American Bar Association House of Delegates, Resolution 604 (adopted Feb. 6, 2023)

American Bar Association

AMERICAN BAR ASSOCIATION

ADOPTED BY THE HOUSE OF DELEGATES
FEBRUARY 6, 2023

RESOLUTION

RESOLVED, That the American Bar Association urges organizations that design, develop, deploy, and use artificial intelligence ("AI") systems and capabilities to follow these guidelines:

1) Developers, integrators, suppliers, and operators **("**Developers**")** of AI systems and capabilities should ensure that their products, services, systems, and capabilities are subject to human authority, oversight, and control;

2) Responsible individuals and organizations should be accountable for the consequences caused by their use of AI products, services, systems, and capabilities, including any legally cognizable injury or harm caused by their actions or use of AI systems or capabilities, unless they have taken reasonable measures to mitigate against that harm or injury; and

3) Developers should ensure the transparency and traceability of their AI products, services, systems, and capabilities, while protecting associated intellectual property, by documenting key decisions made with regard to the design and risk of the data sets, procedures, and outcomes underlying their AI products, services, systems and capabilities.

FURTHER RESOLVED, That the American Bar Association urges Congress, federal executive agencies, and State legislatures and regulators, to follow these guidelines in legislation and standards pertaining to AI.

REPORT

I. LEGAL ISSUES WITH AI

Artificial Intelligence ("AI") systems and capabilities create significant new opportunities for technological innovation and efficiencies to benefit our society, but they also raise new legal and ethical questions. AI enables computers and other automated systems to perform tasks that have historically required human cognition, such as drawing conclusions and making predictions.[1] AI systems operate at much faster speeds than humans.[2]

With AI and machine learning (ML)[3] already changing the way in which society addresses economic and national security challenges and opportunities, these technologies must be developed and used in a trustworthy and responsible manner. As private sector organizations and governments move rapidly to design, develop, deploy, and use AI systems and capabilities,[4] now is a critical time for the American Bar Association (ABA) to articulate principles that are essential to ensuring that AI is developed and deployed in accordance with the law and well-accepted legal standards.[5]

---

[1] AI is not a single piece of hardware or software, but rather a constellation of technologies that give a computer system the ability to solve problems and to perform tasks that would otherwise require human intelligence. National Security Commission on Artificial Intelligence (NSCAI), *Final Report*, Artificial Intelligence in Context, pages 31-40, https://www.nscai.gov/ [hereinafter "NSCAI Final Report"].
National Artificial Intelligence Research and Development Strategic Plan: 2019 Update (Nov. 12, 2020), https://catalog.data.gov/dataset/the-national-artificial-intelligence-research-and-development-strategic-plan-2019-update.

According to the National Institute of Standards and Technology (NIST), AI is:

> (1) A branch of computer science devoted to developing data processing systems that performs functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.
> (2) The capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement.

NIST *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools* (Aug. 2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[2] U.S. Government Accountability Office (GAO), *Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapons Systems*, GAO-22-104765 (Feb. 2022), https://www.gao.gov/assets/gao-22-104765.pdf. [hereinafter "GAO AI Report."]

[3] *Championing ethical and responsible machine learning through open-source best practices*, THE FOUNDATION FOR BEST PRACTICES IN MACHINE LEARNING, v. 1.0.0 (May 21, 2021), https://www.nist.gov/system/files/documents/2021/08/18/ai-rmf-rfi-0010-attachment3.pdf.

[4] NSCAI Final Report at 28, *supra* note 1. ("We now know the uses of AI in all aspects of life will grow and the pace of innovation will accelerate.")

[5] This Resolution does not purport to alter lawyers' obligations under applicable rules of professional conduct. Lawyers may wish to consider the issues raised in Daniel W. Linna Jr. and Wendy J. Muchma, Ethical Obligations to Protect Client Data when Building Artificial Intelligence Tools: Wigmore Meets AI (Oct. 2, 2020),

Fundamental concepts such as accountability, transparency, and traceability play an important role in ensuring the trustworthiness of AI systems. These concepts also play key roles in our legal system.[6] This Resolution presents guidance on how the legal system and its participants, including attorneys, regulators, and stakeholders, such as developers, integrators, suppliers, and operators **("**developers**")** of AI systems and capabilities, should assess these fundamental issues with AI. It states that in the context of AI, individual and enterprise accountability and human authority, oversight, and control are required and it is not appropriate to shift legal responsibility to a computer or an "algorithm" rather than to responsible people and other legal entities.

This Resolution will ensure that courts and participants in the legal process have the capacity to evaluate and resolve legal questions and disputes by specifying the essential information that must be included in the design, development, deployment, and use of AI to ensure transparency and traceability.

By focusing on these principles related to AI, this Resolution will help to ensure that accountability, transparency, and traceability are built into AI products, services, systems, and capabilities "by design" from the beginning of the development process. Following the proposed guidelines will enhance AI by maximizing the benefits from the use of AI in a trustworthy and responsible manner and help to minimize the risks.

Further, the Resolution urges Congress, federal executive agencies, and State legislatures and regulators to follow the guidelines in legislation and standards pertaining to AI.

