Washington State Health Care Authority

Artificial Intelligence Ethics Framework

Background and purpose

Background

The Washington State Health Care Authority (HCA) is responsible for purchasing health care for more than 2.7 million Washington residents through Apple Health (Medicaid), the Public Employees Benefits Board (PEBB) Program, the School Employees Benefits Board (SEBB) Program, and the Compact of Free Association (COFA) Islander Health Care Program. HCA is committed to whole person care, integrating physical health and behavioral health services for better results and healthier residents. As the largest health care purchaser in the state, HCA leads the effort to transform the health care system to ensure residents have access to better health and better care at a lower cost. As part of HCA's efforts, the agency has prioritized using data to achieve health system transformation.

Artificial Intelligence (AI) is a field of developing computers and technology with capability of behaving in ways that can mimic and, over time, exceed human capabilities[1]. HCA has adopted this definition of AI and differentiates it from the term "Machine Learning" (ML). ML, while sometimes used interchangeably with AI, refers to a subset[2] of efforts within the field of AI that use programming and algorithms to learn and recognize patterns from data and automatically apply that learning to improve decision-making.

Use of AI-based components in health care applications has increased significantly with the wide availability of generative AI tools such as ChatGPT, Bard, etc. Although mass awareness of AI tools in health care is fairly recent, the use of AI in health care applications, especially in the fields of biomedical devices and health insurance fraud detection, has been growing rapidly in recent years. AI-based tools such as chatbots are implemented by many national and state health care and health insurance providers to improve customer service responsiveness. The role of AI in improving fraud and abuse detection processes in the health insurance industry is universally understood and acknowledged. AI-based tools are also now being incorporated in electronic health records (EHR) systems that could potentially enable more meaningful interaction between clinicians and EHRs to facilitate patient care.

As the use of AI tools has advanced in the health care sector, so have the level and number of concerns about trust, privacy, and accountability. In the highly regulated health care sector, federal and state governments have started formulating regulatory approaches to ensure adequate protections for consumers while facilitating innovation and investment in this fast-evolving field. Federal Health and Human Services (HHS) agencies have started incorporating AI in their respective data strategies. The White House has released detailed guidance and directives on the use of AI in federal government agencies. Internationally, the Organization for Economic Cooperation and Development (OECD) has spearheaded the development of guidelines for ethical use of AI tools and technologies.

Washington State is home to technology innovators like Microsoft, Amazon, Starbucks, and many others that have been at the forefront of developing, deploying, and using AI systems. Within the health care sector in the state, many provider organizations are also partnering with their EHR providers to incorporate and enable AI tools to augment clinician workflows. Health insurance companies operating in the state are using AI-based technologies to improve customer service processes and their responsiveness, gain operational efficiencies, and prevent and detect potential

[1] **https://ai.engineering.columbia.edu/ai-vs-machine-learning/**
[2] **https://doi.org/10.1016/B978-0-12-818438-7.00012-5**

fraud and abuse. Within state government, the Office of the Chief Information Officer (OCIO) has released guidance[3] for state agencies when using generative AI tools.

Purpose

The purpose of this AI Ethics Framework is to assist the HCA AI Ethics Committee in evaluating the proposed AI use case and provide recommendations to the HCA Data Utilization Committee or the HCA Research Advisory Council, as appropriate. Although there are many approaches to governing the use of AI, HCA has decided to utilize an ethics framework as part of its existing data governance structure to govern the use of AI at HCA, or by HCA partners, involving HCA clients' data. Using an ethics perspective allows HCA to incorporate a balanced approach that is focused on protecting individuals while objectively evaluating the benefits, costs, and risks associated with utilizing AI tools.

This AI Ethics Framework is also intended to proactively provide potential requestors with information to assist them in preparing their request for using AI tools involving HCA clients' data. When used appropriately and with diligence, the HCA AI Ethics Framework can provide guidance to AI system developers, implementers, and users on best practices to maximize client protections.

