Assessments, U.S. National Institute of Standards and Technology at Table H-3 (2012).

[12] Rensis Likert, A Technique for the Measurement of Attitudes, Archives of Psychology 140 at 1–55 (1932).

Defining Factors for Characterizing Benefit Categories

As noted in the high-level steps above, a benefits assessment should occur in parallel with the harms assessment. To determine aspects of the AI system that could encourage development and use, the framework user should explore the extent of potential benefits stemming from the use case.

In a similar approach to the harms assessment, the framework employs ten categories of benefits.[13] Each category includes a non-exhaustive list of specific benefits, located in **Appendix 1.**

The framework user steps through each category of benefits to identify and describe relevant specific benefits that could result from the development and deployment of the AI application being considered.

For each specific benefit, the framework user would evaluate the benefit's potential impact and likelihood. The impact factors are scale, scope, duration, and proportionality. The benefit likelihood factors include probability and frequency. Descriptions for each of these factors are presented in **Table 4** and **Table 5**, respectively.

Table 4. Benefit Factors of Impact

Factor	Description
Scale	How acutely the benefit could impact a population or group throughout the AI lifecycle
Scope	How broadly the beneficial impact could be experienced across populations or groups
Duration	How long the beneficial impact would be experienced by a population or group
Proportionality	Whether a population or group is proportionally affected by the beneficial impact as compared to other populations or groups

Table 5. Benefit Factors of Likelihood

Factor	Description
Probability	The likelihood the benefit could impact a population or group and whether this particular benefit has occurred before
Frequency	How often a population or group would experience the beneficial impact

[13] The non-exhaustive list of benefits was compiled from a literature review.

To analyze magnitudes of benefits based on impact and likelihood factors at either the specific benefit level or at the benefit category level, a framework user has the same options for implementation as assessing harms - either employing different rank order categories or the Likert scale.

The factors are named in such a way that a "very high" or "5" score represents a positive aspect of the benefit, (e.g., "long duration"). Some benefits may be essentially instantaneous (e.g., physical health), others may extend over time (e.g., psychological health). Thus, some factors (e.g., scope) may raise different considerations for those benefit categories. The framework user has flexibility in determining the specific weightings for each benefits category given the importance of specific categories may vary depending on the sector and context.

Analyzing "Magnitude" for Harms and Benefits

Regulators should have flexibility, based on sector-specific expertise and experience, in analyzing the magnitude of harms/benefits based on the severity, impact, and likelihood factors ("magnitude" analyses). This document suggests that the magnitude analyses of harms/benefits focus at the categorical level.[14] One way to make this analysis is by calculating the magnitude for each specific harm/benefit and calculating an average (weighted) across categories. One advantage of assessing the magnitude factors for each specific harm/benefit in **Appendix 1**, which are then analyzed at the categorical level, is that this approach provides a comprehensive view of all relevant harms and benefits with a nuanced understanding of the interplay of all the relevant harms and benefits.

Appendix 2 provides an exemplary method of calculating the magnitude analyses for specific harms and benefits, respectively.

Ways to Make High Consequence Determinations

There are many ways a framework user may choose to apply the resulting magnitude values based on their sector-specific expertise and experience. Magnitudes can be considered within a specific category or across categories. Ultimately, any approach will need to determine whether an AI development or use case being assessed is of sufficiently high consequence to warrant continued regulatory focus.

Examples of ways to combine or compare magnitudes include:

[14] This document suggests analyzing magnitudes at the category level because: (1) specific harms/benefits will be different across sectors; (2) the number of specific harms/benefits and associated factors can result in a comprehensive, but cumbersome number of overall quantifications; (3) analyses at the categorical level account for the interaction across harms/benefits that will alter the impact of any individual harm/benefit assessed independently; and (4) analyses at the categorical level allow for standardization of categories across sectors for comparative analysis and public justified confidence.

- Does any specific harm category have a magnitude greater than threshold X?[15]
- Do one or more harm categories have a magnitude greater than threshold Y?
- Does the total of magnitudes across all harm categories exceed threshold Z?
- Are the total magnitudes across all harm categories greater than the total magnitudes of benefits across all benefit categories?
- Are there one or more harm categories that might overwhelmingly outweigh the assessed categories of benefits?
- Are there one or more benefit categories that might overwhelmingly incentivize development or use, despite the magnitude, number, or type of harm categories identified?
- Does the aggregation of low magnitude level harms/benefits change context to whether the system is considered highly consequential?

Even at the categorical level, regulators should have flexibility in analyzing the magnitude for harms/benefits. For example, a framework user may focus on the most salient harms/benefits as opposed to every possible harm/benefit that exists. The framework also recognizes that even a single harm/benefit or magnitude can signal a high consequence AI use case that requires regulatory focus.