II. OVERVIEW OF AI

AI holds great potential to bring innovation and efficiency across a number of industry sectors. New AI-enabled systems are benefitting many parts of society and the economy, from commerce and healthcare to transportation and cybersecurity. Consider just a few examples of recent AI innovations:

- Artificial intelligence is being deployed as a dialog agent for customer service. Several of these efforts have passed the Turing test – the eponymous idea developed by early computer pioneer Alan Turing which posited that the true test

---

https://www.americanbar.org/groups/professional_responsibility/publications/professional_lawyer/27/1/ethical-obligations-protect-client-data-when-building-artificial-intelligence-tools-wigmore-meets-ai/.
Risks to protect client confidentiality are present in the latest AI-augmented capabilities such as ChatGPT, and are heightened if counsel is unaware of the ways such capabilities involve human reviewers:

> "Ethical concerns arise because the conversations that happen within ChatGPT are not merely an exchange between a user and a computer program—humans are reviewing these ChatGPT conversations."

Foster J. Sayers, *ChatGPT and Ethics: Can Generative AI Break Privilege and Waive Confidentiality*, NYLJ (January 31, 2023), p. 3.

[6] Other important legal issues with AI have been identified, such as intellectual property infringement, algorithmic bias, access to justice, fairness in decision-making, discrimination, unfairness, and privacy and data protection/ cybersecurity. These issues may be appropriate for future ABA resolutions.

of computer intelligence will be met when individuals cannot tell the difference between a computer and a human interaction;

- Self-driving cars are under wide development by virtually every major manufacturer in the world (as well as most of the larger tech companies). While they are still in the testing stage, there is every reason to anticipate that geo-fenced cars will be on the market within 5-10 years;

- The AI product named Watson defeated the human champion in a game of Jeopardy and one named Alpha Go defeated the world Go champion;

- A system known as Deep Patient is now being deployed, successfully, as a diagnostic assistant to clinicians in a hospital setting, helping them make improved diagnoses in difficult cases. It is capable of predicting the onset of certain psychological diseases like schizophrenia in situations where the symptoms are not apparent to human clinicians;

- An artwork created by AI recently sold for over $400,000 at auction;

- More than two years ago a TV station in China began using an AI-powered announcer as the news anchor;

- Recent tests of autonomous self-directed weapons systems have successfully demonstrated that military systems can identify and target adversaries without human intervention; and

- New AI programs that go by the generic name of Deep Fakes can create fake video that can be virtually indistinguishable from reality.

Recently, governments and other organizations have been working on proposed AI governance frameworks and principles with the goal of mitigating the risks that can result through implementation of AI systems and capabilities. For example, NIST has developed an AI Risk Management Framework to provide guidance regarding the trustworthiness of AI systems.[7] Specifically, the framework is intended to help incorporate trustworthiness considerations into the design, development, use, and evaluation of AI systems, and it highlights accountability and transparency as two key guiding principles."[8]

The White House Office of Science and Technology Policy (OSTP) has acknowledged the "extraordinary promise of AI" as well as its pitfalls, and the need to "advance development, adoption, and oversight of AI in a manner that aligns with our democratic

[7] NIST *AI Risk Management Framework,* (AI RMF 1.0) NIST AI 100-1 (Jan. 2023), https://www.nist.gov/itl/ai-risk-management-framework [hereinafter "NIST AI Risk Management Framework"].
[8] *Id.* at 13.

values."[9] In recognition of the importance of ensuring that the American public has appropriate protections in the age of AI, OSTP released its Blueprint for an AI Bill of Rights "for building and deploying automated systems that are aligned with democratic values and protect civil rights, civil liberties, and privacy." [10] OSTP explained:

> Our country should clarify the rights and freedoms we expect data-driven technologies to respect. What exactly those are will require discussion, but here are some possibilities: your right to know when and how AI is influencing a decision that affects your civil rights and civil liberties; your freedom from being subjected to AI that hasn't been carefully audited to ensure that it's accurate, unbiased, and has been trained on sufficiently representative data sets; your freedom from pervasive or discriminatory surveillance and monitoring in your home, community, and workplace; and your right to meaningful recourse if the use of an algorithm harms you.[11]

III. ACCOUNTABILITY AND HUMAN OVERSIGHT, AUTHORITY, AND CONTROL

The ABA urges organizations that design, develop, deploy, and use AI systems and capabilities to follow these guidelines:

- Developers, integrators, suppliers, and operators ("developers") of AI systems and capabilities should ensure that their products, services, systems, and capabilities are subject to human authority, oversight, and control.

- Responsible individuals and enterprises should be accountable for the consequences caused by their use of AI products, services, systems, and capabilities, including any legally cognizable injury or harm caused by their use, unless they have taken reasonable measures to mitigate against that harm or injury.

Accountability and human authority, oversight and control are closely interrelated legal concepts. In the context of AI, they present key concerns, given that AI is increasingly being used in a variety of contexts to make decisions that can significantly impact

---

[9] L. Parker and R. Richardson, *OSTP's Continuing Work on AI Technology and Uses That Can Benefit Us All*, OSTP Blog (Feb. 3, 2022), https://www.whitehouse.gov/ostp/news-updates/2022/02/03/ostps-continuing-work-on-ai-technology-and-uses-that-can-benefit-us-all/.

[10] White House Office of Science and Technology Policy (OSTP), *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (October 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf. The Blueprint focuses on five principles for automated decision-making systems: (1) Safe and effective systems; (2) Algorithmic discrimination protections; (3) Data privacy; (4) Notice and explanation; and (5) Human alternatives, consideration and fallback.