Scope and definitions

The scope of implementation of the HCA AI Ethics Framework within HCA Data Governance structure and processes is limited to use cases where HCA client or member data will be accessed by an AI system, tool, or service. However, this framework may be used to guide the broader use of AI tools by HCA and other state agencies to maximize client protections while facilitating use of innovative technologies.

Definitions of some key terms used in this document are:

AI: Artificial Intelligence (AI) is a field of developing computers and technology with capability of behaving in ways that mimic and, over time, exceed human capabilities.

HCA: Washington State Health Care Authority (HCA).

HCA clients: Washingtonians served by HCA.

HCA client data: Data pertaining to HCA clients related to services provided by, sponsored by, or paid for by HCA, regardless of where the data are housed or hosted.

[3] WA OCIO Interim Guidelines on use of generative AI; **https://ocio.wa.gov/policies/generative-ai-guidelines**

HCA AI Ethics Committee

The HCA AI Ethics Committee is sponsored by the HCA Data Governance and Oversight Committee and is tasked with developing and maintaining an AI ethics framework and applying that framework to the use of AI at HCA or on HCA data. The committee reports to the Data Governance and Oversight Committee and provides recommendations to the HCA Data Utilization Committee regarding any contracts that contain AI services. The committee also reviews information from the HCA Research Advisory Council and provides recommendations or decisions on how to incorporate AI in research projects.

Committee goals

The committee has established the following goals for its work to support governing AI tools at HCA:

- Grow and develop expertise on AI and AI ethics at HCA.
- Create transparent and consistent rules for using AI for HCA and business partners.
- Establish a review process for AI use on HCA member and client data.
- Seek to advance health equity through the use of AI at HCA and business partners.
- Respect tribal sovereignty when applicable through the use of AI at HCA.

Guiding principles

The HCA AI Ethics Committee is guided by the following principles [4] [5] in its application of ethics considerations while evaluating the use of AI tools that involve HCA clients' data:

Public purpose and societal benefit: The use of AI should support HCA's and the state's work in delivering better and more equitable services and outcomes to its residents.

Safety, security, and resilience: AI should be used with safety and security in mind, minimizing potential harm and ensuring that systems are reliable, resilient, and controllable by humans. AI systems used by state agencies should not endanger human life, health, property, or the environment.

Validity and reliability: AI use should produce accurate and valid outputs and demonstrate the reliability of system performance.

Fairness, inclusion, and non-discrimination: AI applications must be developed and utilized to support and uplift communities, particularly those who are historically marginalized. Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination.

Privacy and data protection: AI should be used to respect user privacy, ensure data protection, and comply with relevant privacy regulations and standards. Privacy values such as anonymity, confidentiality, and control generally should guide choices for AI system design, development, and deployment. Privacy-enhancing AI should safeguard human autonomy and identity where appropriate.

Accountability and responsibility: AI should be used responsibly and users should be held accountable for the performance, impact, and consequences of its use in the organization's work.

---

[4] Adapted from Washington Office of Chief Information Officer (OCIO) Interim Guidelines on use of generative AI; **https://ocio.wa.gov/policies/generative-ai-guidelines**

[5] Adapted from NIST AI Risk Management Framework; **https://doi.org/10.6028/NIST.AI.100-1**

Transparency and auditability: Acting transparently and creating a record of AI processes can build trust and foster collective learning. Transparency reflects the extent to which information about an AI system and its outputs is available to the individuals interacting with the system. Transparency answers “what happened” in the system.

Explainability and interpretability: Ensure that AI use in the system can be explained, meaning how the decision was made by the system can be understood. Interpretability of a system means an organization can answer why a decision was made by the system and its meaning or context to the user.

AI ethics considerations

Fairness and bias

Fairness, as used in the context of this document, refers to concerns for ensuring equal and just distribution of both benefits and costs as well as ensuring individuals and groups are free from being treated with bias, discrimination, or stigmatization. At the same time, mitigation of biases alone does not imply the system is fair.