Potential next steps would then depend on responses to the above questions. For example, if the assessed character of the benefits is determined to outweigh the assessed character of the harms, the framework user could specify that development or use of the AI system should proceed, or perhaps even be incentivized. Alternatively, if the assessed character of the harms is determined to outweigh the assessed character of the benefits, then the framework user could specify whether development or use of the AI application should be reconsidered, suggest potential alternatives that have not been considered, or provide recommendations for risk mitigation[16] based on the identified harms. These determinations require sector-specific nuance, taking into consideration aspects that do not have clear answers, such as the willingness to accept the accumulation of minor consequences over time.

Periodic Reassessment of Sectoral AI Use

At the sectoral level, AI impacts should be monitored on a periodic basis to assess if the harmful/beneficial impact(s) of a previously assessed AI system has changed (e.g., environmental change or change in use) or if there has been a contextual change due to socio-techno relationship warranting revisions to the AI classification or regulation as appropriate for that sector. These reassessments may occur at the level of the individual AI system (e.g., an AI-enabled medical

[15] Efforts are underway to establish standards, metrics, and norms for AI development and use. See Biden to Push For Government Standards on AI, Politico (2023); U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools, U.S. National Institute of Standards and Technology (2019).

[16] For example, the NIST AI Risk Management Framework provides a voluntary process for managing AI risks prospectively and continuously throughout the AI lifecycle. Artificial Intelligence Risk Management Framework, U.S. National Institute of Standards and Technology (2023).

device approved by the FDA) or of a class of AI systems (e.g., AI use for credit approval or election campaign use).

Questions a framework user should pose when monitoring an AI system are:

- How is the AI system adapting post classification and deployment to societal and environmental changes within the sector over time? If the AI system does adapt, how regularly will it adapt to such change(s)?
    - Framework users must assess, based on unique sector needs, the best points at which to conduct periodic reassessments.
- What aspects of assessment, classification, or regulation need to be revised given this change?
- Whose obligation it is to conduct periodic reassessments. For example, should the obligation rest on the entity responsible for the AI system to report to the regulator (i.e., modeled after reporting recall patterns and defective systems)? Or should the regulator's responsibility include the periodic reassessments?

Conclusion

This framework aims to provide a tool for regulators to identify which AI use cases and outcomes are or will be highly consequential to society, whether beneficial or harmful. Because it is impractical to govern every AI use or outcome, regulators should focus their efforts on shaping those AI technology uses and outcomes that will be maximally impactful on society. The framework provides latitude for users to identify specific rulesets to employ for determining whether a use case is high consequence, allowing for flexible implementation across sectors. Intra-agency and inter-agency consistency will result from repeatedly applying the same framework template, being transparent through documentation, and building a library of assessments that are publicly available.

Assessments, especially pre-deployment or in early stages of release, might have to be conducted on limited information and based on hypothetical assumptions. New research and data observations post-deployment enable determinations of whether initial classifications are still appropriate and whether revisions to the governance approach are necessary. The framework serves as the first step that a regulator should take in determining if an AI system necessitates further investigation or action. In doing so, the framework also informs policy makers, industry, and civil society on relevant actions to take. It will enable all framework users to focus governance on high consequence use cases, under the assumption that the rapidly expanding application space would overwhelm efforts to address and take action on every use case.

Appendix 1: Harm / Benefit Category Tables

The following appendix is meant to supplement the ID HCAI Framework. It provides a list of example harms and benefits of AI that can be used to help regulators identify use cases of AI that have either significant harm or benefit for society. The specific harms and benefits are organized into the following focus areas:

- Physical and Psychological Health
- Consequential Services
- Human Rights and Liberties
- Social and Democratic Structures
- Performance

Each focus area contains categories and example specific harms/benefits. These examples are intended to guide the framework user through consideration of the types of potentially relevant harms or benefits associated with each category. The listed harms and benefits are meant to be illustrative, and not an exhaustive list. The framework user has flexibility in determining the specific weightings for each category given the importance of specific categories may vary depending on the sector and context.

1. Physical and Psychological Health

The following section describes example harms related to the risk of injury, and example benefits related to improving health or protecting from injury.

Physical Injury / Physical Health

Table 1.1: Physical Injury: How AI systems could hurt people or create dangerous environments.

