



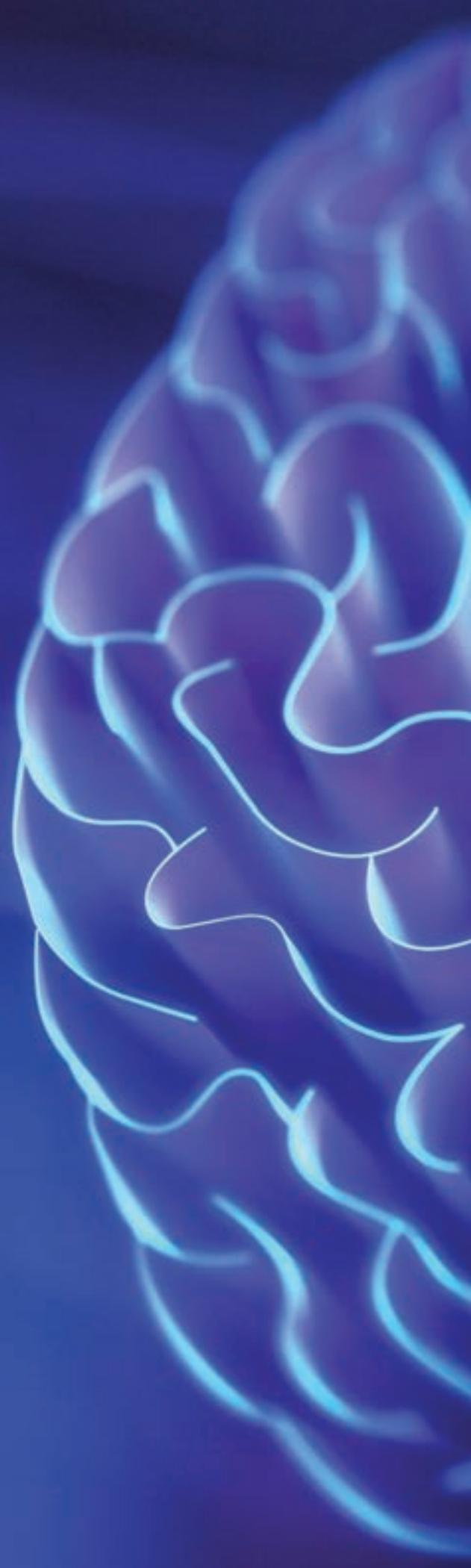
# Controlling AI

**The imperative for  
transparency and explainability**

September 2019

---

[kpmg.com](https://www.kpmg.com)



# Our authors



**Andrew Shefford**  
**Managing Director**  
Emerging Technology Risk Services  
KPMG in the U.K

---

Andrew Shefford leads KPMG's Artificial Intelligence In Control team and is part of the global core team. He was lead author for KPMG's publication 'Trust in AI' and was central to the development of KPMG's AI risk and control framework, both published on the firm's website. Andrew has extensive experience helping organisations adapt their IT risk and control approaches to emerging technologies and is the firm's Global network leader for IT Internal Audit. Like Martin, Andrew believes strongly that overcoming lack of trust in AI is central to its successful exploitation and true technology of the future.



**Martin Sokalski**  
**Principal, Advisory, Emerging  
Technology Risk Services**  
KPMG in the U.S

---

Martin Sokalski is a Global Leader for KPMG's Emerging Technology Risk practice. He helps organisations around the globe embrace the "art of the possible," enabled by emerging technologies like artificial intelligence, by facilitating ideation, innovation, and responsible adoption. Over the years, he has helped many organisations across multiple sectors assess, design, and implement new digital operating and governance models, to help them achieve desired business outcomes while embedding key governance, trust, and value imperatives. Martin regularly speaks at conferences and contributes to thought leadership on artificial intelligence, digital transformation, and emerging technologies. Martin believes that adoption of AI at scale is currently inhibited by lack of trust and transparency, explainability, and unintended bias and he aims to work with industry leaders to solve for that challenge.



**Professor Dr. Sander Klous**  
**Partner, Data & Analytics Lead**  
KPMG in the Netherlands

---

Sander Klous is a Data & Analytics Leader for KPMG in the Netherlands and a professor of Big Data ecosystems for business and society at the University of Amsterdam. He has a PhD in High Energy Physics (HEP) and has worked for over a decade on a number of projects for CERN, the world's largest physics institute, in Geneva. His best-selling book, *We are Big Data*, was runner-up for the management book of the year award in 2015. His new book, *Building Trust in a Smart Society*, is a top-selling management book in the Netherlands. Sander has significant experience in large-scale distributed computing, real-time systems, and data processing technologies. His current focus is on the broad use of reliable analyses, ethical algorithms, and trusted analytics in a way that is valued by clients and society at large.

# Contents

1

Introduction

5

Key developments

11

Governance  
and ethics of AI

15

Key to governing AI: a framework  
that helps enable transparency

17

AI in Control: a framework to  
govern algorithms



# Introduction

**World-changing technologies over human history all involve a common element: Control.**

**Steam and light—and a long list of inventions and technologies—emerged because we were able to guide natural forces into transformative power.**

**Aviation would not exist without the mastery we have attained over flight.**

**Artificial Intelligence (AI) has the potential to be just as world-changing.**



## But we don't know the full extent of what AI can do for the world.

And, like other transformative technologies, the power and promise of AI can only be fully unlocked by our understanding and control of its build and actions. This is why companies need to establish an overall management policy for AI, with a focus on responsibly unleashing the power of these technologies.