Bias, as used in the context of this document, is broader than just demographic balance and data representation. HCA has adopted the assessment of bias as articulated by the National Institute of Standards and Technology (NIST) along three major categories of systemic bias, human bias, and statistical and computational bias (see Appendix C).

Equitability

Equitability in AI refers to AI technologies that humans intentionally design, develop, and implement to result in equitable outcomes for everyone, including for people with disabilities.[6] Simply mitigating harmful biases in systems does not make them inherently fair. It is not feasible to apply bias mitigation universally across the entire system. For instance, consider a system in which predictions that become the basis for decisions are balanced across demographic groups. Even in such a system, there may be accessibility challenges for individuals with disabilities or those affected by the digital divide that may exacerbate existing disparities for such sub-populations.

Transparency

Transparency in AI refers to the extent to which information about an AI system, including its inputs and outputs, is made available to individuals interacting with or impacted by such a system, regardless of whether they are aware that they are doing so.[3] Simply making a system transparent does not necessarily make it an accurate, privacy-enhanced, secure, or fair system; however, it is difficult to determine whether an opaque system possesses such characteristics, and to do so over time as complex systems evolve.

Accountability

Accountability, as used in this document, refers to the system's attributes related to auditability, reporting, minimizing negative impacts, and redress processes.[7] This reflects the mechanisms and processes to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment, and use.

Auditability refers to the mechanisms enabled to assess the algorithms, data, and design processes. This includes evaluation by external auditors and the availability of such evaluation reports.

Reporting refers to the processes to report on actions or decisions that contribute to a certain system outcome, including the processes to respond to the consequences of such outcomes.

Minimizing negative impacts, while self-explanatory, also includes the processes to identify, assess, and document potential negative impacts of AI systems. The use of impact assessments

[6] Adapted from NIST AI Risk Management Framework; **https://doi.org/10.6028/NIST.AI.100-1**
[7] Adapted from Ethics Guidelines for Trustworthy AI by *EU High-Level Expert Group on Artificial Intelligence, 2019*

both prior to and during development, deployment, and use of AI systems is critical to identify and minimize negative impacts.

Redress refers to the mechanisms to ensure that when adverse impact occurs, the processes for individuals to report such impact and seek redress are easily accessible and provide adequate redress.

Morality

Morality, as used in the context of this document, is based on core biomedical moral principles and behavioral norms.[8] These principles include autonomy, beneficence, non-maleficence, and justice.

Autonomy refers to the norms of respecting and supporting individual autonomous decision-making.

Beneficence refers to the norms that prioritize relieving, lessening, or preventing harm. It includes engaging in actions that provide benefits to others and balancing the provision of benefits against the costs and risks of those actions.

Non-maleficence refers to the norms of avoiding actions that would cause harm to others.

Justice refers to the norms that support the fair distribution of benefits, risks, and costs among clients and in society in general.

Compliance — privacy, security, governance

Compliance with privacy, security, and data governance requirements is critical to building trust in the AI system. In the context of this document, compliance refers to the AI system's ability to meet all relevant federal and state laws and regulations that are applicable to the specific data and use cases for which the AI tool is proposed to be used.

Trustworthiness

Trustworthiness, as used in the context of this document, refers to the system's attributes related to accuracy, reliability, and reproducibility.[9]

Accuracy pertains to the AI system's ability to make correct predictions, recommendations, or decisions based on data and/or models. When occasional inaccurate predictions cannot be avoided, it is important that the system can indicate how likely these errors are to occur. A high level of accuracy is crucial in situations where the AI system directly affects human lives.

Reliability refers to the ability of the AI system to work as expected with a range of inputs and in a range of situations.

Reproducibility refers to the ability of the AI system to exhibit the same behavior when repeated under the same conditions.