Harm	Example
Bodily injury or death	Development or use of the AI system could result in severe physical harm. Accidents could be unintended, or the result of misuse, system malfunction, reduced appropriate human oversight, or exposure to unsafe environments or situations
Damage to critical infrastructure	Development or use of the AI system could result in damage or destruction of critical infrastructure such as water, sewage, transportation, or energy systems. Damage or destruction to

Harm	Example
	infrastructure could potentially harm populations or groups, especially vulnerable populations (e.g., power loss to a hospital)
Exposure to unhealthy agents	Development or use of the AI system could result in exposure to unhealthy agents and jeopardize health
Medical misdiagnosis	Use of the AI system could result in wrong drug recommendations for a patient or detection/diagnostic errors (e.g., failure to detect a tumor on a radiological scan)
Technology-facilitated violence	Use of the AI system could incite or enable offline violence
Assessment	Assign a "5" or "Very High" if there are clear risks of physical harm (e.g., due to minimal or no safeguards, or sufficient warnings for potential dangers are not provided). Assign a "1" or "Very Low" if there is minimal risk of harm (e.g., safeguards are provided that are adequate for the functional properties of the system)
Weighting	Depends on whether the system has functionality that could hurt those directly or indirectly subject to it, or lacks relevant characteristics that could conceivably lead to accidents

Table 2.1: Physical Health: How AI systems could protect from injury or dangerous environments.

Benefit	Example
Hazard reduction	Use of the AI system could minimize health risks from tasks that can harm people through repetitive movement, exposure to unhealthy agents, or working in dangerous conditions[17]
Medical diagnosis	Use of the AI system could provide consistent analysis of patient data for more accurate diagnoses of a variety of diseases, or aid in

[17] Foundations of Assessing Harm, Microsoft (2022).

Benefit	Example
	the detection of critical diseases earlier in their progression than conventional methods[18]
Medical access	Use of the AI system could enable more patients to access care and allow medical professionals to reach larger segments of the population, particularly in underserved areas or areas with limited resources[19]
Crime reduction or prevention	AI-driven security or surveillance systems could alert human analysts of patterns or when there is abnormal or suspicious activity[20]
Assessment	Assign a "5" or "Very High" if there are clear signs that physical health and safety can be improved with use of the system, or sufficient warnings for potential dangers are provided. Assign a "1" or "Very Low" if the system provides minimal improvements in health and safety, or sufficient warnings are not provided
Weighting	Dependent on whether the system has functionality that could affect the health and safety of those directly or indirectly subject to it, or lacks relevant characteristics that could conceivably lead to protection from injury or dangerous environments

[18] Darrell M. West & John. R. Allen, How Artificial Intelligence is Transforming the World, Brookings (2018).

[19] Kanadpriya Basu, et al., Artificial Intelligence: How is It Changing Medical Sciences and Its Future?, Indian Journal of Dermatology at 365–370 (2020).

[20] Christian Davenport, Future Wars May Depend as Much on Algorithms as on Ammunition, Report Says, Washington Post (2017); Foundations of Assessing Harm, Microsoft (2022).

Emotional or Psychological Injury / Emotional or Psychological Health

Table 1.2: Emotional or Psychological Injury: How misused AI systems can lead to severe psychological/emotional distress.

Harm	Example
Distortion of reality or gaslighting	Intentional misuse of the AI system that undermines the end-user's trust in established institutions and distorts their sense of reality
Addiction/Attention hijacking	Prolonged interaction with the AI system that leads to addiction that affects the end-user's well-being, potentially at the expense of happiness and life satisfaction, sense of direction or purpose, relationships and human interactions, and personal character
Reputation damage	The AI system could make analyses, recommend actions or use language which disparages a person's characteristics or situation
Identity theft	Use of the AI system leads to loss of control over personal credentials, reputation, or representation
Dehumanization	Use of the AI system could erode, obstruct, or deny the subjectivity, individuality, agency or distinctly human attributes of people
Harassment	Use of the AI system could lead to online abuse (e.g., cyberbullying, deadnaming, doxxing, trolling, hateful or toxic language, gender-based sexual harassment)
Invalidation	Use of the AI system could result in the denial, rejection, or dismissal of a population's or group's feelings or experiences
Misattribution	Use of the AI system could result in misattribution of an action or content to a person or group of individuals
Loss of autonomy	Use of the AI system could lead the end-user to have involuntary thoughts or feelings, or perform actions that are uncharacteristic or against their will
Intrusion on emotional state	Use of technology like face recognition to detect, analyze, process, and interpret non-verbal communication cues (facial expression, eye contact, body language, etc.) in order to intrude upon, harass, or manipulate an individual(s)

Assessment	Assign a "5" or "Very High" if the prevention of unwanted uses and undesirable extensions of the scope of use have not been systematically and successfully addressed in the development and design of the system. Assign a "1" or "Very Low" if these concerns have been addressed during system development and design
Weighting	Depends on whether the system can be used in an undesirable way, if the consequences of misuse or extensions of system scope are serious or not, and if there are interactions with public environments

Table 2.2: Emotional or Psychological Health: How AI systems can improve emotional and psychological health.

