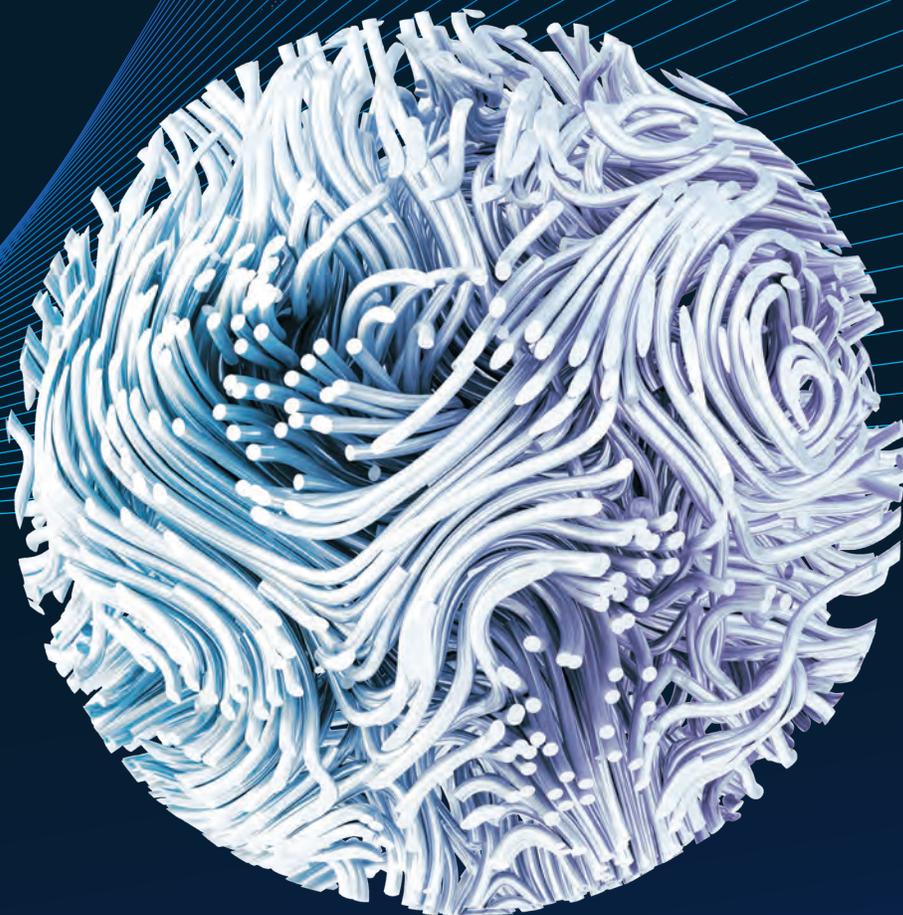


McKinsey  
& Company



# McKinsey on Government

Perspectives: Adopting AI, automation, and advanced analytics in governments

*McKinsey on Government* is a publication of McKinsey's Public and Social Sector Practice and McKinsey Center for Government (MCG). Our objective is to share insights, new approaches, and experiences from our proprietary research and relationships across all aspects of government around the world at the local, state, agency, national, and supranational levels.

This and archived issues of *McKinsey on Government* are available online at [McKinsey.com](https://www.mckinsey.com).

Editorial Contact:  
[McKinsey\\_on\\_Government@McKinsey.com](mailto:McKinsey_on_Government@McKinsey.com)

Cover Image:  
© Westend61/Getty Images

**McKinsey Center for Government Leaders:**

Solveigh Hieronimus, Navjot Singh

**Editorial Board:**

Tera Allas, Alberto Chaia, Michael Conway, Eoin Daly, Jonathan Dimson, Andre Dua, Nora Gardner, Andrew Goodman, Solveigh Hieronimus, Naufal Khan, Adam Kendall, Acha Leke, Diaan-Yi Lin, Tarek Mansour, Gundbert Scherf, Joerg Schubert, Navjot Singh, Sebastian Stern, Ali Ustun, Jonathan Woetzel