[8] Adapted from Beauchamp TL, Childress JF. Principles of biomedical ethics. 7th ed. New York, NY: Oxford University Press; 2013; as interpreted in **http://dx.doi.org/10.3163/1536-5050.102.4.006**

[9] Adapted from Ethics Guidelines for Trustworthy AI by *EU High-Level Expert Group on Artificial Intelligence, 2019*

HCA AI Ethics Framework

The HCA AI Ethics Framework is intended to be applied by the HCA AI Ethics Committee to evaluate AI use requests that involve HCA clients' data. What follows is a high-level representation of the process:

1) Requestor (HCA staff, HCA partner, researcher, or external third party) submits request to HCA for using AI tools that involve HCA clients' data.
    a. The request may be submitted as part of existing data governance processes, or
    b. The request may be submitted directly to the HCA AI Ethics Committee by HCA leadership.
        i. In this case, the data governance team will provide assistance to capture the necessary information substantially similar to that obtained in existing data governance processes.
2) The request is routed to the appropriate HCA Data Team staff who support the HCA AI Ethics Committee.
3) Staff bring the request to the AI Ethics Committee for initial discussion and review. Staff clarify the use case and present it to the committee as part of this step.
4) If indicated by the initial committee review, staff reach out to the requestor to complete the AI Ethics: Screening Questionnaire (see Appendix A).
5) Based on responses from the screening questionnaire, the requestor may be asked to complete the AI Ethics: Assessment Questionnaire (see Appendix B), or the AI Ethics: Screening Questionnaire responses may be brought to the next available AI Ethics Committee meeting for discussion and review.
6) If the requestor is asked to complete the AI Ethics: Assessment Questionnaire, the completed responses will be evaluated by committee staff and then brought for discussion and review at the next AI Ethics Committee meeting.
7) Any recommendations from the AI Ethics Committee are then routed to the appropriate Data Governance Committee—either the Data Utilization Committee or Research Advisory Council—depending on the nature of the request.

AI Ethics: Screening Questionnaire

The AI Ethics: Screening Questionnaire is available in Appendix A. The requestor completes this questionnaire as directed by HCA staff. The AI Ethics Committee then uses it to screen requests. The intent of this screening process is to allow for a quick self-assessment by the requestor that helps the AI Ethics Committee and staff determine whether the request is ready for the much deeper AI Ethics: Assessment Questionnaire.

AI Ethics: Assessment Questionnaire

The AI Ethics: Assessment Questionnaire is available in Appendix B. The requestor completes this questionnaire as directed by HCA staff. The AI Ethics Committee then uses it in a comprehensive assessment of requests. The intent of this detailed assessment is to allow the requestor to clearly articulate how the AI tool would impact Washingtonians. Completion of this assessment will help the AI Ethics Committee and staff evaluate the request in a comprehensive manner and will assist in an informed decision-making process.

Appendix A — AI Ethics: Screening Questionnaire

The HCA AI Ethics Screening Questionnaire is designed to be completed by the requestor with as much detail as feasible. This is in addition to detailed description of the data request and intended use cases as applicable and part of existing HCA data governance processes. Responses to this questionnaire are intended to inform the HCA AI Ethics Committee's review of the request to utilize AI tools on HCA clients' data.

1. Fairness and Bias

a. Describe your strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design.
b. How did you consider diversity and representativeness of users in the data? How did you test for specific populations or problematic use cases?
c. Which processes did you put in place to test and monitor for potential biases during the development, deployment, and use phase of the system?

2. Equitability

a. Which mechanisms did you deploy that allow others to flag issues related to bias, discrimination, or poor performance of the AI system?
b. How did you assess and acknowledge the possible limitations stemming from the composition of the used data sets?

3. Transparency

a. Which measures did you establish to ensure traceability?
b. Describe the methods used to test and validate the algorithmic system:
    i. Rule-based AI systems: the scenarios or cases used in order to test and validate.
    ii. Learning-based model: information about the data used to test and validate.
c. Describe the outcomes of the algorithmic system:
    i. The outcomes of, or decisions taken by, the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).
d. How did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?
e. Describe how you clearly communicate characteristics, limitations, and potential shortcomings of the AI system.
    i. In case of the system's development: to whoever is deploying it into a product or service.
    ii. In case of the system's deployment: to the (end-)user or consumer.