Benefit	Example
Emotional analysis/ intelligence	Use of the AI system could enhance the ability to detect, analyze, process, and interpret non-verbal communication cues (facial expression, eye contact, body language, etc.) in order to better understand social cues or assist with, for example, pediatric pain management or accessibility scenarios
Companionship	Use of AI systems can provide human-like interactions where human-to-human interaction is not otherwise available, or extend the scope of human-to-human interactions by detecting where it is needed and calling on a person to do the interaction [21]
Emotional liberation	Use of AI systems can provide human-like interactions that can help reduce self-restraint, allow people to be more willing to express themselves, reduce the feeling of being judged, and make them feel more at ease[22]
Character improvement	The AI system could make analyses, recommend actions, or use language which improves a person's characteristics or situation[23]
Assessment	Assign a "5" or "Very High" if there are clear signs that emotional and psychological health can be improved with use of the system, or

[21] Laura Donnelly, Digital Assistants Could Alleviate the Loneliness of Elderly, The Telegraph (2018).
[22] Sophie Kleber, 3 Ways AI is Getting More Emotional, Harvard Business Review (2018).
[23] Foundations of Assessing Harm, Microsoft (2022).

	sufficient warnings for potential dangers are provided. Assign a "1" or "Very Low" if the system provides minimal improvements in emotional or psychological health, or sufficient warnings are not provided
Weighting	Depends on whether the system has functionality that could affect the emotional or psychological health of those directly or indirectly subject to it, and if there are interactions with public environments

2. Consequential Services

The following section describes example harms related to the denial of consequential services, and example benefits related to access to consequential services.

Opportunity Loss / Opportunity Access[24]

Table 1.3: Opportunity Loss: How AI systems could lead to decisions that limit access to resources, services, and opportunities.

Harm	Example
Employment discrimination	Use of the AI system could result in discriminatory recommendations or decisions related to employment, where the end-user is denied access to apply for or secure a job based on characteristics unrelated to merit
Housing discrimination	Use of the AI system could result in discriminatory recommendations or decisions related to housing, where the end-user is denied access to housing or the ability to apply for housing
Insurance and benefit discrimination	Use of the AI system could result in inequitable access, cost, or allocation of insurance or social benefits, where the end-user is denied insurance, social assistance, or access to a medical trial due to biased standards

[24] Because assessments of benefits and harms are especially insightful when you know the deployment context (e.g., how AI system has been deployed, what populations are impacted, and the socio-techno relationship of the technology), the "opportunity/loss" assessment for the "Consequential Services" categories might be especially challenging to perform at the anticipatory level before the AI system has been deployed in society. Thus, it is suggested that an initial assessment is conducted based on anticipatory deployment, but this assessment is periodically reassessed based on actual deployment insights.

Harm	Example
Educational discrimination	Use of the AI system could result in inequitable access, accommodations, or other outcomes related to education, where the end-user is denied access to education due to unchangeable characteristics
Assessment	Assign a "5" or "Very High" if problems of bias have not been addressed at any stage of the development, design, and testing of the system, or the system is known for being biased. Assign a "1" or "Very Low" if these problems of bias have been adequately addressed or solved
Weighting	Depends on whether system functions include activities which can negatively affect basic human interests and rights, or access to resources, services, or opportunities

Table 2.3: Opportunity Access: How AI systems could affect decisions that improve access to resources, services, and opportunities.[25]

Benefit	Example
Employment access	Use of the AI system could result in unbiased or reduction in discriminatory recommendations or decisions related to employment, where the end-user is provided access to apply for or secure a job based on merit
Employment opportunity	AI system development results in employment opportunities not otherwise available to a population or group
Housing access	Use of the AI system could result in unbiased or reduction in discriminatory recommendations or decisions related to housing, where the end-user is provided access to housing or the ability to apply for housing
Insurance and benefit access	Use of the AI system could result in more equitable access, cost, or allocation of insurance or social benefits, where the end-user has access to insurance, social assistance, or a medical trial due to unbiased standards

[25] Foundations of Assessing Harm, Microsoft (2022).

Benefit	Example
Educational access	Use of the AI system could result in more equitable access, accommodations, or other outcomes related to education, where the end-user is provided access to education regardless of unchangeable characteristics
Assessment	Assign a "5" or "Very High" if the system significantly reduces existing societal inequities, or has protocols in place to minimize their occurrence. Assign a "1" or "Very Low" if the system amplifies existing societal inequities
Weighting	Depends on whether system functions include activities which can affect basic human interests and rights, or access to resources, services, or opportunities

Economic Loss / Economic Access

Table 1.4: Economic Loss: How AI systems related to financial instruments, economic opportunity, and resources can amplify existing societal inequities.