AI unveils a world hidden in complexity. The insights from algorithms that learn and continuously evolve are changing our businesses and our lives. Many scientists see a future where some of the deepest mysteries and intractable problems facing humanity can be solved. We are already seeing the benefits emerge—from algorithms that discover subatomic particles and help capture the first photograph of a black hole to the enterprise level, where sophisticated data and analytics, driven by AI, are making mission-critical decisions that affect the bottom line and the brand—and the health and safety of consumers.

### AI on the ground.

Picture a line of business owner (LOB) for consumer loans at a large financial institution. A situation involving bias and discrimination has surfaced, along with a headline or two in the news. During a board meeting, one member after another asks this leader and the chief digital officer (CDO) to explain the decisions and rationale behind the denial of loans to applicants of a certain age group or race. At play is an AI algorithm that produced the results or augmented a decision by loan officers in the field. The problem for these two leaders: No one can explain exactly why the algorithm did what it did.

Moments similar to this are playing out in areas across business and the public sector: recruiting, transportation, marketing, healthcare, college admissions, housing, and the management of smart cities. Any organisation that builds or adopts advanced, continuous-learning technologies is tapping into a power for insight and decision-making that far exceeds the capabilities of the human mind. This is a massive opportunity.

But algorithms can be destructive when they produce inaccurate or biased results, an inherent concern amplified by the black box facing any leader who wants to be confident about their use. That is why, in the midst of enormous excitement around AI, there is hesitancy in handing over decisions to machines without being confident in how decisions are made and whether they're fair and accurate. This is a trust gap.

## Gaining confidence.

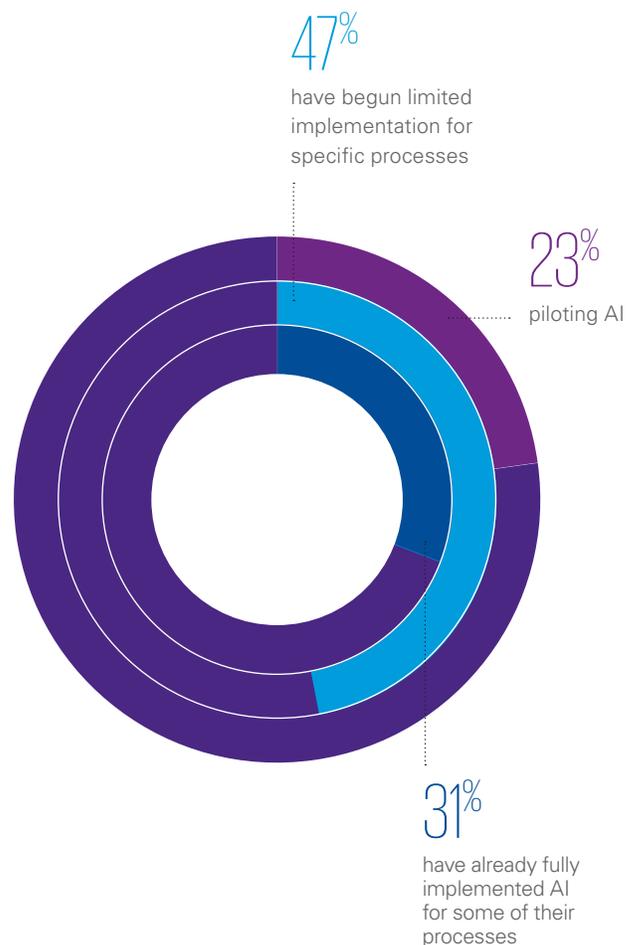
The enormous benefits of AI will fully emerge only when algorithms become explainable (and, hence, understandable) in simple language, to anyone. The trust gap exists because there is no transparency of AI; instead, there is an inherent fear of the unknown surrounding this technology. Gaining trust also involves understanding the lineage of the AI models and protecting them (and data that forms them) from different types of adversarial attacks and unauthorised use. Critical business decisions made by AI affect the brand—and consumer trust in the brand—and they can have an enormous impact on the well-being or safety of consumers and citizens. No one wants to say, “because the machine said so.” No one wants to get AI wrong.

## Closing the trust gap.

Fair and explainable AI is more than a big ask in the C-suite and the boardroom today—it’s a demand. KPMG’s 2019 CEO Outlook<sup>1</sup>, for example, found that 66% of leaders surveyed overlooked insights provided by computer-driven data analysis because they were contrary to their experience or intuition.

For most organisations, AI is still in the lab, so to speak, deployed on a functional level and not yet an integral part of the decision-making in the business—although that is rapidly changing.

## According to the KPMG 2019 U.S. CEO Outlook, organisations are at different levels of their AI deployment journeys.



<sup>1</sup>KPMG 2019 CEO Outlook: Agile or Irrelevant: Redefining Resilience, June 2019

## What's the solution?

For AI to move ahead toward the common good, for leaders to assume responsibility and accountability over the results, it's essential to establish a framework (powered by methods and tools) to facilitate responsible adoption and scale of AI.