4. Accountability

a. Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability in addition to internal initiatives?
b. Describe the processes established for third parties (e.g., suppliers, consumers, distributors/ vendors) or workers to report potential vulnerabilities, risks, or biases in the AI system.
c. Describe your risk or impact assessment of the AI system, which takes into account different stakeholders that are (in)directly affected.

d. Describe the mechanisms in place that allow for redress in case of the occurrence of any harm or adverse impact.

5. Morality

a. Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?
b. Did you ensure that the social impacts of the AI system are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?

6. Compliance [Privacy, Security, Governance]

a. Describe the classification of the data being used. Refer to WA OCIO Data Classification Standards[10] below:
    i. Category 1: Public Information
    ii. Category 2: Sensitive Information
    iii. Category 3: Confidential Information (e.g., PII)
    iv. Category 4: Confidential Information Requiring Special Handling (e.g., PHI)
b. Describe the applicability of all relevant federal and state privacy laws to the use of data for this request.
c. Has client consent been obtained for use of their data for this requested purpose? If not, describe the legal authority to use client data for this requested purpose.
d. Describe the mechanisms or system in place to ensure the integrity and resilience of the AI system against potential cyber-attacks.
e. Describe the security certifications of your data systems that meet relevant data security requirements in compliance with applicable state and federal laws.
f. Describe the purpose of the AI system, and clearly articulate how the data that are being used meet the minimum necessary thresholds as established under the applicable state and federal laws.
g. Describe the protocols, processes, and procedures in place to manage and ensure proper data governance.

7. Trustworthiness

a. Describe the level and definition of accuracy required in the context of the AI system and use case.
b. Describe the process in place to measure whether your system is making an unacceptable amount of inaccurate predictions.
c. Describe the verification methods in place to measure and ensure different aspects of the system's reliability and reproducibility.

[10] See OCIO 141.10, Section 4.10; **http://ocio.wa.gov/policy/securing-information-technology-assets-standards**

Appendix B — AI Ethics: Assessment Questionnaire[11]

The HCA AI Ethics: Assessment Questionnaire is designed to be completed by the requestor with as much detail as feasible. This is in addition to detailed description of the data request and intended use cases as applicable and part of existing HCA data governance processes. Responses to this questionnaire are intended to inform the HCA AI Ethics Committee's review of the request to utilize AI tools on HCA clients' data.

1. Fairness and bias

a. Describe your strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design.
b. How did you consider diversity and representativeness of users in the data? How did you test for specific populations or problematic use cases?
c. Which processes did you put in place to test and monitor for potential biases during the development, deployment, and use phase of the system?
d. Describe in relevant detail your assessment of bias in the AI system as measured against the bias categories and sub-categories in Appendix C.
e. How did you assess possible decision variability that can occur under the same conditions?
    i. If so, did you consider what the possible causes of this could be?
    ii. In case of variability, describe the measurement or assessment mechanism of the potential impact of such variability on fundamental rights.
f. Describe the working definition of "fairness" that you apply in designing AI systems.
    i. Is your definition commonly used? Did you consider other definitions before choosing this one?
    ii. Describe the quantitative analysis or metrics used to measure and test the applied definition of fairness.
    iii. Describe the mechanisms established to ensure fairness in your AI systems. Did you consider other potential mechanisms?

2. Equitability

a. Which mechanisms did you deploy that allow others to flag issues related to bias, discrimination, or poor performance of the AI system?
b. Describe your strategy to consider others potentially indirectly affected by the AI system in addition to the (end)-users.
c. How did you assess and acknowledge the possible limitations stemming from the composition of the used data sets?
d. How did you ensure that the AI system accommodates a wide range of individual preferences and abilities?
    i. Did you assess whether the AI system is usable by those with special needs or disabilities or those at risk of exclusion? How was this designed into the system and how is it verified?
    ii. Did you ensure that information about the AI system is accessible also to users of assistive technologies?