Harm	Example
Credit discrimination	Use of AI the system (e.g., biased recommendation systems) could result in difficulties obtaining or maintaining a sufficiently high credit score, where the end-user is denied access to financial instruments based on characteristics unrelated to economic merit
Price Discrimination	Use of the AI system could result in differential pricing of goods or services for different demographics of people, where the end-user might be offered goods or services at unaffordable prices for reasons unrelated to the cost of production or delivery
Financial loss	Use of the AI system could result in underpricing of goods or services for reasons unrelated to the cost of production or delivery, which might result in financial loss for the service provider
Devaluation of individual occupation(s)	Use of the AI system could result in a broader economic imbalance by minimizing or supplanting the use of paid human expertise or labor

Assessment	Assign a "5" or "Very High" if the system amplifies existing societal inequities. Assign a "1" or "Very Low" if the system partially amplifies existing societal inequities, or has protocols in place to significantly reduce their occurrence
Weighting	Depends on whether system functions include activities which have an effect on existing societal inequities

Table 2.4: Economic Access: How AI systems related to financial instruments, economic opportunity, and resources can reduce existing societal inequities.[26]

Benefit	Example
Credit access	Use of the AI system ensures equitable access to financial instruments where the end-user is provided the ability to obtain or maintain a sufficiently high credit score
Fair pricing	Use of the AI system could ensure consistent and equitable pricing of goods or services for different demographics of people, at price points that result in favorable revenue for the service provider

Assessment	Assign a "5" or "Very High" if the system significantly reduces existing societal inequities, or has protocols in place to minimize their occurrence. Assign a "1" or "Very Low" if the system amplifies existing societal inequities
Weighting	Depends on whether system functions include activities which have an effect on existing societal inequities

3. Human Rights and Liberties

The following section describes example harms related to the infringement on human rights, and example benefits related to upholding or improving human rights.

[26] Foundations of Assessing Harm, Microsoft (2022).

Liberty Loss / Liberty Protection

Table 1.5: Liberty Loss: AI recommendations and influences on legal, judicial, and social systems can reinforce biases and lead to detrimental consequences.

Harm	Example
False accusation	Use of the AI system could result in exacerbating human bias, misattribution of suspicious behavior or criminal intent, wrongful arrest, or unreasonable searches and seizures based on historical records or incorrect inferences
Social control and homogeneity	Use of the AI system could induce conformity or compliance and affect rights to freedom of association, freedom of expression or practice of religion, or personal agency
Loss of effective remedy	The inability to follow, understand, and explain the rationale in decisions made by the AI system could lead to the lack of an ability to contest, question, or trust decisions the AI system makes
Assessment	Assign a "5" or "Very High" if the system could compel conformity or compliance, or otherwise result in loss of individual rights. Assign a "1" or "Very Low" if the system does not affect individual rights, and there is sufficient awareness of the capabilities and limitations of the system
Weighting	Depends on whether system functions include activities which have an effect on legal, judicial, or social systems

Table 2.5: Liberty Protection: How AI system recommendations and influences on legal, judicial, and social systems can reduce biases and detrimental consequences.

Benefit	Example
Criminal justice	Use of AI systems in predictive risk analysis could reduce human bias in law enforcement and sentencing systems[27]

[27] Foundations of Assessing Harm, Microsoft (2022); Darrell M. West & John. R. Allen, How Artificial Intelligence is Transforming the World, Brookings (2018).

Assessment	Assign a "5" or "Very High" if the system reduces bias and results in fairer legal, judicial, or social systems. Assign a "1" or "Very Low" if the system does not reduce biases, or otherwise results in loss of individual rights
Weighting	Depends on whether system functions include activities which have an effect on legal, judicial, or social systems

Privacy Loss / Privacy Protection

Table 1.6: Privacy Loss: The information generated by development or use of the AI system could be used to determine facts or make assumptions about someone without their knowledge.

Harm	Example
Privacy violation	Non-consensual data collection, or other operations could lead to loss of data privacy or inadequate protection of personally identifiable information (PII)
Dignity loss	Exposing, compelling or misleading users to share sensitive or socially inappropriate information, which could influence how people are perceived or viewed
Forced association	Requiring participation in the development or use of the AI system to participate in society or obtain organizational membership
Permanent record	Digital files or records of end-user activity could be retained and remain searchable indefinitely
Loss of anonymity	Data and activity monitoring by the AI system could limit the end-user's ability to navigate the physical or virtual world with desired anonymity

Assessment	Assign a "5" or "Very High" if conditions of data privacy and security are not met, or relevant data is not stored in a safe and secure way. Assign a "1" or "Very Low" if relevant data is stored or managed in a safe and secure way
Weighting	Depends on whether personal or private data is stored or managed by the system

Table 2.6: Privacy Protection: How the AI system can be used to detect or manage sensitive information.