[11] E. Lander & A. Nelson, *ICYMI: WIRED (Opinion): Americans Need a Bill of Rights For An AI-Powered World*, OTSP Blog (Oct. 22, 2022), https://www.whitehouse.gov/ostp/news-updates/2021/10/22/icymi-wired-opinion-americans-need-a-bill-of-rights-for-an-ai-powered-world/.
See, Ben Winters, AI Bill of Rights Provides Actionable Instructions for Companies, Agencies, and Legislators, EPIC (Oct. 11, 2022), https://epic.org/ai-bill-of-rights-leaves-actionable-instructions-for-companies-agencies-and-legislators/.

people's lives, including evaluating applicants for jobs, determining who receives access to loans, assessing criminal defendants' likelihood of being a repeat offender in connection with bail proceedings, screening rental applicants, and determining how self-driving cars should navigate through complex traffic and driving situations.

The Defense Advanced Research Projects Agency (DARPA) recently announced that it is starting a program to evaluate the use of AI to make complex decisions in modern military operations. DARPA explained that this In the Moment (ITM) program "aims to evaluate and build trusted algorithmic decision-makers for mission-critical Department of Defense (DoD) operations."[12]

Various organizations have recognized the importance of accountability with AI systems. In its AI Risk Management Framework (AI RMF 1.0), NIST stated that:

> Organizations need to establish and maintain the appropriate accountability mechanisms, roles and responsibilities, culture, and incentive structures for risk management to be effective. …
>
> Trustworthy AI depends upon accountability. Accountability presupposes transparency. *Transparency* reflects the extent to which information about an AI system and its outputs is available to individuals interacting with such a system – regardless of whether they are even aware that they are doing so. …
>
> When consequences are severe, such as when life and liberty are at stake, AI developers and deployers should consider proportionally and proactively adjusting their transparency and accountability practices.[13]

The Organization for Economic Cooperation and Development (OECD) Principles for AI includes Principle 1.5 on Accountability, which provides:

> Organizations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the OECD's values-based principles for AI.[14]

Australia has issued a voluntary framework of eight AI Ethics Principles which includes accountability, stating:

> People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.[15]

---

[12] *Developing Algorithms That Make Decisions Aligned With Human Expert*, DARPA Notice (March 3, 2022), https://www.darpa.mil/news-events/2022-03-03.

[13] NIST AI Risk Management Framework, at 9, 15, and 16, *supra* note 7,

[14] OECD AI Principles, https://oecd.ai/en/dashboards/ai-principles/P7. [hereinafter "OECD AI Principles."]

[15] *Australia's AI Ethics Principles, Principles at a Glance*,

In addition, large technology companies have also recognized the importance of accountability with regard to their AI products. For example, one of Microsoft's Six Principles for Responsible AI is accountability: "people should be accountable for AI systems."[16] Similarly, Google includes accountability in its Objectives for AI Applications, and states that AI should "be accountable to people. We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control."[17]

Human accountability is of particular importance given that with ML, a subset of AI, computers are able to learn from data sets without being given explicit instructions from humans. Instead, the computer model learns from experience and trains itself to find patterns and make predictions.[18] There has been widespread recognition of the critical role that humans should play in overseeing and implementing AI systems that are making such important decisions. For example, the term "human-centered artificial intelligence" has been used to describe the view that AI systems "must be designed with awareness that they are part of a larger system consisting of human stake-holders, such as users, operators, clients, and other people in close proximity."[19]

Accountability is important given the increasing concern about understanding AI decision-making and ensuring fairness in AI models, including with regard to the potential discriminatory impact of certain AI systems. For example, Amazon started a program to automate hiring by using an algorithm to review resumes. However, the program had to be discontinued after it was discovered that it discriminated against women in certain technical positions, such as software engineer, because the software analyzed the credentials of its existing employee base, which was predominantly male.[20] In addition, researchers found a gender and skin-type bias with commercial facial analysis programs, with an error rate of 0.8 percent for light-skinned men, versus 34.7 for dark-skinned women.[21]

There have been recent efforts to prohibit AI systems from violating anti-discrimination and privacy laws. For example, the Equal Employment Opportunity Commission (EEOC) launched an initiative to ensure that AI used in hiring and other employment

---

https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles.

[16] Microsoft *Responsible AI principles in practice*, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6, [hereinafter "Microsoft *Responsible AI Principles*"].

[17] Google *AI Principles*, https://ai.google/principles/.

[18] S. Brown, *Machine Learning Explained*, MIT Management: Ideas Made to Matter (April 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

[19] M. Riedl, *Human-Centered Artificial Intelligence and Machine Learning*, arXiv:1901.11184[cs.AI].

[20] J. Dastin, *Amazon Scraps Secret AI Recruiting Tool That Shows Bias Against Women*, Reuters (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[21] L. Hardesty, *Study Finds Gender and Skin-Type Bias in Commercial Artificial Intelligence Systems*, MIT NEWS (Feb. 11, 2018), https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-021.

decisions does not violate anti-discrimination laws.[22] New York City passed a new law to take effect in 2023 that prohibits the use of AI machine learning products in hiring and promotion decisions unless the tools have first been audited for bias.[23] In 2018, California passed the California Consumer Privacy Act (CCPA), a consumer protection law intended to protect the privacy of California residents. In 2020, it passed the California Privacy Rights Act (CPRA), amending the CCPA to add measures including the right to limit use and disclosure of sensitive personal information and the right to obtain information about how companies use automated decision-making technology.[24] In addition, questions have also been raised about the protection of privacy because of the processing of personal data in AI systems.[25]

Existing laws and regulations can be used to prevent potential violations of anti-discrimination and privacy laws by AI systems. For example, Federal Trade Commission (FTC) Commissioner Rebecca Kelly Slaughter explained her view that the FTC's existing tools, including section 5 of the FTC Act, the Equal Credit Opportunity Act, the Fair Credit Reporting Act, and the Children's Online Privacy Protection Act, can and should be used to protect consumers against algorithmic harms.[26]

In light of the need to ensure compliance with laws and regulations being used to prevent harms from AI systems, it is essential that the humans and enterprises with responsibility for these AI systems be held accountable for the consequences of the uses of these systems.

