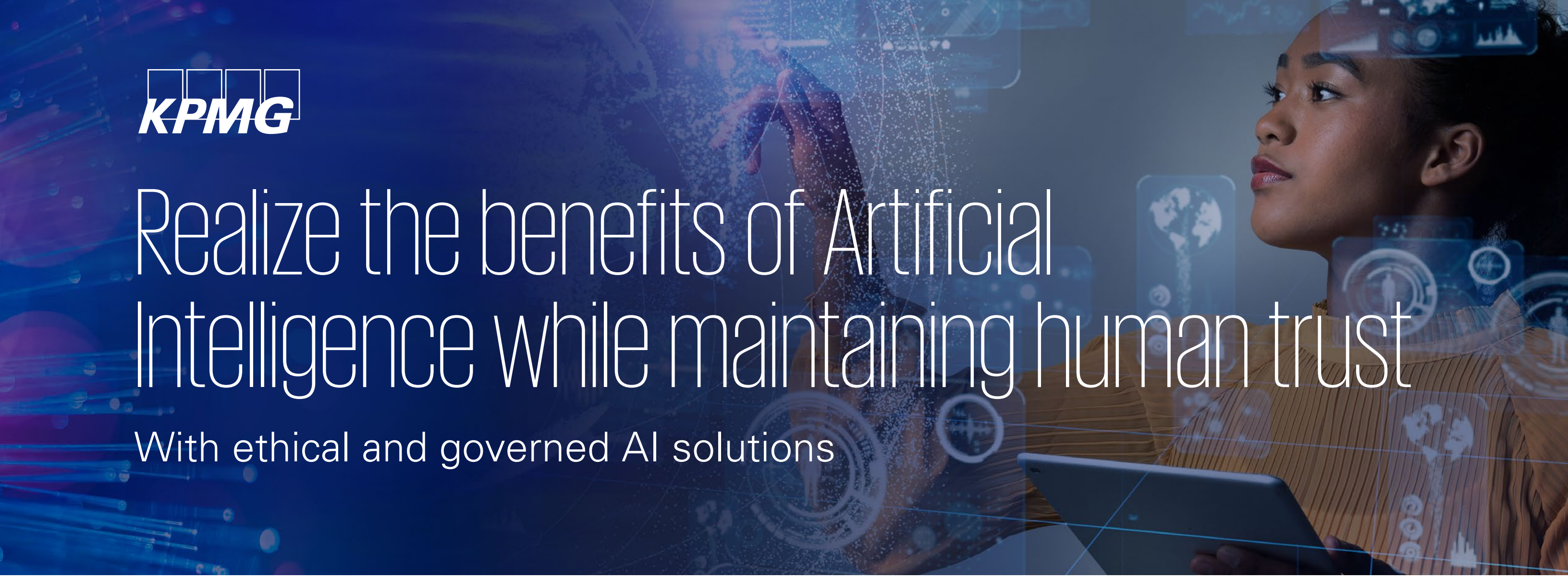


Realize the benefits of Artificial Intelligence while maintaining human trust

With ethical and governed AI solutions

Implement Artificial Intelligence in an ethical, trustworthy manner

Artificial Intelligence has the power to not only be transformative, but completely world-changing. Algorithms which continuously learn and evolve pave the way for rapid innovation with many benefits for business and society as a whole.

But alongside the potential benefits, AI can also pose considerable risks and challenges to society. These risks have raised concerns about whether such powerful AI systems are worthy of trust. In a KPMG study *(Trust in Artificial Intelligence - a five-country study)*, almost all citizens (95%) expect AI systems to meet high standards of performance and ethics, and the majority (more than 57%) would be more willing to use AI systems if assurance mechanisms were in place. But only around a quarter (28%) of citizens are willing to trust AI systems in general. Particularly, only 17% of US citizens have high or complete confidence in federal or state government to develop and use AI in the best interests of the public. To realize the full potential of AI and all its benefits, steps must be taken to build and maintain public trust.

Among the many principles of trustworthy AI, fairness, transparency and explainability, data privacy, and human oversight are the key factors directly related to building and maintaining a framework for ethical AI. This article explores key questions surrounding ethical AI, including why ethics is such an important factor in AI adoption, which decisions should be handed over to machines, and when these decisions are handed over, how to ensure our well-defined ethics regarding fundamental human values are consistently met.