Benefit	Example
Fraud detection	Use of the AI system could provide the ability to precisely identify possible fraudulent activities (e.g., abnormalities, outliers, or deviant cases) that might require additional investigation related to the manipulation, release, or access to sensitive data[28]
Anonymity	Use of the AI system could enhance the end-user's ability to navigate the physical or virtual world with desired anonymity[29]
Sensitive data management	Use of the AI system can provide protection for sensitive data that might accidentally be exposed to humans (e.g., routing requests for healthcare records between providers)[30]
Data tracking	AI-driven data and privacy protection platforms could help organizations identify sensitive data and track and control all data movement within and outside their enterprise[31]
Assessment	Assign a "5" or "Very High" if conditions of data privacy and security are met, or relevant data is stored in a safe and secure way. Assign a "1" or "Very Low" if relevant data is not stored or managed in a safe and secure way
Weighting	Depends on whether personal or private data is stored or managed by the system

Negative Environmental Impact / Positive Environmental Impact

Table 1.7: Negative Environmental Impact: How the environment and populations or groups could be negatively impacted by the AI system life cycle.

[28] Artificial Intelligence, Automation, and the Economy, Executive Office of the President at 27-28 (2016); Foundations of Assessing Harm, Microsoft (2022).

[29] Foundations of Assessing Harm, Microsoft (2022).

[30] David Roe, The Role of AI in Ensuring Data Privacy, CMSWIRE (2020).

[31] Remesh Rachendran, How Artificial Intelligence Is Countering Data Protection Challenges Facing Organizations, Entrepreneur (2019).

Harm	Example
Adverse environmental impacts	Development or use of the AI system could lead to damage of the natural environment, damage to the built environment or property, exploitation or depletion of environmental resources, or displacement of inhabitants where resources are located
Chemical exposure	Development or use of the AI system could expose the environment, populations, or groups to toxic chemicals
Climate change	Development or use of the AI system could lead to unnecessary carbon emissions, or cause other climate harm
Assessment	Assign a "5" or "Very High" if there is risk of long-term impact on the natural or built environment and its inhabitants that cannot be mitigated or prevented. Assign a "1" or "Very Low" if the impact on the natural or built environment and its inhabitants is very low, and short-term environmental policies and regulations have been taken into account
Weighting	Depends on whether there are relevant characteristics that can reasonably influence an ecosystem

Table 2.7: Positive Environmental Impact: How the environment and populations or groups could be positively impacted by the AI system life cycle.

Benefit	Example
Environmental impacts	Use of the AI system could improve conservation and environmental efforts, including improving recycling systems, managing renewable energy for maximum efficiency, forecasting energy demand in large cities, making agricultural practices more efficient and environmentally friendly, and protecting endangered habitats[32]
Weather and environmental forecasting	Use of the AI system could increase the accuracy of weather and environmental condition forecasts, which would be important for

[32] Foundations of Assessing Harm, Microsoft (2022).

Benefit	Example
	agriculture, utility, transportation, and shipping / logistics industries[33]
Natural disaster prediction	AI-driven systems could help experts predict when and where disasters may strike with more accuracy, allowing people more time to keep themselves and their homes safe in the case of a natural disaster, and improve emergency relief response times[34]
Assessment	Assign a "5" or "Very High" if the system provides a positive impact on the natural or built environment and its inhabitants, and short-term environmental policies and regulations have been taken into account. Assign a "1" or "Very Low" if there is minimal impact, or the risk of long-term negative impact(s) that cannot be mitigated or prevented
Weighting	Depends on whether there are relevant characteristics that can reasonably influence an ecosystem

4. Social and Democratic Structures

The following section describes example harms related to the erosion of social and democratic structures, and example benefits related to the improvement of social and democratic structures.

Manipulation / Incentivization

Table 1.8: Manipulation: How the AI system's ability to create highly personalized and manipulative experiences can undermine an informed citizenry and trust in societal structures.

Harm	Example
Misinformation	Use of the AI system could result in the unintentional release of false or incorrect information

[33] Archer Charles, Top Benefits of Artificial Intelligence, Koenig (2023).
[34] Archer Charles, Top Benefits of Artificial Intelligence, Koenig (2023).

Harm	Example
Disinformation	The AI system could be exploited to deliberately release false or incorrect information or disguise it as legitimate or credible in order to deceive people
Malinformation	The AI system could be used to maliciously share genuine information that is designed to stay private to the public sphere
Behavioral exploitation, Coercion	Use of the AI system could result in exploitation of personal preferences or patterns of behavior beyond that of typical marketing or advertising to induce a desired reaction
Fraudulent behavior	Use of the AI system to intentionally conduct a deceptive action for unlawful gain
Assessment	Assign a "5" or "Very High" if there are clear risks that end-users can be harmed by the system due to the absence of safeguards against exploitation or manipulation (e.g., guidelines for data and consumer protection). Assign a "1" or "Very Low" if safeguards are provided that are adequate for the functional properties of the system
Weighting	Depends on whether the system has functionality that could result in exploitation or manipulation

Table 2.8: Incentivization: How the AI system can encourage a decision or performance of a specific individual or societal beneficial action.