Under our legal system, in order to be held accountable, an entity must have a specific legal status that allows it to be sued, such as being an individual human or a corporation. On the other hand, property, such as robots or algorithms, does not have a comparable legal status.[27] Thus, it is important that legally recognizable entities such as humans and corporations be accountable for the consequences of AI systems, including any legally cognizable injury or harm that their actions or those of the AI systems or

---

[22] EEOC Artificial Intelligence and Algorithmic Fairness Initiative (2021), https://www.eeoc.gov/ai; EEOC *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees,* https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

[23] N. Lee and S. Lai, *Why New York City Is Cracking Down on AI in Hiring*, BROOKINGS TECHTANK (Dec. 20, 2021), https://www.brookings.edu/blog/techtank/2021/12/20/why-new-york-city-is-cracking-down-on-ai-in-hiring/.

[24] B. Justice, *CPRA Countdown: It's Time to Brush Up on California's Latest Data Privacy Law*, NATIONAL LAW REVIEW (Dec. 18, 2021), https://www.natlawreview.com/article/cpra-countdown-it-s-time-to-brush-california-s-latest-data-privacy-law.

[25] C. Tucker, *Privacy, Algorithms and Artificial Intelligence*, in *The Economics of Artificial Intelligence: An Agenda*, NATIONAL BUREAU OF ECONOMIC RESEARCH (2019), https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/privacy-algorithms-and-artificial-intelligence.

[26] R. Slaughter, *Algorithms and Economic Justice*, ISP DIGITAL FUTURE WHITEPAPER & YALE JOURNAL OF LAW & TECHNOLOGY SPECIAL PUBLICATION (Aug. 2021)

[27] Michalski, Roger (2018), *How to Sue a Robot*, UTAH LAW REVIEW: Vol. 2018: No. 5, Article 3, https://dc.law.utah.edu/ulr/vol2018/iss5/3.

capabilities cause to others, unless they have taken reasonable measures to mitigate against that harm or injury.[28]

IV. TRANSPARENCY AND TRACEABILITY

The ABA urges organizations that design, develop, deploy, and use artificial intelligence ("AI") products, services, systems and capabilities to follow this guideline:

- Developers should ensure the transparency and traceability of their AI products, services, systems, and capabilities, while protecting associated intellectual property, by documenting key decisions made with regard to the design and risk of the data sets, procedures, and outcomes underlying their AI products, services, systems, and capabilities.

A. Transparency

In the context of AI, transparency is about responsible disclosure to ensure that people understand when they are engaging with an AI system, product, or service and enable those impacted to understand the outcome and be able to challenge it if appropriate.[29] NIST stated that "explainable AI" is one of several properties that characterize trust in AI systems.[30]

---

[28] In developing rules of liability, the supplier/component part doctrine would apply. Under that doctrine, the manufacturer of a non-defective component is not liable for harm caused by a defect in a larger system sold by a manufacturer into which the component was integrated.

[29] OEDC adopted Transparency and Explainability Principle 1.3 that states:

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

- to foster a general understanding of AI systems,
- to make stakeholders aware of their interactions with AI systems, including in the workplace,
- to enable those affected by an AI system to understand the outcome, and,
- to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

OECD AI Principles, *supra* note 12.

[30] NIST *Artificial Intelligence*, https://www.nist.gov/artificial-intelligence; NIST *Four Principles of Explainable Artificial Intelligence*, NIST Interagency/Internal Report (NISTIR) - 8312, https://doi.org/10.6028/NIST.IR.8312.

Four principles of explainable AI – for judging how well AI decisions can be explained:

- *Explanation* – AI systems should deliver accompanying evidence or reasons for all their outputs.
- *Meaningful* – Systems should provide explanations that are meaningful or understandable to individual users.
- *Explanation Accuracy* – The explanation correctly reflects the system's process for generating the output.
- *Knowledge Limits* – The system only operates under conditions for which it was designed or when the system reaches a sufficient confidence in its output. (The idea is that if a system has insufficient confidence in its decision, it should not supply a decision to the user.)

See, https://www.nist.gov/artificial-intelligence/ai-fundamental-research-explainability.

Lack of transparency with AI can negatively affect individuals who are denied jobs, refused loans, refused entry or are deported, imprisoned, put on no-fly lists or denied benefits. They are often not informed of the reasons other than the decision was processed using computer software. Human rights principles that may be impacted are rights to a fair trial and due process, effective remedies, social rights and access to public services, and rights to free elections. [31]

OECD has explained that the term transparency carries multiple meanings:

> In the context of this Principle [1.3], the focus is first on disclosing when AI is being used (in a prediction, recommendation or decision, or that the user is interacting directly with an AI-powered agent, such as a chatbot). Disclosure should be made with proportion to the importance of the interaction. The growing ubiquity of AI applications may influence the desirability, effectiveness or feasibility of disclosure in some cases.
>
> Transparency further means enabling people to understand how an AI system is developed, trained, operates, and deployed in the relevant application domain, so that consumers, for example, can make more informed choices. Transparency also refers to the ability to provide meaningful information and clarity about what information is provided and why. Thus transparency does not in general extend to the disclosure of the source or other proprietary code or sharing of proprietary datasets, all of which may be too technically complex to be feasible or useful to understanding an outcome. Source code and datasets may also be subject to intellectual property, including trade secrets.
>
> An additional aspect of transparency concerns facilitating public, multi-stakeholder discourse and the establishment of dedicated entities, as necessary, to foster general awareness and understanding of AI systems and increase acceptance and trust.

