# American Statistical Association Statement on Ethical AI Principles for Statistical Practitioners

October 8, 2024

Artificial Intelligence (AI) is rapidly gaining popularity, accessibility, and applicability. Continued advances in technology and demand for rapid assimilation of large volumes of data are likely to accelerate the adoption of AI. Despite the promising benefits of AI, implementing it in a safe and responsible manner presents numerous challenges.

This is particularly relevant for statistical practitioners who encounter issues distinct from standard statistical practice. For instance, AI applications rely on enormous data sets, often collected in ways that may compromise their integrity, resulting in uncertainties about the accuracy and reliability of the data. Additionally, models used in AI applications are substantively different from traditional statistical models, featuring potentially billions of parameters and a high risk of overfitting, necessitating novel and nonstandard assessment methods. Furthermore, AI applications, frequently deployed in autonomous systems, interact with the world more directly than typical statistical models, thus having a greater potential to influence decision-making processes.

As a result, the ASA provides the following ethical considerations to statistical practitioners for the development and use of AI.

The ASA Ethical Guidelines for Statistical Practice (2022) describes "statistical practice" as including activities such as designing the collection of, summarizing, processing, analyzing, interpreting, or presenting data and model or algorithm development and deployment, where the ethical guidelines also apply to the use of AI.

The ASA thinks **trustworthiness** should be the overarching objective for an ethical AI framework. Characteristics of trustworthy AI include explainability, equitability, privacy, safety, reliability, validity, human-centricity (prioritizing human values and well-being), human-in-the-loop (ensuring human oversight is integrated into AI decision-making), and responsibility. To embrace these characteristics, we focus on **accountability**, **transparency**, and **fairness** as guiding principles. These principles can serve as a guideline for members and other statistical practitioners in the ethical development and use of AI and help maintain public trust in statistical practitioner's contributions to these systems.

**Accountability Principle: Statistical practitioners and their organizations should be capable of auditing and justifying the AI systems they use or help develop.**

AI systems can have complex and unpredictable behaviors and outcomes and pose significant legal, ethical, and social challenges. As statistics plays a central role in data science and AI, statistical practitioners should seek to ensure AI is subject to appropriate oversight and governance; it can be audited, monitored, and evaluated; and these measures contribute to the trustworthiness of the technology. For example,

- Statistical practitioners should not rely blindly on AI and become lax in their statistical practice. This means statistical practitioners should apply and maintain their professional competence and keep up to date on the latest developments in AI and statistical practice. Understanding how AI generates output is important for evaluating decisions based on AI and supporting accountability.
- Statistical practitioners should clearly define operating constraints and communicate what statistical operations can and cannot be appropriately done with AI. In addition, they should promote model governance, which involves setting up quality control review points, monitoring the models over time, and formulating a plan to cope with possible review outcomes.
- Statistical practitioners have a responsibility to ensure an appropriate ecosystem is present to assess, monitor, and mitigate risks of AI.

**Transparency Principle: Statistical practitioners and their organizations should promote transparency to build trust in the AI systems they use or help develop.**

AI systems can have hidden and obscure processes and mechanisms, and they can influence people's perceptions, behaviors, and choices. It is important to communicate the tasks AI is performing versus those performed by humans, as well as whether AI is making decisions. AI systems should generate relevant and meaningful information for users and other affected parties, and this information should be accompanied by a description of benefits, limitations, and potential risks. Transparency is a vital aspect for statistical practitioners using or developing AI, as it not only provides clear and accessible information for the public and stakeholders, but also fosters trustworthiness in the system. For example,

- Statistical practitioners should seek to disclose known biases inherent in the data collection process, as well as the limitations and implications of the use of AI they use or develop, so they can communicate effectively in their practice.
- Statistical practitioners are encouraged to share the extent to which the AI systems they use or develop are validated, how widely they are adopted, and the degree of maturity of the implemented algorithms. This is particularly important for high-stakes decisions made by AI systems.

**Fairness Principle: Methods and technologies should be developed in ways that promote fairness and equity among those affected by the use of AI technologies.**

AI systems can have significant effects on people's lives, such as access to opportunities, resources, and services. Therefore, it is crucial to ensure AI does not discriminate, exclude, or harm anyone based on their identity, background, or circumstances. Fairness is a fundamental

value in statistical practice, as it aims to provide accurate and unbiased information to promote fairness in decision-making. Ensuring fairness in AI systems not only upholds ethical standards but also enhances the trustworthiness of the technology in the eyes of those it affects. For example,

- Statistical practitioners, in the AI they use or develop, should be aware of potential sources of bias and unfairness, such as unrepresentativeness of the data, within-data quality, algorithm design, model evaluation, and deployment. They should also apply appropriate methods and tools to assess and mitigate the risks of unfair outcomes and disproportionate impacts.
- Statistical practitioners should be aware of the balance between individual and collective interests, as well as those of diverse stakeholder groups. These responsibilities extend to ensuring neutrality and independence in data mining procedures and minimizing risks of exploitation, discriminatory practices, and other harms AI can facilitate or create.

The ASA Ethical Guidelines for Statistical Practice offer many specific elements that can promote ethical development and use of AI. Additionally, certain applications (e.g., the use of AI in authoring publications) may be governed by more specific guidelines, but the principles outlined in this document remain applicable. It is also important for practitioners to be aware of evolutions in ethical, regulatory, and legal guidance on the development and use of AI systems.

AI systems can affect the rights, interests, and well-being of people and society. It is essential that statistical practitioners help ensure AI is developed and used in ways that respect human dignity, values, and autonomy; demonstrate care for privacy, confidentiality, the common good, the environment, and future generations; transparently reflect when personal information is being used for AI; and communicate the appropriate uses and limitations of the AI. Our ability to embed trustworthiness into AI development and use is a key factor in building and maintaining the credibility and reputation of statistical practitioners, who rely on the trust and confidence of the public and stakeholders.

**Additional Resources**

Many guidelines exist for ethical development and use of AI. For example, one might leverage guidance related to GDPR for data privacy or DMA for marketing systems, as well as frameworks for developing AI systems that minimize the risks from these systems.

In maintaining accountability**,** practitioners may look to the National Institute of Standards and Technology's 2024 AI Risk Management Framework to help manage the risks arising from the use of generative AI, the US Government Accountability Office's 2021 AI Accountability Framework for federal agencies and other entities, or guidance from the National Telecommunications and Information Administration's AI Accountability Report.

For ensuring transparency and fairness, practitioners may look to the US Department of Justice's Guidance on AI and Civil Rights, US Equal Employment Opportunity Commission's Artificial Intelligence and Algorithmic Fairness Initiative, or US Department of Housing and Urban Development's guidance on employing the use of AI in housing as examples.

Finally, to develop general trustworthiness in AI applications, practitioners might look to recent reports such as "Toward Trustworthy AI: Mechanisms for Supporting Verifiable Claims."

These examples are not exhaustive. Statistical practitioners should recognize that scientific, government, and industry/business may have different trusted frameworks and familiarize themselves with—and consult—appropriate sources and frameworks in their day-to-day practice.