Why modern government is important

Government agencies in the U.S. must modernize in order to keep up with changing user needs, regulations, and health and public safety requirements. Leaders of modern governments rethink business processes and service delivery models to more effectively achieve their mission. This article is one of a series that features how modernizing affects the government workforce and the user experience, improves security and public trust, and accelerates the digital journey. KPMG team members offer insights intended to help guide governments in their modernization efforts to encompass all processes, technologies, policies, and the workforce so each works together to create connected, powered, and trusted organizations.

Why is Ethical AI so important?

AI is an increasingly ubiquitous part of our everyday lives, and continues to transform how we live and work day-to-day. With the potential power and scale of AI, and its ability to make autonomous decisions based on evolving algorithms, ensuring artificial intelligence and the algorithms it relies on are built upon an ethical foundation is vital. Ethical AI focuses on the fundamental human values, such as including individual rights, privacy, non-discrimination, and non-manipulation, associated with artificial intelligence, and is a specific piece in the broader picture of Trusted AI.

If used unethically - or even incautiously - AI can cause severe harmful effects for individuals, the environment, and society as a whole. Without considering ethical implications, artificial intelligence can cause biased hiring decisions, privacy violations when facial detection is used for surveillance, and the potential ethical consequences of implementing predictive policing, to name just a few examples.

For government and agencies, ethical AI is even more critical, as the majority of the AI systems in use would involve human lives - directly and indirectly - either to share information with AI systems, or to be impacted by results from AI. On the other hand, as policy-makers, the government needs to be the role model in order to set examples for the industry on practising ethical AI, especially with the absence of AI laws and regulations.

AI solutions that aren't ethical ultimately won't be trustworthy. And if it is not trustworthy, widespread acceptance and adoption will be hindered, and the benefits of AI will not be fully realised.

Continuous monitoring and mitigation: ensuring unbiased processes and decisions across the entire AI lifecycle

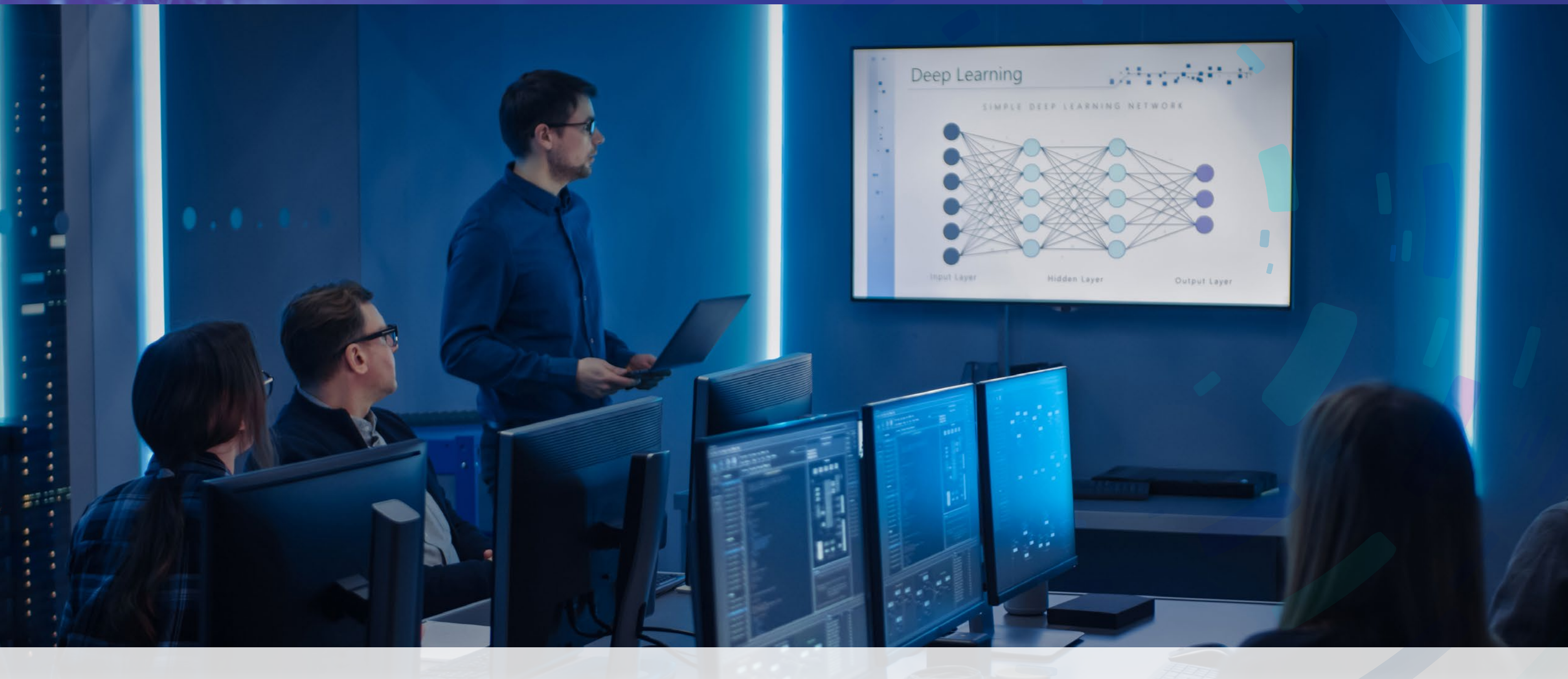
Building AI and algorithms as free from bias as possible is crucial for maintaining ethical practices in artificial intelligences, and this fairness must be maintained as they continue to learn and evolve. Governments and agencies must be confident about how decisions are made using AI, and whether these decisions are fair and accurate, in order to avoid undermining the trust in AI. One real-world example is the risk assessment software COMPAS, used to forecast future criminals and guide judges in the courtrooms to determine everything from bail amounts to sentences.[1] The algorithm was found to be almost twice as likely to label blacks as a higher risk, but not actually re-offend; and the algorithm made the opposite mistake among whites. What should we do to handle bias in AI systems better?