Numerous organizations around the world have developed AI principles. A researcher who reviewed them reported that "[f]eatured in 73/84 sources, transparency is the most prevalent principle in the current literature."[32] Varied terminology is used to express this concept of transparency, comprising efforts to increase explainability, interpretability, intelligibility or other acts of communication and disclosure.

---

[31] Rowena Rodrigues, *Legal and human rights issues of AI: Gaps, challenges and vulnerabilities*, JOURNAL OF RESPONSIBLE TECHNOLOGY, Vol. 4, Dec. 2020, 100005, https://doi.org/10.1016/j.jrt.2020.100005.

[32] Anna Jobin, *et. al., Artificial Intelligence: the global landscape of ethics guidelines*, HEALTH ETHICS & POLICY LAB, ETH Zurich, 8092 Zurich, Switzerland (2019), https://www.researchgate.net/profile/Anna-Jobin/publication/334082218_Artificial_Intelligence_the_global_landscape_of_ethics_guidelines/links/5d19ec7d299bf1547c8d2be8/Artificial-Intelligence-the-global-landscape-of-ethics-guidelines.pdf?origin=publication_detail.

European Union member state reports on AI can be found at https://futurium.ec.europa.eu/en/european-ai-alliance/pages/official-documents-and-reports.

Intelligibility can uncover potential sources of unfairness, help users decide how much trust to place in a system, and generally lead to more usable products. It also can improve the robustness of AI systems by making it easier for data scientists and developers to identify and fix bugs.[33]

The FTC published guidance regarding the commercial use of AI technology, acknowledging that while AI has significant positive potential, it also presents negative risks, such as unfair or discriminatory outcomes or the entrenchment of existing disparities.[34] The FTC urged companies to:

- Be transparent with consumers;
- Explain how algorithms make decisions;
- Ensure that decisions are fair, robust, and empirically sound; and
- Hold themselves accountable for compliance, ethics, fairness and non-discrimination.

B. Traceability

It is important to ensure that the complex processes in data science — from data processing through modeling to deployment in production — can be documented in a way that is understood easily.[35] Traceability is considered a key requirement for trustworthy AI. It would allow companies to better understand the entire reasoning process, and builds trust with AI implementations.[36]

According to NIST, "[t]rustworthy AI refers to AI capabilities that exhibit characteristics such as resilience, security, and privacy so that relevant people can adopt them without fear."[37] An AI capability must be traceable, meaning that it is developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including

---

[33] Microsoft *Responsible AI principles, supra* note 14. Microsoft Research Collection: *Research Supporting Responsible AI* (April 13, 2020), https://www.microsoft.com/en-us/research/blog/research-collection-research-supporting-responsible-ai/.

[34] FTC *Using Artificial Intelligence and Algorithms* (April 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms; FTC, *Aiming for truth, fairness, and equity in your company's use of AI* (April 19, 2021), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

[35] Andreas Gödde, *Traceability for Trustworthy AI: A Review of Models and Tools*, SAS, https://www.mdpi.com/2504-2289/5/2/20/htm.
https://blogs.sas.com/content/hiddeninsights/2018/03/12/interpretability-traceability-clarity-ai-mandate/.
See, Association for Computing Machinery, *Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems,* FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (March 2021), pages 758–771, https://dl.acm.org/doi/10.1145/3442188.3445937.

[36] Sanjay Srivastava, *The path to explainable AI*, CIO (May 21, 2018), https://www.cio.com/article/221668/the-path-to-explainable-ai.html.

[37] NIST, Draft –Taxonomy of AI Risk (Oct. 2021), https://www.nist.gov/system/files/documents/2021/10/15/taxonomy_AI_risks.pdf*; see* GAO AI Report, *supra* note 2.

with transparent and auditable methodologies, data sources and design procedures and documentation.[38]

C. Documenting key decisions made with regard to the design and risk of the data sets, procedures, and outcomes.

As AI algorithms become more complex, the need for greater transparency grows. Experts are developing software tools that will address the "black box" problem[39] – not knowing how algorithms arrive at their final output – by analyzing complex AI systems and documenting how the system processes information, answers questions, and provides results.[40]

Traceability is related to the need to maintain a complete account of the provenance of data, processes, and artifacts involved in the production of an AI model – and it should encompass all elements of an AI system, product or service, namely the data, the system, and the business model. It requires documentation of the data sets, procedures, and outcomes for the AI system or capability.[41]

Practical Considerations – In establishing traceability for AI products, services, systems, and capabilities, developers should create contemporaneous records that document key decisions made with regard to the design and risk of the AI data sets. This means using automated tools when appropriate and available, or otherwise using documentation techniques (online or manual) appropriate for the software development lifecycle and for

---

[38] The Department of Defense (DoD) adopted *5 Principles of Artificial Intelligence Ethics* that commits the Department to this principle of traceability. U.S. Department of Defense, *5 Principles of Artificial Intelligence Ethics*, https://www.defense.gov/News/News-Stories/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/. *See AI Principles: Recommendations on the Ethical Use of Artificial Intelligence* by the Department of Defense, Defense Innovation Board, *available at* https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF.
Similarly, the *Principles of Artificial Intelligence Ethics for the Intelligence Community*[38] provide:

> *Transparent and Accountable* – We will provide appropriate transparency to the public and our customers regarding our AI methods, applications, and uses within the bounds of security, technology, and releasability by law and policy, and consistent with the Principles of Intelligence Transparency for the IC. We will develop and employ mechanisms to identify responsibilities and provide accountability for the use of AI and its outcomes.

