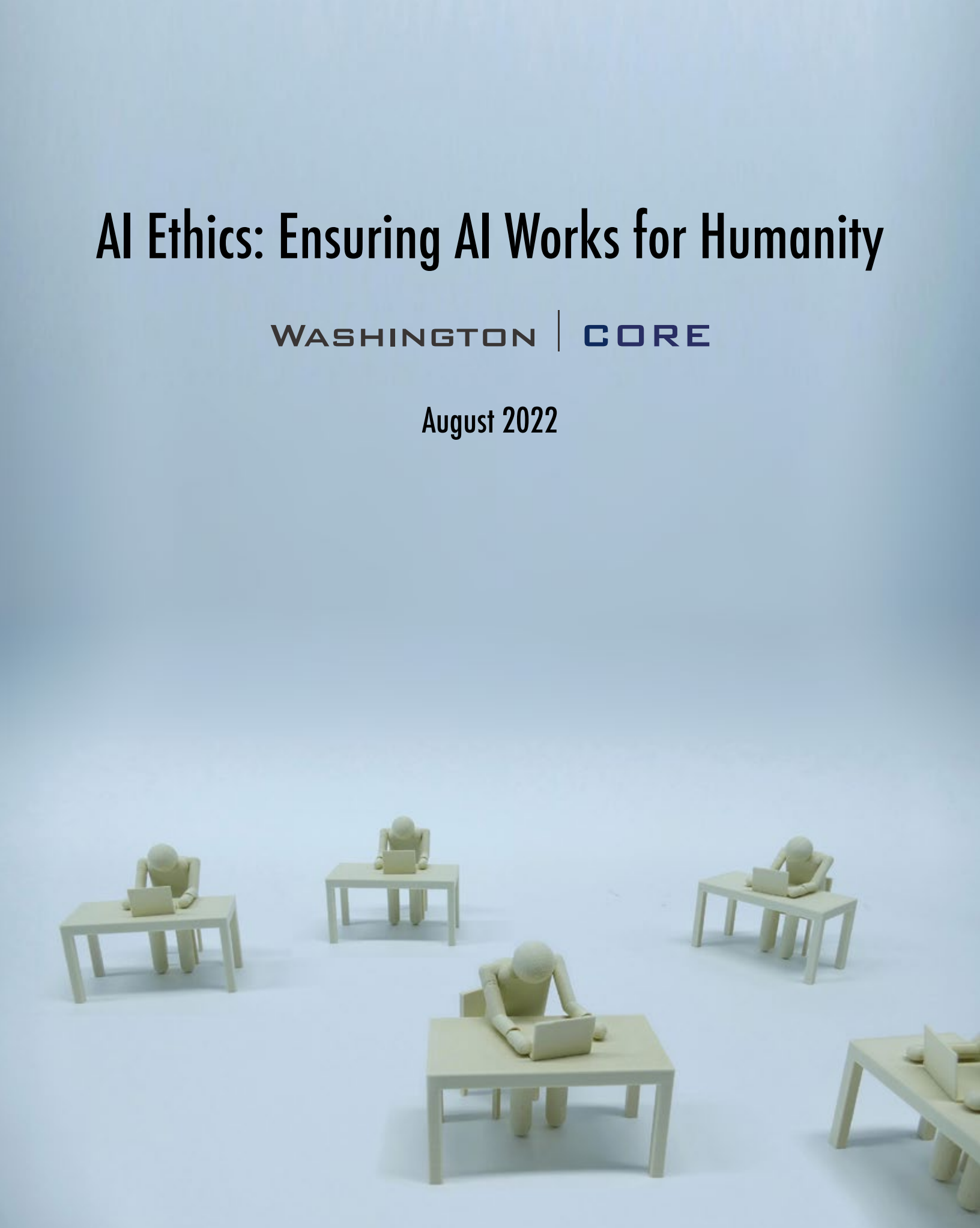
AI Ethics: Ensuring AI Works for Humanity

WASHINGTON | CORE

August 2022



The number of artificial intelligence (AI) use cases has risen rapidly in recent years, and the technology has become prevalent in services and tools affecting most facets of life, including customer service, healthcare, human resource management, transportation, and many others.

Alongside the efficiency and convenience that AI can provide, there are risks relating to the collection and use of data. One of the major concerns about the use of AI in real-world decision-making is that AI systems gather large quantities of data about individuals to make decisions, and that those decisions are susceptible to biases about age, race, ability, and gender that are embedded in the data used by the algorithms.