[11] Adapted from: Ethics Guidelines for Trustworthy AI *by EU High-Level Expert Group on Artificial Intelligence, 2019*

iii. Did you involve or consult this community during the development phase of the AI system?

e. How did you take the impact of your AI system on the potential user audience into account?
   i. Did you assess whether the team involved in building the AI system is representative of your target user audience? Is it representative of the wider population, considering also other groups who might tangentially be impacted?
   ii. Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?
   iii. Did you get feedback from other teams or groups that represent different backgrounds and experiences?

3. Transparency [Traceability]

a. Which measures did you establish to ensure traceability?
b. Describe the methods used for designing and developing the algorithmic system:
   i. Rule-based AI systems: the method of programming or how the model was built.
   ii. Learning-based AI systems: the method of training the algorithm, including which input data was gathered and selected, and how this occurred.
c. Describe the methods used to test and validate the algorithmic system:
   i. Rule-based AI systems: the scenarios or cases used in order to test and validate.
   ii. Learning-based model: information about the data used to test and validate.
d. Describe the outcomes of the algorithmic system:
   i. The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).

4. Transparency [Explainability]

a. How did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?
b. Describe how you assessed:
   i. to what extent the decisions and hence the outcome made by the AI system can be understood.
   ii. to what degree the system's decision influences the organization's decision-making processes.
   iii. why this particular system was deployed in this specific area.
   iv. what the system's business model is (for example, how does it create value for the organization)?
c. Describe how the AI system was designed with interpretability in mind from the start:
   i. Did you assess whether you can analyze your training and testing data? Can you change and update this over time?
   ii. Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?

5. Transparency [Communication]

a. Describe the communication to (end-)users – through a disclaimer or any other means – that they are interacting with an AI system and not with another human? Did you label your AI system as such?

b. Describe the mechanisms established to inform (end-)users on the reasons and criteria behind the AI system's outcomes.
    i. Did you communicate this clearly and intelligibly to the intended audience?
    ii. Did you establish processes that consider users' feedback and use this to adapt the system?
    iii. Did you communicate around potential or perceived risks, such as bias?
    iv. Depending on the use case, did you consider communication and transparency towards other audiences, third parties, or the general public?
c. Describe how the purpose of the AI system is clarified and who or what may benefit from the product/service.
    i. Did you specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience?
    ii. Did you think about human psychology and potential limitations, such as risk of confusion, confirmation bias, or cognitive fatigue?
d. Describe how you clearly communicate characteristics, limitations, and potential shortcomings of the AI system.
    i. In case of the system's development: to whoever is deploying it into a product or service.
    ii. In case of the system's deployment: to the (end-)user or consumer.

6. Accountability [Auditability]

a. Describe the established mechanisms that facilitate the system's auditability, such as ensuring traceability and logging of the AI system's processes and outcomes.
b. How did you ensure, in applications affecting fundamental rights (including safety-critical applications), that the AI system can be audited independently?
c. Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability, in addition to internal initiatives?

7. Accountability [Reporting]

a. Describe the processes established for third parties (e.g., suppliers, consumers, distributors/ vendors) or workers to report potential vulnerabilities, risks, or biases in the AI system.

8. Accountability [Minimizing negative impact]

a. Describe your risk or impact assessment of the AI system, which takes into account different stakeholders that are (in)directly affected.
b. Describe the training and education to help developing accountability practices:
    i. Which workers or branches of the team are involved? Does it go beyond the development phase?
    ii. Do these trainings also teach the potential legal framework applicable to the AI system?
    iii. Did you consider establishing an ethical AI review board or a similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas?

9. Accountability [Redress]

a. Describe the mechanisms in place that allow for redress in case of the occurrence of any harm or adverse impact.

b. Did you put mechanisms in place to provide information to both (end-)users as well as third parties about opportunities for redress?

10. Morality [Autonomy]

a. Does the AI system interact with decisions by human (end-)users (e.g., recommended actions or decisions to take, presenting of options)?
   i. Could the AI system affect human autonomy by interfering with the (end-)user's decision-making process in an unintended way?
   ii. Did you consider whether the AI system should communicate to (end-)users that a decision, content, advice, or outcome is the result of an algorithmic decision?
   iii. In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?
b. Does the AI system enhance or augment human capabilities?
c. Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?