[39] Cliff Kuang, *Can A.I. Be Taught to Explain Itself*? THE NEW YORK TIMES MAGAZINE (Nov. 21, 2017), https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html
[40] Neil Savage, *Breaking into the black box of artificial intelligence: Scientists are finding ways to explain the inner workings of complex machine-learning models*, NATURE (Mar. 29, 2022), https://www.nature.com/articles/d41586-022-00858-1.
[41] The assessment for traceability includes:

- *Procedures*: Methods used for designing and developing the algorithmic system: how the algorithm was trained, which input data was gathered and selected, and how this occurred.
- *Data*: Methods used to test and validate the algorithmic system: information about the data used to test and validate.
- *Outcomes*: The outcomes of the algorithms or the subsequent decisions taken on the basis of these outcomes, as well as other potential decisions that would result from different cases (*e.g.,* for other subgroups of users).

conducting AI risk assessments. Computer scientists are developing data models and tools to fully document data, procedures and outcomes for AI systems. They enable some form of automated repetition of the construction of the artifacts.[42]

Examples of the types of key decisions to be documented throughout the AI lifecycle include:

- *Business* – business-oriented requirements, expected uses and outcomes, key performance features (including when AI is used or relied upon in decision making). Human control over the selection of inputs and generation of outputs in order to reduce the risks of unintended adverse consequences.
- *Data* – types, quantities, and sources of data to be used in training the AI systems and capabilities; modeling, analysis, evaluation.[43]
- *AI risk assessment* – risks assessed, unintended bias, or hazardous use.
- *Cybersecurity risks* – risks of unauthorized access to, and compromise of the integrity of, the AI algorithms, software, training data, and/or model.
- *Design and development* – key design trade-offs, risks mitigated by the design. Review of algorithm(s), software code and the AI model.
- *Testing* – involvement of humans with detailed understanding of AI processes and industry domain issues. Testing of implementing software, model with data sets, and adjustments and correction of errors. Problems observed in generating desired outputs. Performance deficiencies, malfunctions, unintended outputs, and discovered risks observed.
- *Deployment*
- *Developers should respond promptly* to avert or mitigate AI risks that are identified at any point in the AI system/product life cycle.

In the event of a gap between actual and desired performance with an AI system, capability, product, or service, recurring errors or failures with specific processes and undesirable events reoccurring, traceability will enable root cause analysis, a process for understanding 'what happened' and solving a problem through looking back and drilling down to find out 'why it happened' in the first place. Then, looking to rectify the issue(s) so that it does not happen again, or reduce the likelihood that it will happen again.[44]

The many benefits of root cause analysis include reducing risk and preventing recurring failures, improving performance, as well as the potential for cost reduction. It provides a logical approach to problem solving using data that already exist and a learning process

---

[42] *Traceability for Trustworthy AI: A Review of Models and Tools*, https://www.mdpi.com/2504-2289/5/2/20/htm.

[43] The key is to fully understand the data's behavior. Best practices include documenting assumptions around completeness of the data, addressing data biases, and reviewing new rules identified by the machine before implementing. If AI is being used to identify anomalies, companies can put checks and balances in place to manually test and determine if the results make sense.

[44] Chartered Institute of Internal Auditors, *Root Cause Analysis* (Sept. 22, 2020), https://www.iia.org.uk/resources/delivering-internal-audit/root-cause-analysis?downloadPdf=true.

for better understanding of relationships, causes and effect, and solutions. The process should lead to more robust AI systems and capabilities.

V. EXISTING ABA POLICY

The ABA House of Delegates passed two Resolutions that address AI. This Resolution builds on and is consistent with those existing ABA policies.

- ABA urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ("AI") in the practice of law, including (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI. 19A112.

- ABA urges federal, state, local, territorial and tribal governments to:
  - Ensure due process and refrain from using pretrial risk assessment tools unless the data supporting the risk assessment is transparent, publicly disclosed, and validated; and
  - Recognize that an individual's criminal history and other criteria may reflect structurally biased application of laws, policies or practices, as well as conscious or unconscious bias. 22M700.

VI. CONCLUSION

This Resolution addresses important legal issues concerning AI by focusing on the principles of accountability, transparency and traceability. It states that in the context of AI, human and enterprise accountability and human authority, oversight, and control are required and it is not appropriate to shift legal responsibility to a computer or an "algorithm" rather than to responsible people and other legal entities.

It will ensure that courts and participants in the legal process have the capacity to evaluate and resolve legal questions and disputes by specifying the essential information that must be included in the design, development, deployment, and use of AI to ensure transparency and traceability. Passage of this Resolution will enhance AI by maximizing the benefits from the use of AI in a trustworthy and responsible manner and help to minimize the risks.

Respectfully Submitted,

Claudia Rast and Maureen Kelly, Co-Chairs
Cybersecurity Legal Task Force

February 2023

APPENDIX

LAWS, COURT DECISIONS, AND LEADING REPORTS

An exhaustive analysis of federal, state, and international laws applicable to AI is outside the scope of this Report. Below are some of the highlights:

National Conference of State Legislatures (NCLS) ***State AI Legislation***
https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx

General AI bills or resolutions were introduced in at least 17 states in 2021-22, and were enacted in Alabama, Colorado, Illinois, Mississippi, Vermont, and Washington.