**This report is for leaders involved in the world of Artificial Intelligence and Machine Learning algorithms.**

**The business and compliance imperative to understand and be confident in AI technologies has reached critical mass.**

**This paper explains the urgency and describes methods and tools that can help leaders govern their AI programs.**



“The true art of the possible for Artificial Intelligence will become unlocked as soon as there is more trust and transparency. This can be achieved by incorporating foundational AI program imperatives like integrity, explainability, fairness and resilience.”

---

**Martin Sokalski**

Principal, Advisory, Emerging  
Technology Risk Services  
KPMG in the U.S.

# Key developments

**Based on interviews with executives driving AI strategy at large companies, we heard a consistent message. Many companies are just beginning to invest in AI control frameworks, compared to other AI deployment priorities<sup>2</sup>.**

---

<sup>2</sup>KPMG 2019 Enterprise AI Adoption Study



### **Gaining trust around AI is a top goal of leaders.**

45% of surveyed executives say that trusting AI systems was either challenging or very challenging<sup>3</sup>.



### **New policy initiatives and regulations around data and AI signal the end of self-regulation and the rise of a new oversight model<sup>4</sup>.**



### **Most leaders aren't clear on what an AI governance approach should be.**

Some 70% say they don't know how to govern algorithms<sup>5</sup>.



### **Companies are struggling to decide who is accountable for AI programs and results.**

During our interviews, we heard that most companies are still trying to determine who has authority over AI deployment. Some companies have established a central authority in an AI council or Centre of Excellence; others have assigned responsibility to different leaders, like the Chief Technology Officer or Chief Information Officer.



### **A framework that includes technology-enabled methods can help address the inherent risks and ethical issues in AI.**

The objective is to help business users gain control over their AI programs by enabling four trust anchors: integrity, explainability, fairness, and resilience.

<sup>3</sup>Forrester Research, Q2 2018 Global AI Online Survey

<sup>4</sup>AI, Internet Policy Proposals Signal Shift Away From Self-Regulation. Wall Street Journal, WSJ Pro: Artificial Intelligence, April 9, 2019

<sup>5</sup>Source: KPMG, Why AI Must Be Included in Audits, 2018

# The need to know: trust anchors

**The cost of getting AI wrong extends beyond the financials—lost revenue, fines from compliance failures—to reputational, brand, and ethical concerns.**

---

We just pictured a CDO and LOB leader trying to explain the outcome of a single model before the board. The layers of accountability extend from the C-suite and the line of business owner for the entire credit card division of a bank (someone who owns everything related to this business, including the AI models) all the way to the customer level, with a loan officer who may face accountability and who, in many ways, represents the brand.





## Key business decisions at scale have a determining effect on success, as an example:

Should the division approve a credit card for a customer?

Among the decisions for each customer: the annual percentage rate, the spending limit, and a long list of other factors. Machine learning models are typically making these decisions for millions of customers. In a very real sense, given the scale, the business is in the hands of a handful of smart data scientists—and the machines they build and train—using ground truth created from historical loan data.

## Autonomous algorithms: then vs. now.

Most algorithms today are relatively simple and deterministic: They produce the same output from a predetermined set of states and a fixed number of rules. The approaches for evaluating them for validity and integrity are largely established and adopted. In fact, in our estimation, over 80% of the leading practices needed to maintain their accuracy and effectiveness are known.

Think of expert systems in manufacturing. Think of actuarial science that uses deterministic rules or decision tables in insurance. Think of robotic process automation in financial services.

It isn't that hard to determine whether the conclusions they reach are acceptable—and sound and scalable supervision is relatively easy.

These rules can get very complex, especially when the number of attributes (also known as features, or variables) in the data or the number of records increases.

Machine learning and deep learning—and other types of AI—are creatures of a different kind. They are trained to learn from data (commonly referred to as ground truth) instead of being explicitly programmed, which means they can “understand-learn-uncover” the nuances and the patterns in the data, they can handle a very large set of attributes, and are often significantly more complex in how they do what they do.

Think of training a prediction model from a set of a million past loan applications, which in turn uses 100 attributes. Think of detecting a tumour from a million MRI images. Think of classifying emails. Once trained and evaluated, these models can be provided with new or unseen data from which they can make predictions. They are probabilistic in nature and respond with a degree of confidence.

While all of these aspects are good, it can be unclear what the models are doing: what they learn, particularly when employing opaque deep learning techniques such as neural nets, how they will behave, or whether they will develop unfair bias over time as they continue to evolve. That's why understanding which attributes in the training data influence the model's predictions has become very important.

## Algorithmic Risk: Trust in the Machine

Let's take a closer look at a potential problem for the CDO and line of business owner for the loan division of a big financial firm.

If an error hides within an algorithm (or the data feeding or training the algorithm), it can influence the integrity and fairness of the decision made by the machine. This could include adversarial data or data masking as ground truth. The business leaders are on the hook for preserving the brand reputation for the firm, even as the AI models increasingly make decisions that might not be understood or in line with corporate policies, corporate values, guidelines, and the public's expectations. Multiply these issues by the number of algorithms the loan division is utilising. This is when trust weakens or actually evaporates.