Fairness should be considered not only in the algorithms which AI rely on, but in a government or agency's processes. The organization shall proactively identify and document inherent bias in the data, features and inference results, as well as focusing on understanding, documenting and monitoring bias in the development and production.

The ability to govern ethics becomes a key factor for the responsible adoption and scaling of AI. Appropriate technologies and tools for continuous monitoring and governance are essential to help ensure models are continuously trained to learn from data, while ensuring neither the original data nor feedback data cause bias to creep in. While end-to-end automation may be the ultimate goal of operationalizing AI, organizations must address the risk that complex algorithms could take a wrong turn, assess the impact of unfair predictions and, where necessary, design systems with human-in-the-loop review processes.

Once the bias is detected, debiasing technical tools and approaches should be applied to avoid and mitigate the problem. For example, the AI Fairness 360 library (AIF360) by IBM and Amazon SageMaker Clarify provides a range of state-of-the-art bias mitigation techniques that enable the developer or data scientist to reduce any discovered bias.

[1] https://www.technologyreview.com/2017/06/12/105804/inspecting-algorithms-for-bias/



Ethical AI solutions can be unlocked through introducing transparency at a foundational level

Citizens want to know more about AI, but currently have low awareness and understanding of AI and its uses, according to the KPMG article *Trust in Artificial Intelligence - a five-country study*. Various stakeholders involved in the lifecycle of AI systems require a different type of understanding, of why, why not, and how. For example, when a government agency uses AI models to decide which citizens' benefit applications would be approved or denied, caseworks might wonder why a citizen's benefits application was rejected. Meanwhile, a compliance department might want to know how the model is working across demographics, and a front office worker may be interested in which variables to validate on borderline decisions. In order to build an ethical AI model in which the output can be trusted, the "black-box model" should be opened up to meet the transparency demand of all stakeholders.

Some opacity around any artificial intelligence system is inevitable. The best-performing models in many domains - known as "black-box models" - can be complex to comprehend by human brains, such as deep neural networks. Thanks to emerging Explainable AI (XAI) techniques, it is possible to open up black-box AI to certain extents without scarifying model performance and accuracy. This is the foundation of understanding how decisions are made and assessing whether they are fair.

Explainable AI techniques facilitate algorithm transparency, one of the key enablers for ethical and trusted AI.

Abiding by data laws and using data ethically

Although the US does not currently have a nationwide data privacy law, with the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) both recently setting the stage, we are firmly moving towards establishing an oversight of data privacy issues that could impact the data usage for AI. Numerous other states have followed suit, proposing data privacy legislation similar to the CCPA in 2021, with more states expected to follow in the near future.