General Data Protection Regulation (GDPR) Article 22 – AI Requirements[45]

GDPR imposes legal requirements on whoever uses an AI system for profiling and/or automated decision-making (regardless of the *means* by which personal data are processed), even if they acquired the system from a third party. These requirements include Fairness; Transparency, including meaningful information about the logic involved in the AI system; and the right to human intervention, enabling the individual to challenge the automated decision.

Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) (25 November 2022), approved by the Council on December 6, 2022.

2021/0106(COD), https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

The Regulation introduces new obligations for vendors of AI systems, and includes requirements for high-risk AI systems and users.

European Parliament, The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, PE 641.530 (June 2020),
https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.

Holbrook v. Prodomax Automation Ltd.**,** 2021 U.S. Dist. LEXIS 178325 (Sept. 20, 2021) U.S. Dist. Ct., W.D. Mich.

[45] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) (OJ L 119 04.05.2016, p. 1, CELEX: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679).

Man Whose Wife Was Killed by Factory Robot Settles Mid-Trial, BLOOMBERG (Nov. 9, 2021), https://news.bloomberglaw.com/product-liability-and-toxics-law/man-whose-wife-was-killed-by-factory-robot-settles-mid-trial.
Eric L. Alexander, *Unintended Consequences for Software Liability*? REED SMITH (Nov. 26, 2021), https://www.lexology.com/library/detail.aspx?g=54e4a579-500d-4db0-adc2-065bc9b06263.

Leading Reports

White House Office of Science and Technology Policy (OSTP), *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (October 2022)
https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

The Blueprint focuses on principles for automated decision-making systems: (1) Safe and effective systems; (2) Algorithmic discrimination protections; (3) Data privacy; (4) Notice and explanation; and (5) Human alternatives, consideration and fallback.

National Security Commission on Artificial Intelligence (NSCAI), *Final Report*
https://www.nscai.gov/.

Presents the strategy for the U.S. to win in the AI era by responsibly using AI for national security and defense, defending against AI threats, and promoting AI innovation. *Blueprints for Action* provide plans to implement the recommendations.

House Committee on Transportation and Infrastructure

Boeing 737 MAX Investigation, https://transportation.house.gov/committee-activity/boeing-737-max-investigation.
Final Committee Report on the Design, Development, and Certification of the Boeing 737 MAX (Sept. 2020).

NIST AI Risk Management Framework: Second Draft (August 2022)
https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.

Intended for voluntary use "in addressing risks in the design, development, use, and evaluation of AI products, services, and systems."

Artificial Intelligence and the Courts: Materials for Judges, American Association for the Advancement of Science (AAAS) (Sep. 2022)
https://www.aaas.org/ai2/projects/law/judicialpapers.

604

With the support of NIST, this AAAS project is developing resources to support judges as they address an increasing number of cases involving AI.

Stanford HAI, *Artificial Intelligence Index Report 2021*, Stanford Human-Centered Artificial Intelligence
https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf.

Presents unbiased, globally sourced data that will enable policy-makers, researchers, executives, and the public to develop intuitions about AI.

Industry IoT Consortium, *Industrial IoT Artificial Intelligence Framework* (Feb. 22, 2022)
https://www.iiconsortium.org/pdf/Industrial-AI-Framework-Final-2022-02-21.pdf.

Provides guidance in the development, training, documentation, communication, integration, deployment, and operation of AI-enabled industrial IoT systems.

OECD AI Principles (May 2019)
https://oecd.ai/en/ai-principles.

Promotes the use of innovative and trustworthy AI and respects human rights and democratic values.

European Commission, European AI Alliance
https://futurium.ec.europa.eu/en/european-ai-alliance/pages/official-documents-and-reports.

Council of Europe, Karen Yeung, Responsibility and AI, DGI(2019)05
https://rm.coe.int/responsability-and-ai-en/168097d9c5.

A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework.

Katherine B. Forrest, When Machines Can Be Judge, Jury, And Executioner: Justice In The Age Of Artificial Intelligence (2021)

GENERAL INFORMATION FORM

Submitting Entity: Cybersecurity Legal Task Force

Submitted By: Claudia Rast and Maureen Kelly, Co-chairs

1. Summary of Resolution(s).

This Resolution presents guidance on how the legal system and its participants, including attorneys, regulators, and stakeholders – developers, integrators, suppliers, and operators **("**developers**")** of AI systems and capabilities – should assess three fundamental issues with AI: accountability, transparency and traceability.

The Resolution will ensure that courts and participants in the legal process have the capacity to evaluate and resolve legal questions and disputes by specifying the essential information that must be included in the design, development, deployment, and use of AI to ensure transparency and traceability.

2. Indicate which of the ABA's four goals the resolution seeks to advance (1-Serve our Members; 2-Improve our Profession; 3-Eliminate Bias and Enhance Diversity; 4-Advance the Rule of Law) and provide an explanation on how it accomplishes this.

This Resolution meets Goal 4 – Advance the Rule of Law. The Resolution is designed to help mitigate the risks that can result through implementation of AI systems and capabilities and enhance the use of AI in a trustworthy and responsible manner.

3. Approval by Submitting and Co-sponsoring Entities.

The Cyberspace Legal Task Force voted to sponsor this Resolution on December 2, 2022.
The Antitrust Law Section voted to co-sponsor this Resolution on December 2, 2022.
The Tort, Trial & Insurance Practice (TIPS) Section voted to co-sponsor this Resolution on November 16, 2022.
The Science & Technology Law Section voted to co-sponsor this Resolution on December 20, 2022.
The Standing Committee on Law and National Security voted to co-sponsor this Resolution on November 19, 2022.