This has given rise to a global AI ethics movement. Recognizing the hazards, the European Union (EU) began work in 2021 on an AI Act to provide a regulatory framework for AI, which may become law as early as 2023. In the U.S., however, progress toward national regulation of AI has remained slow, although multiple related bills have been introduced in Congress. Merve Hickok, founder of AIethicist.org, shared her insights with Washington CORE on the importance of the development of ethics guidelines and laws for AI.

Ethics guidelines are important to enable acceptance of AI tools

The lack of AI regulations in the U.S. has led to a "Wild West" environment, where innovation is rapid but few guardrails exist to ensure that new applications do not cause harm to the public.

Bias can emerge at many points in the lifecycle of AI systems, from the design and development stages through deployment and operation. A clear regulatory framework is needed to protect individual rights and provide transparent legal guidance for businesses, which will help to accelerate public acceptance of new AI services.

AIethicist.org website

Source: AIethicist.org[7]

AI is increasingly being used to evaluate employees

One growing area of concern is human resource management, where firms use AI to screen job applicants and monitor employee activity. Cameras are used to track eye movement and facial expressions to make judgments about character, attentiveness, and productivity. Such emotion and facial analysis tools are criticized for lacking scientific validity. Nevertheless, the adoption of worker productivity assessment tools in particular is on the rise. In 2020, Microsoft introduced a product called Productivity Score which evaluates employee activity on Microsoft apps, such as the number of meetings attended. This generated a lot of privacy complaints, so Microsoft was compelled to modify the product so that individual workers could not be identified.[1]

In today's hybrid work environments, monitoring employee activities could result in the unintentional collection of sensitive information about individuals, for example, if background objects in an employee's home are captured by a web camera. The datafication of online activity can also lead to spurious inferences about behavior and productivity. Additionally, it has been shown that built-in biases in facial recognition and language processing algorithms can result in higher error rates for non-males, people with darker skin, people with disabilities, and people with accents.

Recognizing the higher risk of bias and discrimination in recruitment and performance management tools, a number of regulators have prioritized the regulation of AI systems in employment settings. In 2021, the New York City Council enacted Local Law 144 - aka the NYC bias audit - which requires employers using automated recruitment tools to ensure these systems are audited annually for bias. Similar bills aimed at preventing AI bias in the workplace have been introduced in California and Washington DC, and more can be expected to follow.

Meanwhile, the U.S. Equal Employment Opportunity Commission (EEOC) published a guidance document in May 2022 - "The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees" – which lays out how AI systems should be developed considering the needs and representation of people with disabilities.[2]

In the EU, a draft AI Act was released in 2021 that categorizes AI systems used in recruitment and access to employment opportunities as high-risk. The draft legislation requires vendors to provide technical documentation for AI systems, conduct conformity assessment, and have an ongoing risk management system in place.

Public awareness of the risks of AI is rising

Initially, concerns about AI were concentrated in academia, but the rising number of high-profile AI scandals has increased awareness among the general public. Most incidents, like Facebook/Cambridge Analytica in 2018 and Google Nightingale in 2019, have focused on the unauthorized collection of private data for analytical purposes. A smaller number of cases have found actual misuse of algorithms, such as Meta (formerly Facebook) using algorithms to determine which Facebook users were shown housing ads. In June 2022 Meta settled with the U.S. Department of Justice for violating the Fair Housing Act.[3] A week later, Meta-owned Instagram attracted scrutiny when it announced it would conduct a survey of users asking for their race and gender. Although Instagram claims this data would be anonymized and used only to assess how different types of users experience their services, the information could be re-identified and eventually be used in housing and recruitment databases.[4]

Most alarmingly, the 2019 childcare benefits scandal in the Netherlands demonstrated that algorithm bias could result in significant financial and legal hardships for victims. The Dutch tax administration used a self-learning algorithm to create risk profiles to help identify childcare benefits fraud. Based on the algorithm designating citizens with dual nationality as higher risk, the administration mistakenly penalized many families for potential fraud. This resulted in tens of thousands of families – many with lower incomes or belonging to ethnic minorities – sinking deep into debt, and over one thousand children were taken away into foster care. In response, the Dutch privacy regulator fined the tax administration more than €6 million in total for processing data without a legal basis, keeping the information for too long, and using the data in an "unlawful, discriminatory and therefore improper manner."

Momentum is building for AI-related guidelines and regulation in the U.S.

Currently, three bills in Congress address the potential privacy and/or bias risks of AI: the American Data Privacy and Protection Act, the Algorithmic Accountability Act, and the Facial Recognition and Biometric Technology Moratorium Act. It is uncertain whether any of these will pass in their current forms, but the American Data Privacy and Protection Act appears to have a good chance. Bipartisan support is growing for holding companies responsible for infringing on users' privacy.