**Contributing Editors:**

Frances Catanio, Bill Javetski, Dennis Swinford

**External Relations:**

Sharmeen Alam, Alison Burke

**Art Direction and Design:**

Nicole Esquerre,  
Leff Communications

**Data Visualization:**

Richard Johnson, Jonathon Rivait

**Managing Editors:**

Heather Byer, Venetia Simcock

**Editorial Production:**

Elizabeth Brown, Roger Draper, Gwyn Herbein, Pamela Norton, Katya Petriwsky, Charmaine Rice, John C. Sanchez, Dana Sand, Katie Turner, Sneha Vats, Pooja Yadav, Belinda Yu

**McKinsey Practice Publications**

**Editor in Chief:** Lucia Rahilly

**Executive Editors:**

Michael T. Borruso,  
Allan Gold, Bill Javetski,  
Mark Staples

Copyright © 2019 McKinsey & Company. All rights reserved.

This publication is not intended to be used as the basis for trading in the shares of any company or for undertaking any other complex or significant financial transaction without consulting appropriate professional advisers.

No part of this publication may be copied or redistributed in any form without the prior written consent of McKinsey & Company.

# Table of contents



3

## **How governments can harness the power of automation at scale**

Process automation and technologies based on artificial intelligence can bring benefits across numerous functions of government.

---

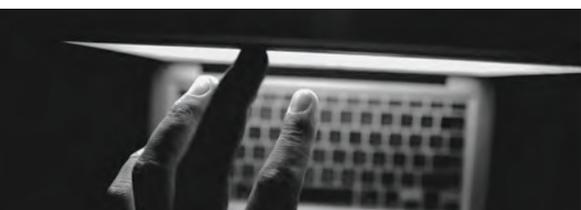


8

## **Cracking down on government fraud with data analytics**

New data tools are giving government agencies the upper hand in taming fraud, waste, and abuse. Lessons from scaled approaches show how to capture the hundreds of billions of dollars at stake.

---



15

## **Defense of the cyberrealm: How organizations can thwart cyberattacks**

Governments and companies have much work to do to protect people, institutions, and even entire cities and countries from potentially devastating, large-scale cyberattacks.

---



27

## **The ethics of artificial intelligence**

Executives and companies can enjoy the benefits of artificial intelligence while also being aware of potential drawbacks and taking careful steps to mitigate their effects.

---

# Introduction

We are delighted to present this latest compendium of articles from *McKinsey on Government*, the journal offering McKinsey's global perspective and strategic thinking on innovation in government productivity and performance. This compendium focuses on how governments can tackle long-standing problems and achieve breakthrough innovations leveraging next-generation solutions—such as artificial intelligence (AI) and automation—in pragmatic and risk-conscious ways. The five articles in this compendium present deep insights and practical approaches grounded in practical, hands-on experience from the front line.

The digital revolution across governments is just beginning, and it can help governments innovate faster to improve citizen experience, engage employees better, and achieve a step change in operational efficiency and effectiveness. For example, a recent study reveals that AI has the potential to deliver additional global economic activity of around \$13 trillion by 2030, or about 16 percent higher cumulative GDP compared with today.<sup>1</sup> In addition, we estimate automation alone could raise productivity growth on a global basis by 0.8 to 1.4 percent annually.<sup>2</sup> And as you will see in the articles, some of these technologies can also be used to plug leakages in government operations. For example, the US government each year makes more than \$140 billion<sup>3</sup> in improper payments—about 40 percent of which result in a monetary loss to the government. Technologies such as AI can be used to reduce these monetary losses arising from fraud, waste, and abuse.

The compendium begins with a piece about how different government functions can harness the power of automation at scale. We then look to share how governments can crack down on fraud with data analytics and discuss the benefits, risks, and ethical considerations associated with using AI in the public sector. We conclude by sharing best practices on how governments can protect people, institutions, and even entire cities and countries from large-scale cyberattacks—a critical topic for governments to consider.