A number of techniques, including those based on renormalisation group theory, have been proposed<sup>6</sup>. As models across AI tasks—including computer vision, speech recognition, and natural language processing—become more sophisticated and autonomous, they take on a higher level of risk and responsibility. When left untrained for long periods, things can go awry: runtime bias creep, concept drift, and issues such as adversarial attacks can compromise what these models learn. Imagine compromised MRI scans or traffic lights being manipulated in a smart city.

Continuous-learning algorithms also introduce a new set of cybersecurity considerations. Early adopters are still grappling with the magnitude of risks presented by these issues on the business.

Among the risks are adversarial attacks that hit the very foundation of these algorithms by poisoning the models or tampering with training data sets, potentially compromising privacy, the user experience, intellectual property, and any number of other key business aspects. Consider the impact on lives or an environment of an adversarial attack in medical devices or industrial control systems. Tampering with data could disrupt consumer experiences by providing inappropriate suggestions in retail or financial services. Such attacks might ultimately erode the competitive advantage that the algorithms were intended to create.

With complex, continuous-learning algorithms, humans need to know more than just the data or attributes and their respective weights to fully realise the implications of the AI getting it wrong or going rogue; they need to understand aspects such as the context and intended purpose under which the model was developed, who trained them, provenance of the data and any changes made to it, and how the models were (and are) served and protected. And they need to understand what questions to ask and what key indicators to look for around an algorithm's integrity, explainability, fairness, and resilience.

**The No. 1 challenge for AI adopters is quality data. The CTO of a government agency specifically stated in our Global AI survey that if they can't trust data, they can't use AI<sup>7</sup>.**

<sup>6</sup>An Exact Mapping Between the Variational Renormalisation Group and Deep Learning, Pankaj Mehta, David J. Schwab. 2014

<sup>7</sup>Forrester Research, Q2 2018 Global AI Online Survey

## The anchors of trust.

When you break down all the actions and capabilities needed to secure trust in your algorithms and models—and hence your brand—KPMG believes that four dimensions emerge.



### Algorithm integrity.

Think of a home inspection that checks the ‘bones’ of a house as a metaphor for determining the structural flaws and integrity of an algorithm. What leaders need to know is this: the provenance and lineage of training data, controls over model training, build, model evaluation metrics and maintenance from start to finish, and the verification that no changes compromise the original goal or intent of the algorithm. Also key would be continuous monitoring of the model performance metrics, including concept drift detection.



### Explainability.

Understanding the reasons a model made a prediction—and being able to interpret the reasons—is essential in trusting the system, especially if one has to take an action based on those probabilistic results. This is a subjective capability in AI. Being able to explain why and how a model produced an output (insight, decision) depends on the definition of success established and the overall governance of the algorithm, from the assemblage of ground truth that is clean, sufficient, and appropriate to the continuous assessment of results. Several approaches exist—including LIME, an explanation technique that focuses on local, or isolated, aspects of decision making,<sup>8</sup> and Defense Advanced Research Projects Agency (DARPA) Explainable AI (XAI) program, which aims to create a suite of machine learning techniques leading to more explainable models, with an explanation interface.



### Fairness: ethics and accountability.

AI and algorithms won’t be trusted if they’re not fair. For them to be fair, they need to be designed and built as free from bias as possible—and they need to maintain fairness as they evolve. Attributes used to train algorithms need to be relevant, appropriate for the goal, and must be allowed for use. In some instances, however, personal information is relevant to the model, as in healthcare when gender or race can be a critical part of studies or treatment. Careful oversight and governance is needed to make sure proxy data doesn’t train a model. A postal code, for example, can be a proxy for ethnicity or income and inadvertently produce biased results and downstream risks—just one being regulatory violations. Techniques must be applied to understand bias that inherently exist in the data, and mitigate them using approaches such as rebalancing, reweighting, or adversarial removal of bias..

Tools for continuous monitoring as well as governance are essential to help ensure models that are continuously trained with usage and feedback data don’t cause bias to creep in during runtime.



### Resilience.

Here is where we’re talking about the robustness and resilience of the models or algorithms that are deployed or served. The served models are typically exposed as APIs or embedded within applications, and they need to be portable and operate across diverse and complex ecosystems.

Resilient AI should cover all the aspects of secure adoption and holistically address risks through securely designed architecture and the detection of anomalies using AI concepts like generative adversarial networks that pit algorithms against each other to produce better and more nuanced outcomes. The goal is to help ensure all the components are adequately protected and monitored. Why? External circumstances can lead to errors when algorithms are unable to correct or compensate for data that is inaccurate or anomalous. Protecting the usage and feedback data that could be used to continuously train the models is also critical. Basic actions include continuously monitoring models endpoints and controlling access to the models.



### A central question needs to be resolved: Who among the humans is accountable for the results of AI?

Accountability is a crucial governance issue that must be established across all AI initiatives, down to each individual model. We found significant variation among KPMG’s 2019 Enterprise AI Adoption Study in assigning authority and accountability. Some organisations have created a centralised authority such as an AI council; others have assigned it to functions such as the office of the chief technology or chief information officer. But few organisations have solid accountability practices in place, a leadership gap that can weaken trust internally and among external stakeholders. A big reason for this missing link: Most organisations lack tools and expertise to gain a full understanding and introduce transparency into their algorithms.