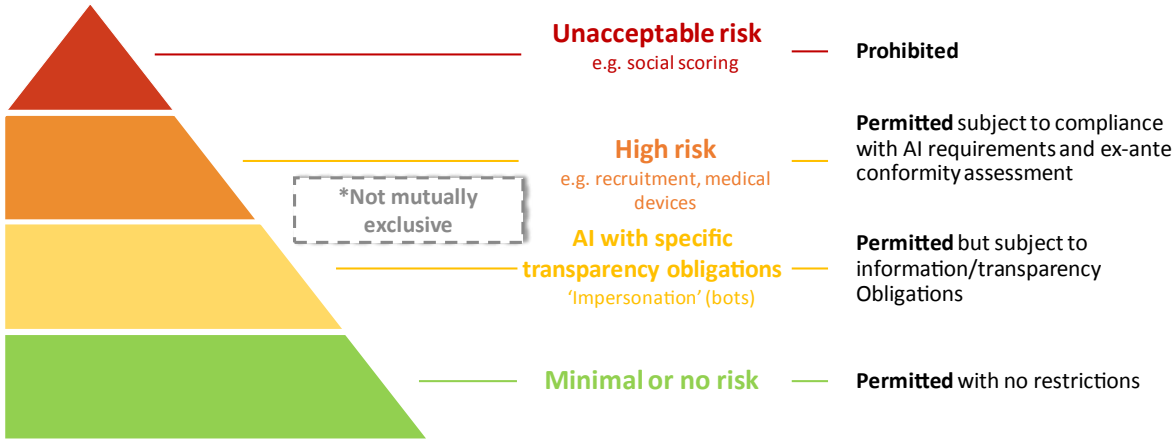
The National Institute of Standards and Technology (NIST) is driving public discussion about AI ethics and product standards in the U.S., alongside the IEEE[5] and IEC[6]. NIST is developing the Artificial Intelligence Risk Management Framework to better manage dangers to individuals, organizations, and society. The framework is intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems.

One size does not fit all: low and high-risk AI

U.S. industry – including both AI innovators and users – has lobbied heavily against the aforementioned bills. One reason for this opposition is the proposal to require algorithmic audits of all organizations. IBM, for example, has urged that this audit requirement be restricted to "high risk" AI use cases and not apply to "low risk" cases such as customer service chatbots. For now, there is no good definition of a "high risk" use case either in the U.S. or in the EU, and establishing clear criteria will be critical to the success of any future regulations. The draft EU AI Act identifies an initial list of high-risk use cases, which includes employment and workers management, education and vocational training, access to public services and benefits, and others.

The need for independent audits of AI systems is clear. The AI ethics and policy field is now building an ecosystem of audit criteria, standards, code of conduct, capacity building and enforcement mechanisms that will work to keep AI systems operating in an ethical manner that benefits humanity.

EU proposed legal framework for AI: Risk-based regulation approach



Source: European Commission[8]

Merve Hickok

Founder of AIethicist.org, is a globally renowned Artificial Intelligence (AI) policy, ethics, and governance expert. She has been recognized by several global organizations for her work - most recently as one of the 100 Brilliant Women in AI Ethics™ – 2021. She provides consulting and training to organizations on responsible and trustworthy AI and how to operationalize these concepts. She is also a data science ethics lecturer at the University of Michigan, and the Research Director at the Center for AI & Digital Policy, which contributes to the development of AI policy, advocating for trustworthy AI which protects fundamental rights, the rule of law, and democratic values. She helps build capacity towards similar goals and co-instructs AI policymakers, advocates, and practitioners from more than 50 countries. She also contributes to several global standard-setting bodies working toward Trustworthy AI.

Endnotes

[1] https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic
[2] https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence
[3] https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known
[4] https://techcrunch.com/2022/07/28/instagram-race-survey/
[5] Institute of Electrical and Electronics Engineers
[6] International Electrotechnical Commission
[7] https://www.aiethicist.org/
[8] https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf?

WASHINGTON | CORE

Author: Chris Wood, Senior Research Analyst

Washington CORE, L.L.C. is an independent consulting & research firm providing strategic research, analysis and advisory services. Founded in 1995, Washington CORE leverages in-depth research capabilities coupled with extensive global networks in both the public and private sectors, to deliver clarity and insight to prepare our clients for success in an ever-changing global landscape.

Please visit https://www.wcore.com for more information.