Benefit	Example
Beneficial default actions	Use of the AI system could increase the chances of a specific beneficial outcome (e.g., automatic enrollment or selection of default options)
Increased/personalized knowledge	AI system could increase informed citizenry. Use of AI systems could improve educational quality by providing personalized instruction (e.g., development of instructional strategies, resources, tutoring, and evaluations tailored for each student's capabilities and limitations), real-time feedback to student replies, or freeing up additional instructional time by expediting administrative tasks (grading, scheduling, record-keeping, etc.)

Assessment	Assign a "5" or "Very High" if use of the system increases the chances of a good or positive outcome. Assign a "1" or "Very Low" if system operations have limited impact on outcome selection
Weighting	Depends on whether the system has functionality to encourage specific outcomes

Social Detriment / Social Improvement

Table 1.9: Social Detriment: At scale, the way AI systems can negatively impact people and shape social and economic structures within communities.

Harm	Example
Loss of freedom of thought, movement, or assembly	Use of AI system could impact freedom of movement, freedom of thought, rights to association, peaceful assembly, or democratic participation in government
Erosion of democracy	Use of the AI system could result in election interference, censorship, and harm to civil liberties
Stereotype reinforcement	Use of the AI system could reinforce or amplify existing harmful social norms, cultural stereotypes, or undesirable representations about historically or statistically underrepresented demographics of people
False perception	Use of the AI system could result in the proliferation of false perceptions about individuals or groups
Loss of representation/ individuality	The AI system could make use of broad categories of generalization for individuals or groups, which can constrain, obscure, suppress unique forms of expression, or diminish individuality, identities, or designations
Social erasure	Use of the AI system could result in unequal visibility of certain social groups
Social alienation	Use of the AI system could result in a failure to acknowledge an individual or group's membership in a culturally significant social group

Harm	Example
Loss of individuality	Use of the AI system could suppress unique forms of expression and amplify majority opinions or "groupthink"
Denial of self-identity	Use of the AI system could result in non-consensual classifications or representations of a person or groups of people
Assessment	Assign a “5” or “Very High” if the system can be used to influence or erode existing democratic or socioeconomic structures for a given population, or forcefully impede the ability to improve their lives. Assign a “1” or “Very Low” if system operations have limited impact on existing democratic or socioeconomic structures for members of a given population or group
Weighting	Depends on whether the system has relevant characteristics that can influence the democratic or socioeconomic structures

Table 2.9: Social Improvement: At scale, the way AI systems can positively impact people and shape social and economic structures within communities.

Benefit	Example
Transparency and accountability	The AI system could help streamline the ability to collect and analyze large amounts of publicly available material which can be used to keep organizations and governments accountable[35]
Bias detection	Use of the AI system to process data at scale could be used to detect biases in policing and legislative actions[36]
Fact checking	Use of the AI system could automate fact-checking for identifying deepfakes and misleading information if used in combination with detection algorithms and AI classifiers.[37] AI-enabled fact-checking could also provide information to end-users to inform content engagement

[35] Khari Johnson, How AI Can Empower Communities and Strengthen Democracy, Venture Beat (2020).
[36] Darrell M. West & John. R. Allen, How Artificial Intelligence is Transforming the World, Brookings (2018).
[37] Artificial Intelligence, Automation, and the Economy, Executive Office of the President at 27-28 (2016).

Assessment	Assign a "5" or "Very High" if the system can be used to influence or improve existing democratic or socioeconomic structures for a given population, or assist in the ability to improve their lives. Assign a "1" or "Very Low" if system operations have limited impact on existing democratic or socioeconomic structures for members of a given population or group
Weighting	Depends on whether the system has relevant characteristics that can influence the democratic or socioeconomic structures

5. Performance

The following section describes example harms related to the reduction of operational performance, and example benefits related to the improvement of processes, operations, and productivity.

Operational Degradation / Operational Improvement

Table 1.10: Operational Degradation: How the AI system might worsen processes, performance, output, or the work environment.

Harm	Example
Skills atrophy	Over reliance on the AI system could lead to degradation of skills necessary for fulfilling life, complacency, and reduced accessibility and ability to use manual controls
Temporal degradation	Temporal data drifts or lack of model retraining and evaluation could result in performance degradation of the AI system over time[38]
Reduced efficiency	The AI-assisted system could provide a reduction in efficiency or workflow over the current system / state[39]
Job simplification	The adoption of AI could simplify the tasks performed by employees, and potentially result in lower wages, particularly for those who are already in a lower income bracket
Work pace	Implementation of AI to reduce tedious or dangerous tasks could increase stress to workers completing more tasks of greater intensity at a higher pace

[38] Daniel Vela, et al., Temporal Quality Degradation in AI Models, Scientific Reports (2022).

[39] Andrew Green, et al., Artificial Intelligence, Job Quality and Inclusiveness, OECD Employment Outlook 2023 (2023).