11. Morality [Justice]

a. Did you carry out a human rights impact assessment where there could be a negative impact on fundamental rights?

12. Morality [Beneficence]

a. Did you ensure that the social impacts of the AI system are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?
b. In case the AI system interacts directly with humans:
   i. Did you assess whether the AI system encourages humans to develop attachment and empathy towards the system?
   ii. Did you ensure that the AI system clearly signals that its social interaction is simulated and that it has no capacities of "understanding" and "feeling"?

13. Morality [Non-maleficence]

a. Did you assess the broader societal impact of the AI system's use beyond the individual end user, such as potentially indirectly affected stakeholders?
b. Did you ensure measures to reduce the environmental impact of your AI system's life cycle?
c. Did you ensure that your system has a sufficient fallback plan if it encounters adversarial attacks or other unexpected situations (e.g., technical switching procedures or asking for a human operator before proceeding)?

14. Compliance [Privacy and security]

a. Describe the classification of the data being used. Refer to WA OCIO Data Classification Standards[12] below:
   i. Category 1: Public Information
   ii. Category 2: Sensitive Information

[12] See OCIO 141.10, Section 4.10; **http://ocio.wa.gov/policy/securing-information-technology-assets-standards**

iii. Category 3: Confidential Information (e.g., PII)
iv. Category 4: Confidential Information Requiring Special Handling (e.g., PHI)

b. Describe the applicability of all relevant federal and state privacy laws to the use of data for this request.
c. Has client consent been obtained for use of their data for this requested purpose. If not, describe the legal authority to use client data for this requested purpose.
d. Describe the processes and mechanisms in place to flag issues related to privacy and data protection, both for the data collection and data processing aspects of the AI system.
e. Describe the mechanisms in place to log when, where, how, by whom, and for what purpose data was accessed.
f. Describe the mechanisms or system in place to ensure the integrity and resilience of the AI system against potential cyber-attacks.
g. Describe the security certifications of your data systems that meet relevant data security requirements in compliance with applicable state and federal laws.

15. Compliance [Governance]

a. Describe the purpose of the AI system, and clearly articulate how the data that are being used to meet the minimum necessary thresholds as established under the applicable state and federal laws.
b. Describe the protocols, processes, and procedures in place to manage and ensure proper data governance.
c. Describe the oversight mechanisms established for data collection, storage, processing, and use.
d. Describe the processes in place to ensure data retention protocols are enforced for each type of data collected and used in the AI system.

16. Trustworthiness [Accuracy]

a. Describe the level and definition of accuracy required in the context of the AI system and use case:
    i. How will accuracy be measured and assured?
    ii. What measures will be put in place to ensure that the data used is comprehensive and up to date?
    iii. What measures will be put in place to assess whether there is a need for additional data, for example to improve accuracy or to eliminate bias?
b. Describe what harm would be caused if the AI system makes inaccurate predictions.
c. Describe the process in place to measure whether your system is making an unacceptable amount of inaccurate predictions.
d. What steps have been put in place to increase the system's accuracy?

17. Trustworthiness [Reliability and reproducibility]

a. Describe the strategy in place to monitor and test if the AI system is meeting the goals, purposes, and intended applications.
b. Describe how you tested for specific contexts or particular conditions that need to be taken into account to ensure reproducibility.
c. Describe the verification methods in place to measure and ensure different aspects of the system's reliability and reproducibility.

d. Describe the processes in place to notify when an AI system fails in certain types of settings.
e. Have the processes for the testing and verification of the reliability of AI systems been clearly documented and operationalized?
f. Describe the mechanisms of communication established to assure (end-)users of the system's reliability.