We hope you enjoy these articles and find in them ideas worthy of application. For further information, and to download our publications, please visit [mckinsey.com/industries/public-sector/our-insights](https://mckinsey.com/industries/public-sector/our-insights).

We welcome comments: please email us at [McKinsey\\_on\\_Government@McKinsey.com](mailto:McKinsey_on_Government@McKinsey.com)

Sincerely,

**Solveigh Hieronimus**

Global Co-lead  
McKinsey Center for Government  
Partner, Munich

**Naufal Khan**

Digital and Analytics Lead  
McKinsey Center for Government  
Senior Partner, Chicago

**Navjot Singh**

Global Co-lead  
McKinsey Center for Government  
Senior Partner, Boston

---

<sup>1</sup> Bughin, Michael Chui, Raoul Joshi, James Manyika, and Jeongmin Seong, "Notes from the AI frontier: Modeling the impact of AI on the world economy," September 2018, [McKinsey.com](https://mckinsey.com).

<sup>2</sup> *What's now and next in analytics, AI, and automation*, McKinsey Global Institute, May 2017, [McKinsey.com](https://mckinsey.com).

<sup>3</sup> "Government efficiency and effectiveness: Opportunities to address pervasive management risks and challenges while reducing federal costs," US Government Accounting Office, May 17, 2017, [gao.gov](https://gao.gov).

# How governments can harness the power of automation at scale

Process automation and technologies based on artificial intelligence can bring benefits across numerous functions of government.

*by Jens Riis Andersen, Matthias Daub, Andrew Goodman, and David Taylor*



© BreatheFitness/Getty Images

**Governments around the world** are under pressure to operate more efficiently, serve citizens better, and provide more satisfying working environments for their employees. Lessons from the private sector show automation at scale has the potential to serve those purposes, but to get there governments must become more strategic in their approach, embrace new technologies, and be prepared to act at scale. Process automation and technologies based on artificial intelligence can bring benefits across numerous functions of government, including much lower operating costs, more efficient processes, and less wastage and errors. McKinsey estimates that as many as four out of five processes in HR, finance, and application processing are at least partially automatable, with the potential to reduce costs by at least 30 percent.

The benefits of automation can be achieved relatively quickly. Many of the solutions can be built on existing IT systems without significant additional investment. The approach to rolling out automation at scale is intuitive, starting with an assessment of the opportunity, launching pilots, building the infrastructure required, and then scaling.

Many governments have already made significant progress in creating online processes for citizens to complete applications and communicate with providers of services. The next stage is to use automation at scale to bring internal operations and processes up to date, helping them become digital organizations at their core.

### **The private sector has taken a lead**

Many private-sector companies have implemented automation at scale, helping them reduce operating costs, improve service offerings, work faster, and cut mundane manual tasks and processes. A leading oil company has cut four days from its financial-close process through automation of more than 10,000 tasks. One insurer has automated 120,000 transactions per month across 14 processes, realizing cost savings of about 30 percent per process. A large telecom operator uses automated processes

for more than 400,000 transactions per month. Governments can take lessons from their approach.

Private firms have married lean process design, which is focused on minimizing waste and maximizing value, with robotics and machine learning to push automation into new activities, many of which previously required human input. Processes such as procurement, from purchase request to order, are now automated to operate around the clock and at around a third of the cost of manual approaches. A key private-sector focus has been administrative activities, which account for around a quarter of public-sector employment, suggesting governments can make significant efficiency and productivity gains in that area.

Automation offers accuracy, consistency, scalability, and traceability. The impact in government is likely to be an improved service offering, more transparency, and more consistent data and analysis for tasks such as crime prevention. Automation can also boost employee satisfaction—repetitive manual work is frequently cited as one of the main sources of public-sector job dissatisfaction. Six technologies in particular are likely to be useful in driving the change process (exhibit).