Assessment	Assign a "5" or "Very High" if the system can be used to worsen the operational performance or output of a given organization, group, or team. Assign a "1" or "Very Low" if system operations have limited impact on performance or output
Weighting	Depends on whether the system has relevant characteristics that can influence operational performance or output

Table 2.10: Operational Improvement: How the AI system might improve processes, performance, and output.[40]

Benefit	Example
Enhanced productivity	Use of the AI system could improve productivity and cost savings (time and labor), and promote the human workforce to higher-value tasks through the replacement of manual or repetitive and routine processes with automation
Customer personalization	Use of the AI system could provide personalized recommendations based on pattern recognition in customer data, which could in turn improve marketing return on investment (ROI) and boost sales
Increased revenue	Use of the AI system could aid in identifying and maximizing sales opportunities
Constant availability	The AI system can run constantly and consistently with 24/7 availability, and can theoretically work endlessly to the same standard without breaks
Faster data management, decision-making	The AI system has the ability to analyze and manage massive amounts of data and recognize patterns that aren't apparent to humans, which could reduce the time associated with making decisions and performing subsequent action(s)
Value above replacement	The AI-assisted system provides an improvement in efficiency or workflow over the current system / state

[40] Dimitri Antonenko, Business Benefits of Artificial Intelligence, Business Tech Weekly (2020).

Assessment	Assign a “5” or “Very High” if the system can be used to improve the operational performance or output of a given organization, group, or team. Assign a “1” or “Very Low” if system operations have limited impact on performance or output
Weighting	Depends on whether the system has relevant characteristics that can influence operational performance or output

Appendix 2: Calculating Magnitude of Each Specific Harm/Benefit

A. Calculating Magnitude for Each Specific Harm

For each specific harm, the framework user reviews each harm magnitude factor, assigning a score using rank order categories or using a Likert scale.[41]

For the Likert scale option, the framework user would then calculate an overall magnitude for each harm by:

1. Calculating a proportional severity score,

$$Proportional\ Harm\ Severity = \frac{\Sigma(severity\ factor\ scores)}{\Sigma(max\ possible\ severity\ scores)}$$

2. Calculating a proportional likelihood score, and

$$Proportional\ Harm\ Likelihood = \frac{\Sigma(likelihood\ factor\ scores)}{\Sigma(max\ possible\ likelihood\ scores)}$$

3. Multiplying these two scores together.

$$Harm\ Magnitude = Proportional\ Harm\ Severity \times Proportional\ Harm\ Likelihood$$

The user can then use this magnitude to identify specific aspects of the system that create harm above a specific threshold and might require regulation, provide good awareness for areas which might necessitate risk reduction or mitigation, and use it as a numerical basis of comparison for corresponding benefits of the system.

B. Calculating Magnitude for Each Specific Benefit

For each specific benefit, the framework user reviews each benefit magnitude factor, assigning a score using rank order categories or using a Likert scale.[42]

[41] See assessment and weighting guidance for each harm category in Appendix 1.

[42] See assessment and weighting guidance for each benefit category in Appendix 1.

For the Likert scale option, the framework user would then calculate an overall magnitude for each benefit by:

1. Calculating a proportional impact score,

$$Proportional\ Benefit\ Impact = \frac{\Sigma(impact\ factor\ scores)}{\Sigma(max\ possible\ impact\ scores)}$$

2. Calculating a proportional likelihood score, and

$$Proportional\ Benefit\ Likelihood = \frac{\Sigma(likelihood\ factor\ scores)}{\Sigma(max\ possible\ likelihood\ scores)}$$

3. Multiplying these two scores together.

$$Benefit\ Magnitude = Proportional\ Benefit\ Impact \times Proportional\ Benefit\ Likelihood$$

These benefit magnitudes can then be used to compare to corresponding harms of the system, and identify aspects of the system that could help allow or incentivize development.

C. Calculating magnitudes for harm/benefit categories:

Under the Likert scale approach, one way to calculate aggregate magnitudes for each harm/benefit category is to average together all the specific harm/benefit magnitudes belonging to that category; thus, arriving at a maximum of 10 magnitudes for harm categories and 10 magnitudes for benefit categories.

For example, calculation of the aggregate magnitude for the physical injury category would include the following:

$$Aggregate\ Harm\ Magnitude_{PI} = \left(\frac{\Sigma(Harm\ Magnitudes)}{total\ number\ of\ harms}\right)_{PI}$$

where *PI* = the physical injury category.

Similarly, calculation of the aggregate magnitude for the physical health category would include the following:

$$Aggregate\ Benefit\ Magnitude_{PH} = \left(\frac{\Sigma(Benefit\ Magnitudes)}{total\ number\ of\ benefits}\right)_{PH}$$

where *PH* = the physical health category.

Alternatively, the framework user can apply a weighted sum or weighted average of the harm or benefit magnitudes to obtain the aggregate magnitudes for a more accurate reflection of the level of harm or benefit imparted by each category. This would allow for flexibility in the harm/benefit analysis, and better reflect differing sector priorities.