Appendix C — Categories of AI Bias[13]

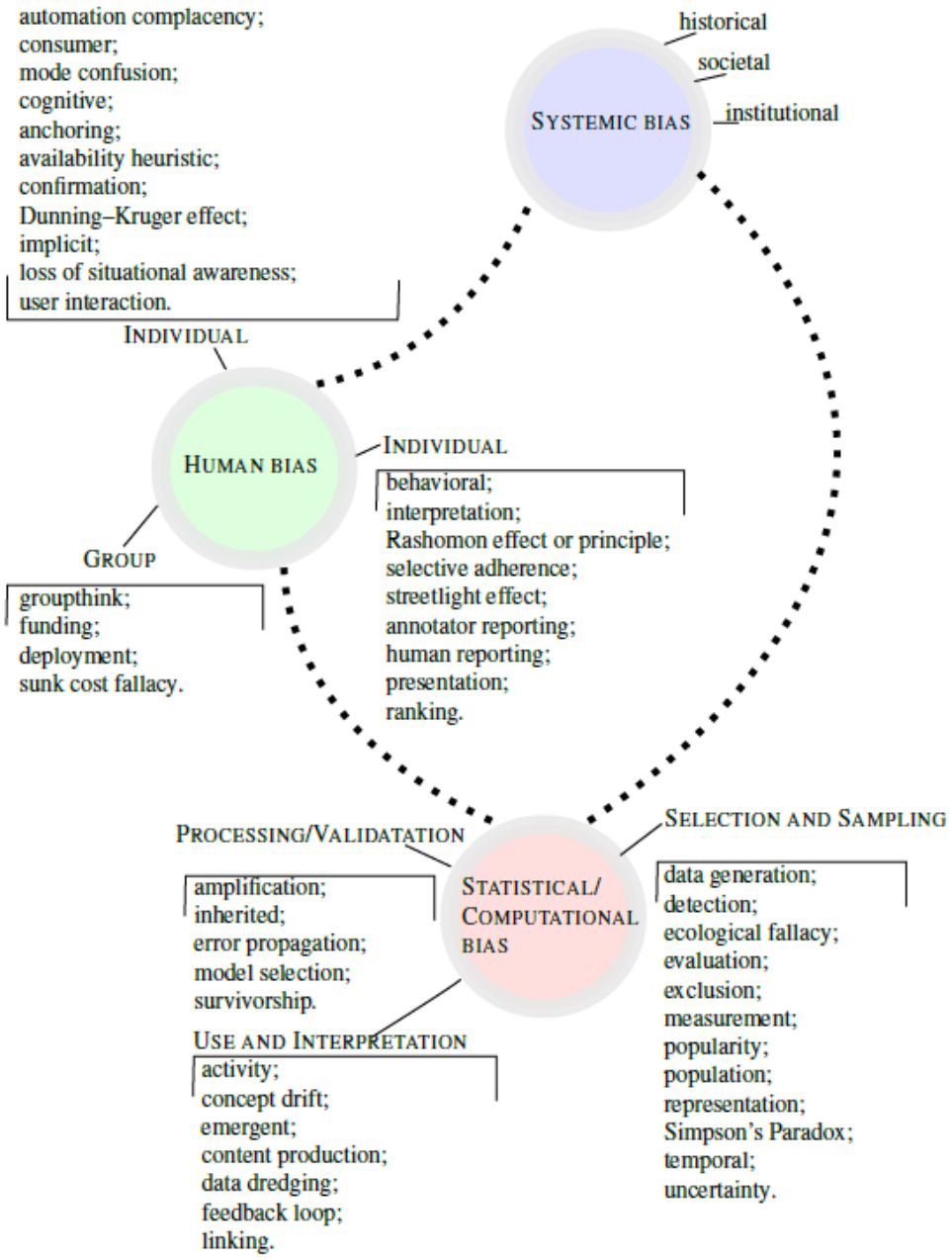

Fig. 2. Categories of AI Bias. The leaf node terms in each subcategory in the picture are hyperlinked to the GLOSSARY. Clicking them will bring up the definition in the Glossary. To return, click on the current page number (8) printed right after the glossary definition.

[13] NIST Special Publication 1270; https://doi.org/10.6028/NIST.SP.1270