<sup>8</sup>Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, “Why Should I Trust You? Explaining the Predictions of Any Classifier.” 2016

# Governance and ethics of AI

**Governance and ethics become the 'how' of responsible adoption of AI by addressing the risk that complex algorithms could take a wrong turn.**

---

Look at the rules and regulations that govern the aviation industry, and the internal best practices that dominate procedures at each individual airline, from the C-suite to the cockpit. Look to the trust placed in the experience by everyone involved—the crew, the passengers, and businesses that transport valuable assets by air. This is what industry must aim for with AI.





A tipping point has arrived in terms of the need for effective governance and responsible adoption and scale of AI. In many cases, organisations are developing internal policies and governance functions to oversee any matters relating to AI in an effort to engender trust and transparency across the enterprise and external stakeholder groups, including consumers.

In the U.K. and in the E.U., with its evolving General Data Protection Regulation, the tide is now firmly moving toward the establishment of oversight. And the timing is a good thing, as the seeds of AI are firmly in the ground and growing. The scale is not there yet, but these technologies are set to expand within the enterprise and across industry sectors and assume greater autonomy and responsibilities. Now is the time to set a framework for governance and ethics around the anchors of trust. Controlling AI will help enable a responsible expansion of power.

## Governance and ethics.

Assessing and securing the trust anchors of AI can come from a new set of leading practices and methods aimed at maintaining control over AI and machine learning algorithms. An effective governance strategy lays a foundation of trust and transparency by putting in place the mechanisms and tools that will continuously measure AI. Leaders will be able to make informed decisions, and their organisations will build a culture of accountability that is stronger and consciously representative of an organisation's ethical compass.

## Governance.

A long list of questions emerges when one digs deep into the workings of AI—and many of them are human issues. Why and how were certain use cases chosen as candidates for AI? Why did the team choose the features it chose (and exclude what it excluded)? How do we measure and demonstrate success (or explain failures)? Why did the algorithm do what it did and who is responsible for the outcome? “Because the algorithms said so,” will not work for leaders and the general public as these systems become ever more powerful and pervasive.

**The need:** Seeing the big picture and setting the right tone at the beginning. If you don't have a governance or an operating model construct for AI, it will be difficult to achieve the desired business outcomes or have confidence in your AI's integrity, explainability, fairness, or resiliency. Governing AI is also the right thing to do in terms of trust and visibility. That means looking at enterprise frameworks and governance through a new lens around people, process, and technology across the entire lifecycle: from a model's early stages through strategy, delivery, monitoring, training and capabilities, and continuing measurement.

## Ethics.

This is an immense topic in AI, both in terms of the issues and dilemmas facing business and society and in the steps and guardrails needed to control AI. Ethics and trust are entwined. And both are the fuel needed for AI to go forward in ways that benefit society in general. Resolutions and regulations are being implemented. The General Data Protection Regulation is a prominent example; and others are setting the stage, an example being the ethics guidelines for trustworthy artificial intelligence recently issued by the European Commission<sup>9</sup>.

Another aspect of ethics is personal autonomy. What decisions are we comfortable handing over to machines—and what decisions should remain in the human realm? This is a vibrant part of the discussion within the scientific community and in governments, notably Europe's Committee of Ministers, which recently declared that AI and machine learning technologies "must not be used to unduly influence or manipulate individuals' thought and behavior."<sup>10</sup>

**The need:** Establishing ethical guardrails from the early stages in an enterprise AI program, which requires visibility into—and the monitoring of—the AI lifecycle, from strategy to execution to continuous evolution.

<sup>9</sup>Ethics Guidelines for Trustworthy AI. European Commission, High-Level Expert Group on AI, April 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

<sup>10</sup>Council of Europe, Committee of Ministers, Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes. February 2019 [https://search.coe.int/cm/pages/result\\_details.aspx?objectid=090000168092dd4b](https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b)



# AI in Control

## The City of Amsterdam

KPMG is working with the City of Amsterdam to assess a digitised municipal service that allows residents to file requests online for matters such as trash in the street. The machine learning algorithm identifies the issue type, the priority, and the specific city service that should respond.

Amsterdam officials use KPMG's AI in Control framework to get an effective and continuous evaluation of evolving AI applications to keep them from inadvertently using patterns of learning that could lead to wrong or biased decisions.

The success of cities will increasingly depend on how smart and ethical they are with data.

### **The targeted outcome**

Enhance the public's confidence in a safe and well-maintained city; assist the city in its mission to protect the digital rights of residents.

"The City of Amsterdam aims to protect the digital rights of our citizens, and we have a responsibility to be inclusive and transparent about the machine learning algorithms we put in place to support our municipal services and programs, so we sought to develop an approach with a partner, KPMG, to help us develop the screening method to verify and approve these algorithms."