The data and attributes required for training AI and algorithms need to be simultaneously relevant, appropriate for the goal and allowed for use. This does not mean all data is off limits; personal information can be necessary in some use cases under restricted governance and in protected environments, such as healthcare research. The data used to train an AI model should not leak any personal information by itself, through the proxy or linked datasets.

Privacy-preserving AI can be achieved by limiting the use of sensitive information in the training dataset and emerging technologies designed to protect the privacy of input, output, and the AI model. For example, homomorphic encryption allows training and prediction to be performed directly on encrypted data; differential privacy measures reduce the risk of leaking individual details in model training and later stages.

Clear policies should be established about the development and deployment of AI, including the use of data, standards of privacy, the use of privacy-preserving technologies, and governance of leading practices.

Control the power of AI in your organization

Human oversight is necessary to maintain control over artificial intelligence and maintain trust at a stakeholder level. But several questions need to be asked for an effective framework to be put in place:

- Which decisions are we comfortable handing over to machines, and which decisions should remain in the human realm?
- Why and how were certain use cases chosen as candidates for AI?
- Will the results of AI algorithms impact live (e.g., hiring), or other objects (e.g., asset register)? If impacting live, the higher ethical standard should be in place, even if we need to sacrifice model accuracy with more transparency.
- Why did the team, or the feature engineering algorithm, choose features they chose, or exclude what they excluded?
- How do we measure and demonstrate success or explain failures?
- Why did the algorithm do what it did, and who was responsible for the outcome?

As systems become ever more powerful, decisions and blame cannot be focused squarely on the algorithm, and AI must not be used to unduly influence or manipulate thoughts and behavior. Organizations must have governance systems in place from the foundations of AI implementation to achieve the desired outcomes while remaining ethical.

Ethical AI solutions from KPMG

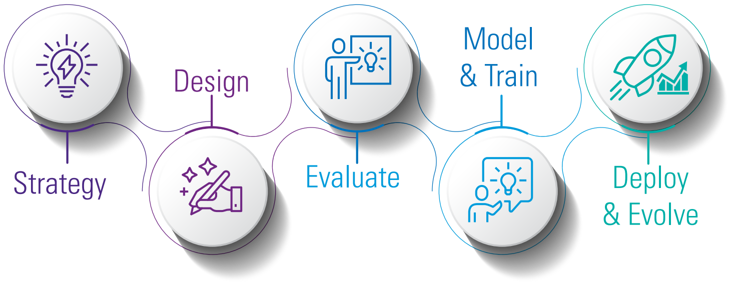
With a lack of laws and official regulations, a government or agency must define what is and is not ethical. KPMG can help your organization establish a framework which helps you evaluate and resolve the ethical issues around the subject of AI, enabling you to address the issues which matter most and ultimately maintain ethics.

Establishing an effective ethical AI framework from the ground up can help organizations gain confidence in the use of AI technology and its many benefits. We can help organizations dig deep into AI at both the enterprise and individual model level, ensuring key trust imperatives and ethical considerations are integrated at a foundational level and controlled throughout.

An effective governance strategy lays the foundation of ethical AI by putting in place mechanisms and tools that will continuously measure, assess, and maintain control over AI and evolving algorithms. This governance model should include data and process standards for key factors such as privacy, security, bias and transparency. As ethics is a key piece in the broader picture of Trusted AI, this ethical AI framework needs to be easily incorporated into the organization's broader Trusted AI governance model.

Since AI is a complex and fast-evolving technology, it requires a sophisticated and robust technology layer to complement and enable the governance model, addressing bias, transparency, privacy, human oversight, etc. The technology layer needs to be well-architected and carefully customized to fit the specific AI use cases and the enterprise-level digital transformation roadmap.