4. Has this or a similar resolution been submitted to the House or Board previously? No.

5. What existing Association policies are relevant to this resolution and how would

they be affected by its adoption?

The ABA House of Delegates has passed resolutions that address issues with AI. This Resolution builds on and is consistent with those ABA policies.

- ABA urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ("AI") in the practice of law, including (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI. 19A112.

- ABA urges federal, state, local, territorial and tribal governments to:
  - Ensure due process and refrain from using pretrial risk assessment tools unless the data supporting the risk assessment is transparent, publicly disclosed, and validated to demonstrate the absence of conscious or unconscious racial, ethnic, or other demographic, geographic, or socioeconomic bias; and
  - Recognize that an individual's criminal history and other criteria may reflect structurally biased application of laws, policies or practices, as well as conscious or unconscious bias. 22M700.

6. If this is a late report, what urgency exists which requires action at this meeting of the House?

This is not a late report. As private sector organizations and governments move rapidly to design, develop, deploy, and use AI systems and capabilities, now is a critical time for lawyers to articulate principles that are essential to ensuring that AI is developed and implemented in accordance with the law and well-accepted legal standards.

7. Status of Legislation. (If applicable)

S. 1605, FY 2022 National Defense Authorization Act – enacted
Legislation to strengthen the U.S. government's artificial intelligence (AI) readiness, support long-term investments in AI ethics and safety research, and increase governmental AI transparency, were passed as part of the FY 2022 *National Defense Authorization Act (NDAA)*.

Artificial Intelligence Capabilities and Transparency (AICT) Act.
The A/CT Act would implement recommendations of the National Security Commission on Artificial Intelligence's (NSCAI) final report. Congress established the NSCAI through the FY 2019 *National Defense Authorization Act (NDAA)* in order to consider the methods and means necessary to advance the development and improve the government's use of AI and related technology.

S. 2551 — Artificial Intelligence Training for the Acquisition Workforce Act or the AI Training Act
This bill requires the Office of Management and Budget (OMB) to establish or otherwise provide an AI training program for the acquisition workforce of executive agencies (*e.g.*, those responsible for program management or logistics) to ensure that the workforce has knowledge of the capabilities and risks associated with AI.

U.S. States

General AI bills or resolutions were introduced in at least 17 states in 2021-22, and were enacted in Alabama, Colorado, Illinois, Mississippi, Vermont, and Washington.

National Conference of State Legislatures (NCLS), *State AI Legislation,* https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx.

8. Brief explanation regarding plans for implementation of the policy, if adopted by the House of Delegates.

This Resolution will be disseminated to members of Congress and State legislators in coordination and cooperation with the ABA Governmental Affairs Office, as well as executives of large and small companies that design, develop, deploy, and use AI systems, capabilities, products, and services.

It will alert them to the ABA's newly-adopted policy and encourage them to take action consistent with the ABA policy. We also encourage its use in Amicus Curiae briefs by the ABA.

9. Cost to the Association. (Both direct and indirect costs).
None.

10. Disclosure of Interest. (If applicable)
Not Applicable.

11. Referrals.

Sections:
Business Law
Civil Rights & Social Justice
Criminal Justice
Environment, Energy & Resources
Intellectual Property
International Law
Litigation
Public Contract Law
Science & Technology Law

604

State and Local Government Law
Tort, Trial & Insurance Practice

Standing Committees:
Cybersecurity Legal Task Force
Professional Responsibility

Divisions:
Young Lawyers
Senior Lawyers
Law Practice

12. Contact Name and Address Information. (Prior to the meeting)

Lucy L. Thomson, Delegate, District of Columbia Bar
Livingston PLLC, Washington, D.C.
lucythomson1@mindspring.com, (703) 798-1001

Roland Trope
Trope Law, New York, New York
rltrope@tropelaw.com, (917) 370-3705

13. Contact Name and Address Information. (Who will present the report to the House?)

Lucy L. Thomson, Delegate, District of Columbia Bar
Livingston PLLC, Washington, D.C.
lucythomson1@mindspring.com, (703) 798-1001

EXECUTIVE SUMMARY

1. Summary of the Resolution

This Resolution presents guidance on how the legal system and its participants, including attorneys, regulators, and stakeholders, such as developers, integrators, suppliers, and operators **(“**developers**”)** of AI systems and capabilities, should assess fundamental issues with AI by addressing the principles of accountability, transparency and traceability.

2. Summary of the Issues that the Resolution Addresses

This Resolution states that in the context of AI individual and enterprise accountability and human authority, oversight, and control is required and it is not appropriate to shift legal responsibility to a computer or an “algorithm” rather than to responsible people and other legal entities.

By focusing in the context of AI on the key issues accountability, transparency and traceability, passage of this Resolution will help mitigate the risks that can result through implementation of AI systems and capabilities and enhance the use of AI in a trustworthy and responsible manner.

3. Please Explain How the Proposed Policy Position Will Address the Issue

This Resolution presents guidance on how the legal system and its participants, including attorneys, regulators, and stakeholders, including developers, integrators, suppliers, and operators **(“**developers**”)** of AI systems and capabilities, should assess fundamental issues with AI by addressing the principles of accountability, transparency and traceability. It states that in the context of AI individual and enterprise accountability and human authority, oversight, and control is required and it is not appropriate to shift legal responsibility to a computer or an “algorithm” rather than to responsible people and other legal entities.

Further, this Resolution would ensure that courts and participants in the legal process will have the capacity to evaluate and resolve legal questions and disputes by specifying the essential information that must be included in the development, deployment and use of AI to ensure transparency and traceability.

4. Summary of Minority Views

None.