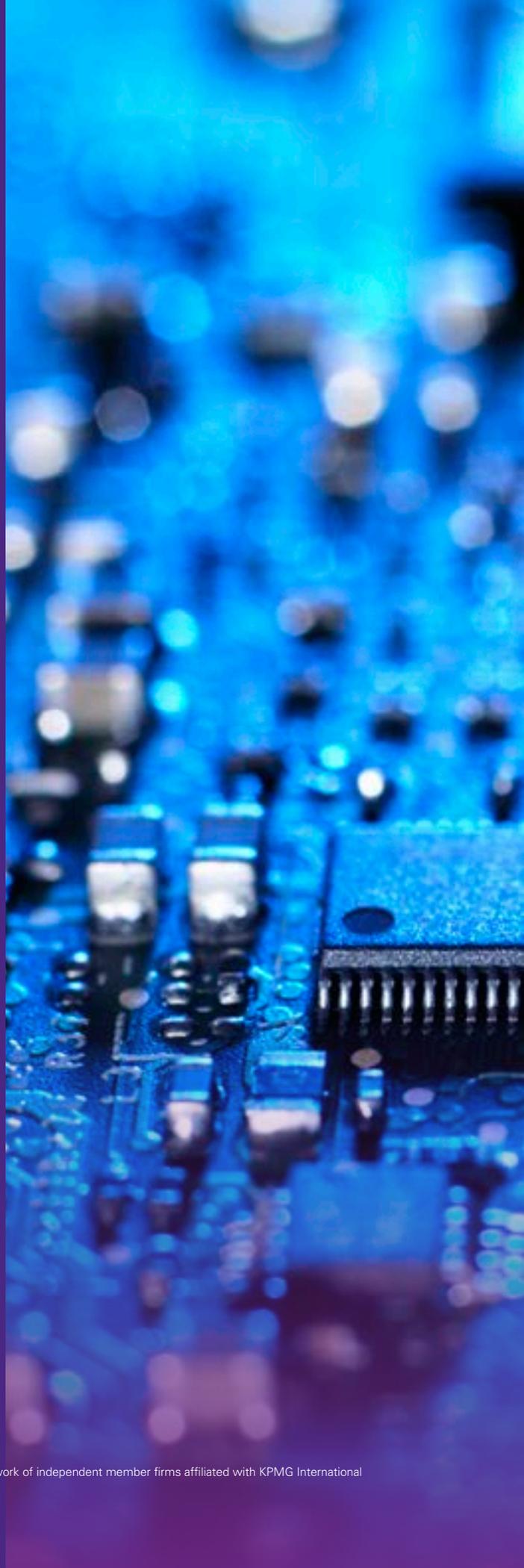
---

### **Ger Baron**

Chief Technology Officer,  
City of Amsterdam

# Key to governing AI: a framework that helps enable transparency

**Putting in place the standards to help business transparently and effectively govern algorithms—and gain trust in the quality of their business decisions.**



## Most leaders don't know how to close the trust gap because they don't know how to govern AI and see the big picture of its operation.

What leaders need is a ground-level view that reveals both the key performance metrics as well as key risk indicators. Bias in training data is one big concern. The overall model risk is another. So is compliance and security—and a long list of other items. And it leads back to that uncomfortable boardroom scene in the introduction of this report: How did a program produce the wrong result? In some cases, it can be due to underlying errors in the code or the data. Only a rigorous approach can prevent or detect them, removing doubt and closing the trust gap.

## Controlling AI

An effective framework can help organisations gain confidence in their AI technology. Such an approach should dig deep into AI at the enterprise and individual model level, to help ensure that key trust imperatives are integrated and controlled throughout. It should continuously assess and maintain control over sophisticated, evolving algorithms by putting in place methods, controls, and tooling that secure the trust anchors along the lifecycle, from strategy through evolution. It should also provide clear guidance for the organisation—stakeholders across various management and oversight functions, to clearly and consciously manage end-to-end lifecycle of AI. Examples of how this can be accomplished and some key considerations are:



### Strategy

Governance of AI begins at the beginning. Setting the right strategy for enterprise AI or for a specific model begins with a clear vision and aspiration and intended outcomes. Here we touch upon the concepts of ethics and accountability.



### Design

Helping ensure that the intended purpose of the algorithm is clearly defined and that models are designed to achieve that intended purpose through feature engineering, data bias, and ground truth. Design needs to align with principles (values and ethics), security and quality standards and guidelines, and compliance requirements.



### Model & Train

Once the design criteria are met, model building and training are initiated. In this phase, to maintain model integrity, fairness, explainability, and resilience one needs to consider bias detection, hyper parameters, feature provenance, among other variables. Model features and data need to be in compliance with organisational principles, policies, business requirements, and regulations.



### Evaluate

This is a key step in the AI lifecycle and it has to do with the ability to verify that the AI models and the outcomes they produce meet the requirements around integrity, fairness, explainability, and resilience. It's about knowing what questions to ask and what key performance and risk indicators to look for and having the capability to execute. The effectiveness and integrity of AI evaluation and monitoring capability will directly drive the confidence an organisation (or an external regulator) has in enterprise AI.



### Deploy & Evolve

AI and ML models are not static and will continue to evolve even after they are in production through interaction with new data sets or other models. Therefore, key factors to consider in this phase include runtime monitoring and reporting of controls, compliance and key performance and risk indicators and metrics for model accuracy, integrity, fairness, explainability, and resilience. Ability for the enterprise to react to those indicators (including dynamic model calibration) is also a needed capability.