### **Opportunities for automation at scale in government**

In recent years, some governments have made significant progress in digitizing citizen-facing services. Among many examples, the UK Government Digital Service has successfully migrated departmental publishing onto the GOV.UK platform and digitized services, including passport and driver's-license applications. Estonia's tax-filing system allows 95 percent of residents to file their tax returns online, and the US Digital Service has built a digital College Scorecard.

Still, while some interfaces have gone digital, automation of core processes has proved to be more challenging, denying governments the chance to achieve efficiencies at scale. In finance, HR, and procurement, some 60 to 80 percent

## Six technologies in particular are likely to be useful in driving the change to process automation.

### 6 keys to driving the change process

 <p><b>Robotic process automation</b></p> <p>Automates routine tasks (eg, data extraction, cleaning), using solutions such as Automation Anywhere and Blue Prism</p>	 <p><b>Smart work flows</b></p> <p>Integrates tasks performed by humans and machines (eg, month-end processes), with providers such as Appian and Pegasystems</p>	 <p><b>Optical character recognition</b></p> <p>Converts typed, handwritten, or printed text into machine-encoded text (eg, PDF/JPG image to text/CSV), using tools such as ABBYY and IRIS</p>	 <p><b>Machine learning</b></p> <p>Uses algorithms to enable systems to learn from experience without being explicitly programmed, using platforms such as SparkBeyond</p>	 <p><b>Natural-language processing</b></p> <p>Analyzes text and speech using learning algorithms (eg, data-to-story translation), with providers such as Arria NLG and Narrative Science</p>	 <p><b>Cognitive agents</b></p> <p>Creates a virtual workforce capable of supporting employees and customers (eg, employee service centers), using tools such as IBM Watson and IPsoft</p>
---	--	---	---	---	---

of tasks are automatable, creating potential for net long-term savings (after accounting for implementation and ongoing software costs) of at least 30 percent:

- **Finance.** Government departments conduct many of the same core finance processes as private-sector firms—they disperse cash, manage month- and year-end financial-close processes, and conduct budgeting, financial planning, and analysis—and often use the same enterprise-resource-planning and financial systems. In the United Kingdom, 11,000 civil servants work in the finance function across 25 government departments, many as large in budget terms as FTSE 100 companies. Some 80 percent of tasks performed in private-sector finance functions have at least some potential for automation, according to McKinsey research, and a similar proportion can be assumed in government. General accounting operations, cash disbursement, and financial control are standout areas of opportunity. One large European utility piloted the automation of new vendor creation and of period closing and found the processes were 70 and 50 percent automatable, respectively. McKinsey’s own Finance Shared Services Center has realized 50 percent automation of expense allocation.

- **HR.** Government is the single largest employer in many countries. In the United States, the Office of Personnel Management alone employs more than 5,000 people (in addition to HR staff in departments and agencies) to manage the federal government’s workforce of approximately 2.8 million people. Some 80 percent of HR processes are at least somewhat automatable, with payroll administration, record keeping, benefits administration, and recruitment administration—functions that many governments still perform in-house—at the top of the list. In the private sector, one large energy provider was able to automate 90 percent of its onboarding process, including ordering and delivery of passes, phone, and office equipment ready for day one, leading to more than 20 percent cost savings.

- **Applications and processing.** Governments process applications and claims for a range of payments and services, from social welfare to visas and tax returns. There has been significant innovation in the citizen-facing front end over the past two to three years. However, in part because of legacy IT systems, digital front ends find themselves interfacing with clunky back offices and siloed databases, leading to rekeying requirements. Automation can help by

reading and writing data between applications, checking consistency and completeness, and even sending and interpreting emails, bridging the gap between digital front ends and legacy systems. The efficiency potential is significant. One insurance firm used software to automate subrogation claims processing and reduced time per claim from 10.0 minutes to 3.5 minutes, increasing the volume of claims processed per week by a third.