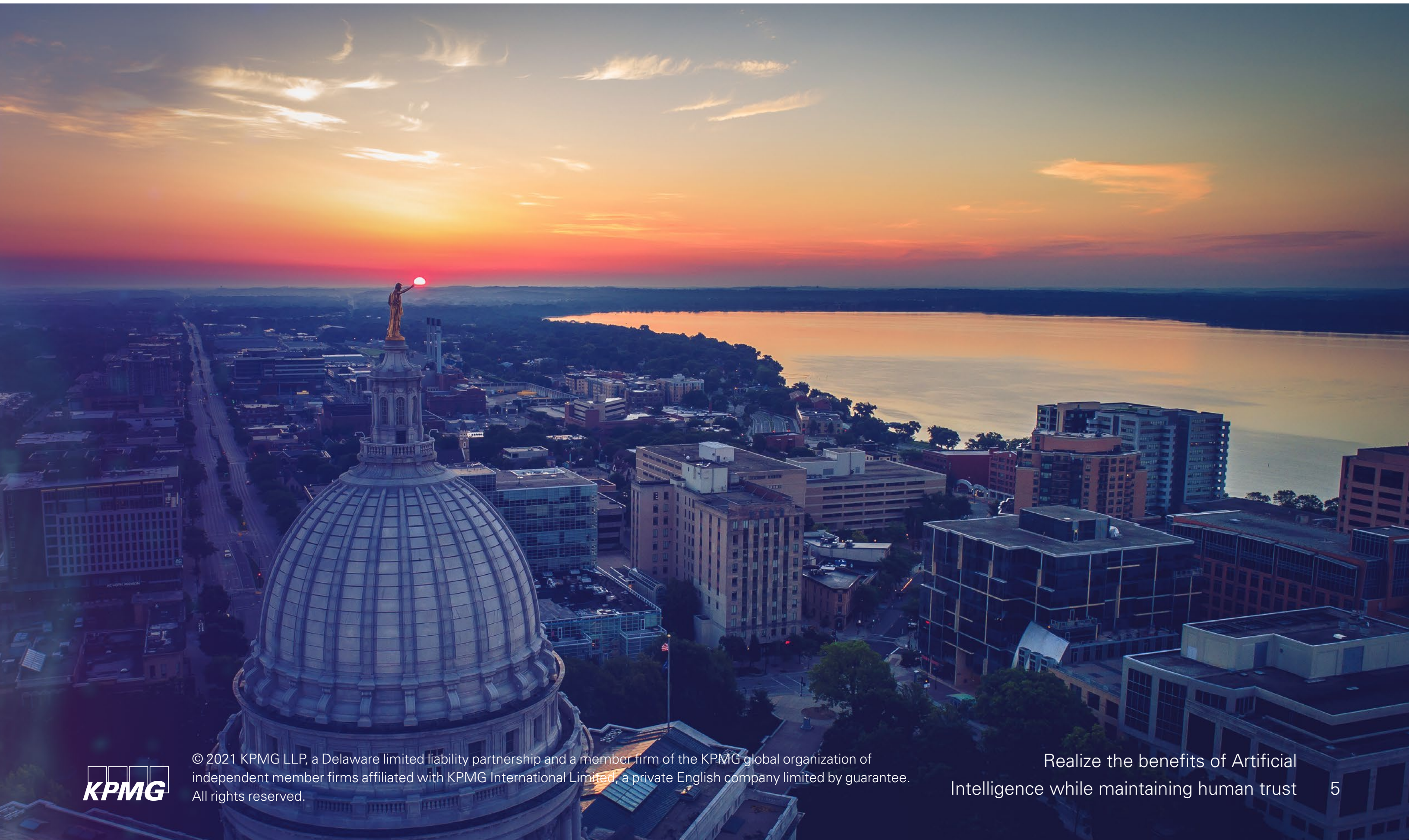
Finally, we address the need of maintaining AI ethics to clearly manage the end-to-end lifecycle of AI. Enterprise-wide policies and processes should be established to govern the entire lifecycle of developing and deploying AI, including the continuous monitoring, evaluation and evolution after deployment.



About KPMG

KPMG has worked with federal, state, and local governments for more than a century, so we know how agencies work. Our team understands the unique issues, pressures, and challenges you encounter in the journey to modernize. We draw on our government operations knowledge to offer methodologies tailored to help you overcome these challenges and work with you from beginning to end to deliver the results that matter.

The KPMG team starts with the business issue before we determine the solution because we understand the ultimate mission. When the way people work changes, our team brings the leading training practices to make sure your employees have the right knowledge and skills. We also help your people get value out of technology while also assisting with cloud, advanced analytics, intelligent automation, and cybersecurity. Our passion is to create value, inspire trust, and help government clients deliver better experiences to workers, citizens, and communities.



Contact us

Viral Chawda

Principal, Advisory
Digital Lighthouse
KPMG LLP
214-840-2000
vchawda@kpmg.com

Ning Wang

Specialist Director
KPMG LLP
214-840-2409
ningwang1@kpmg.com

read.kpmg.us/modgov

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.