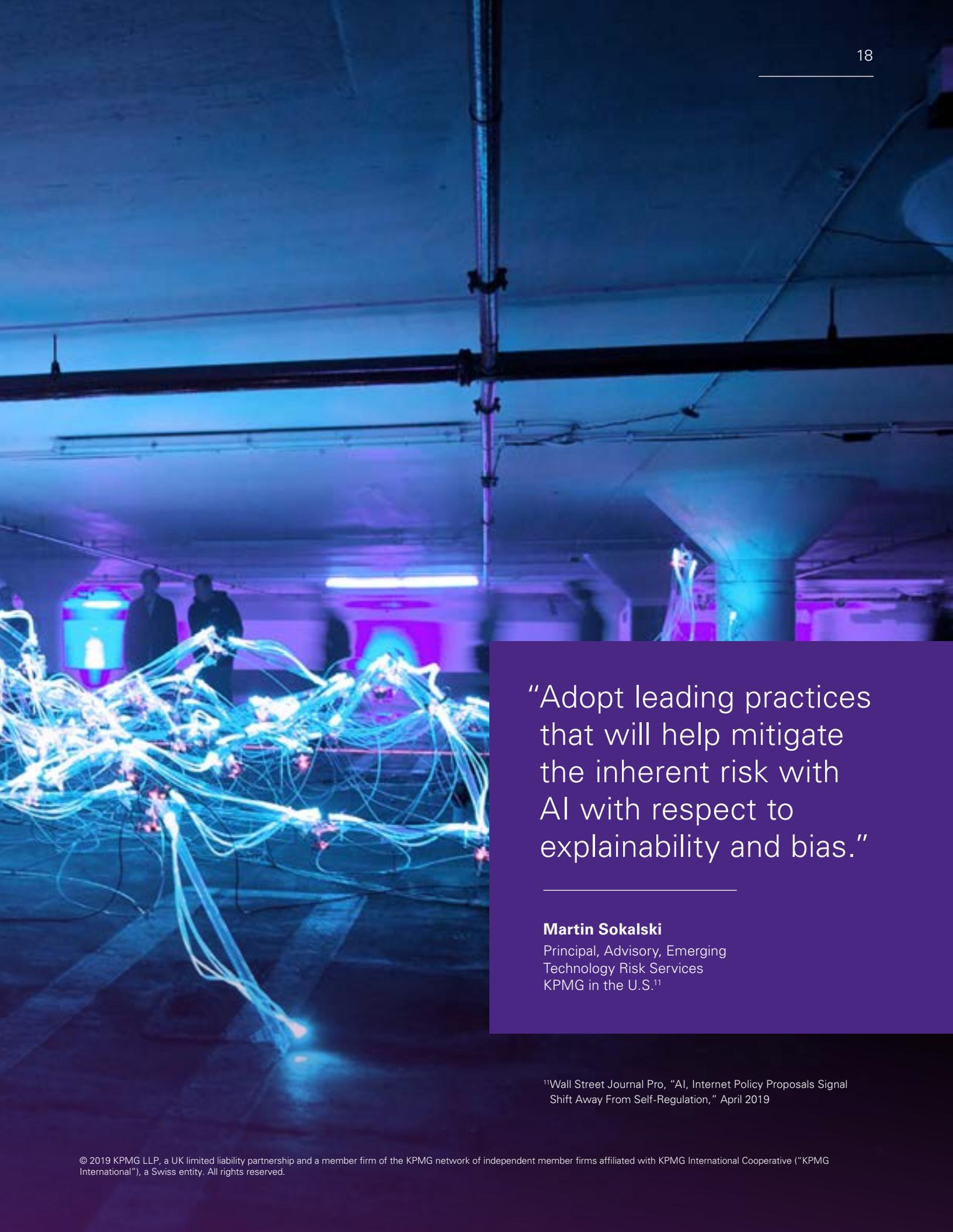
## Transparency from a solid framework of methods and tools is the fuel for trusted AI—and it creates an environment that fosters innovation and flexibility

---

Organisations that build and deploy AI technologies are tapping into a power for insight and decision-making that far exceeds human capability. It's a massive opportunity for business and society at large. But algorithms can be destructive when they produce inaccurate or biased results. That's why leaders are hesitating to hand decisions over to machines without knowing why they were made or whether they're fair and accurate.

The power and potential of AI will fully emerge only when the results of algorithms become understandable in clear, straightforward language. Companies that don't prioritise AI governance and the control of algorithms will likely jeopardise their overall AI strategy, putting their initiatives and potentially their brand at risk.





“Adopt leading practices that will help mitigate the inherent risk with AI with respect to explainability and bias.”

**Martin Sokalski**

Principal, Advisory, Emerging  
Technology Risk Services  
KPMG in the U.S.<sup>11</sup>

<sup>11</sup>Wall Street Journal Pro, “AI, Internet Policy Proposals Signal Shift Away From Self-Regulation,” April 2019

## AI governance

KPMG developed the AI in Control framework to help organisations drive greater confidence and transparency through tested AI governance constructs, as well as methods and tooling along the AI lifecycle, from strategy through evolution. By design, this framework addresses the inherent risks outlined in the sections above and it includes some of the key recommendations and leading practices for establishing AI governance, performing AI assessments, and building continuous AI monitoring and visualisations.



**Develop AI design criteria and establish controls** in an environment that fosters innovation and flexibility.



**Design and implement an end-to-end AI governance and an operating model** across the entire lifecycle: strategy, building, training, evaluating, deploying, operating, and monitoring AI.



**Assess current governance framework and perform gap analysis** to identify opportunities and areas that need to be updated.



**Design a governance framework that delivers AI solutions and innovation** through guidelines, templates, tooling, and accelerators to quickly, yet responsibly, deliver AI solutions.



**Integrate a risk management framework to identify and prioritise business-critical algorithms** and incorporate an agile risk mitigation strategy to address cybersecurity, integrity, fairness, and resiliency considerations during design and operation.



**Design and set up criteria to maintain continuous control over algorithms** without stifling innovation and flexibility.

“First, we need to make sure the data is clean, sufficient, and appropriate. Next, we need to make sure the algorithm gave consistent results and did not depend on small changes in starting assumptions. Finally, we need to make sure that the overall goal was achieved without having overly negative consequences for any particular stakeholder.”

---

**Cathy O’Neil**

Consultant and author of *Weapons of Math Destruction*, from the introduction to *Building Trust in a Smart Society* (Sander Klous, KPMG Netherlands)



## AI assessment

**Conduct a diagnostic review of an enterprise AI program and governance**

to evaluate the current state and applicability of existing governance elements to AI as well as current operating models and readiness for AI at scale. This will include a capability and maturity assessment, as well as a roadmap and recommendations for helping to achieve the target state.



**Conduct assessment of individual AI and ML algorithms:**

testing of controls, evaluation of design, implementation and operation of the algorithm based on four trust anchors—integrity, explainability, fairness, and resilience.

## Continuous monitoring and dashboards



**Create full visibility** into metrics related to the trust imperatives, including key performance and risk indicators such as **Board, Executive, and Program** level reporting focused on key relevant AI KPIs and KRIs

---



**Enable continuous monitoring of key controls and metrics** – what is working (or not) across your AI/ML models

---



**Provide view of the upward/downward trend** over a time period, based on controls and testing

---



**Have ability to respond and correct issues** as they arise. For example, bias is introduced in the learning model, or prohibited features are being used in decision-making.

---



**Conduct an assessment** of your AI model(s) or a health check of your broader enterprise AI program



## Key differentiators of KPMG AI in Control



**Platform agnostic**



**Continuous monitoring**, including for bias and accuracy



**Continuous protection and security**, including training data, to prevent adversarial and other cyber attacks



**Ability to map** the data science terms and concepts to key business risk indicators



**Full visibility** into what the AI models are doing



**End-to-end framework** that governs the build, deployment and evolution of models



Will help drive **greater adoption and scale** across the enterprise

## How does AI in Control work?

The core set of components includes:



### Comprehensive AI framework

The AI Framework helps organisations build trust in their technology performance by transparently and effectively governing algorithms



### AI Knowledge mapping expertise

Looking at the overall governance and management framework for an AI and map it back to the corporate policies and guidelines from a risk perspective



### Prototype-architecture capability

An environment that fosters greater AI control, which is digitised and flexible, to gauge algorithmic risk



### Visibility and risk management dashboard

Allows the user visibility into the various metrics related to the trust imperatives

## Uncovering the full potential of your AI.

Today's organisations rely heavily on algorithm-based applications to make critical business decisions. While this unlocks opportunities, it also raises questions about trustworthiness. As we enter an age of governance by algorithms, organisations must think about the governance of algorithms to build trust in outcomes and achieve the full potential of artificial intelligence.

That's where KPMG AI in Control comes into play. KPMG member firms believe that the governance of AI is just as important as the governance of people. KPMG professionals operate in a technology-agnostic environment, and their recommendations are based on what is best for your needs. Our member firms work to provide a holistic, broad-ranging approach to help you along your AI journey and to achieve your business objectives, now and in the future.

***[read.kpmg.us/  
Alincontrol](https://read.kpmg.us/Alincontrol)***



# Contact us

## **Andrew Shefford**

Managing Director

Emerging Technology Risk Services  
KPMG in the UK

---

T: +44 7775 704613

E: [andrew.shefford@kpmg.co.uk](mailto:andrew.shefford@kpmg.co.uk)

## **Martin Sokalski**

Principal, Advisory, Emerging  
Technology Risk Services

KPMG in the U.S.

---

T: +1 312 665 4937

E: [msokalski@kpmg.com](mailto:msokalski@kpmg.com)

## **Professor Dr. Sander Klous**

Partner, Data & Analytics Lead

KPMG in the Netherlands

---

T: +31206 567186

E: [klous.sander@kpmg.nl](mailto:klous.sander@kpmg.nl)

**[kpmg.com/uk](https://kpmg.com/uk)**

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 endeavour 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 on such information without appropriate professional advice after a thorough examination of the particular situation.

© 2019 KPMG LLP, a UK limited liability partnership and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative ("KPMG International"), a Swiss entity. All rights reserved.

The KPMG name and logo are registered trademarks or trademarks of KPMG International.

CREATE. | CRT114031A | September 2019