## Getting started on automation at scale

The fact that much of automation software sits on top of existing technology stacks means it is possible to get started quickly—with the initial phases completed in weeks. To make early gains and promote internal buy-in, we suggest governments or ministries take a three-step approach, comprising start-up, launch, and scaling.

### 1. Start-up: Understanding automation potential and proving the concept

**Assess the opportunity.** Governments or individual ministries embarking on a program to automate at scale should begin with a top-down assessment of potential gains across administrative and corporate functions, citizen-facing services, and public-leadership roles. The private sector provides useful benchmarks to inform the process. Project leaders should set out their plans to develop use cases over a given period, with targets defined in terms of financial impact or volume, or both. The vision should be transformational, rather than incremental or tactical, aiming, for example, to impact more than 50 percent of tasks and processes. In the private sector, one global bank set a goal of \$1 billion of financial impact from enterprise-wide automation over three years.

**Select technology partners.** Once the assessment phase is complete, leaders should launch a process to select a portfolio of technologies and vendors. Procurement may be governed by a central framework, but it is likely to leverage the services of a range of vendors and include a number of platforms and tools catering to different use cases identified in the initial assessment.

### *Pilot proofs of concept based on use cases.*

Having put the requisite vendor relationships in place, ministries should conduct proof-of-concept sprints for specific use cases. The sprints should be run on an agile basis, allowing flexible development and decision-making. Common initial use cases include invoice verification, vendor procurement, expense allocation, and month-end close processes.

This start-up phase can be completed in as little as eight to 12 weeks, assuming access to the right data, and with a team of fewer than ten people. The Danish government has launched pilots to test both robotic process automation in shared-services centers and machine learning.

### 2. Launch: Building the infrastructure for automation at scale

The physical, operational, and human infrastructure required for automation includes new capabilities and cross-functional teams, governance frameworks, and suitable IT and data models.

**Build cross-functional teams.** Cross-functional teams are necessary to work across capabilities and may include a delivery lead responsible for multiple processes (to plan, manage delivery, and remove obstacles), an automation architect (to design target solutions), developers (to code, configure, test, and improve), and a business analyst (to manage specifications and coordinate handover to end users). The teams need access to subject-matter experts and to IT and legal support. It is a good—but rare—practice to designate an owner for each process to be automated.

**Establish clear governance.** As more processes are automated and delivery teams grow, dedicated governance frameworks are required. Most private-sector firms opt for a federated model, with a center of excellence (COE) coordinating automation across lines of business or activity and creating links to continuous-improvement teams. During the start-up phase, development activity typically takes place in the COE. When moving to the launch phase, development and delivery teams are commonly located in the relevant line of business or activity.

*Agree on the IT and data model.* Successful automation at scale requires a coherent IT and data model that maximizes flexibility. However, it's important not to get locked in to one protocol. This should include clear direction about where to involve the IT department—for example, in installing platforms, deploying software, and ensuring information security.

Moving from start-up to a successful and sustainable launch of automation at scale—for example, automating ten or more processes and with the relevant infrastructure in place—can take six to 12 months, building a team of 15 to 20 people along the way. As use cases develop, some employees who have been co-opted into COEs may return to ministerial duties.

### **3. Scale: Sustaining and delivering value from automation at scale**

Governments have found it particularly challenging to scale automated solutions. This is in part structural—ministries tend to work in silos—but can also be cultural (for example, manifested in risk aversion in relation to IT projects) and a result of talent shortage, amid intense competition for expertise.

*Establish centers of excellence.* An essential step in making the move to sustainable scaled solutions is to work programs based on a road map of priority processes and planned workforce changes. Typically, the COE will set out a standardized approach and provide program-management resources, but individual ministries or teams will drive delivery. Senior-management teams can provide support and encouragement by setting targets for back-office (such as finance, HR, and procurement) and citizen-facing functions (such as applications and claims).

*Invest in capability building.* Private-sector firms that have successfully launched automation at scale have made it a priority to invest in capability

building, in particular, ensuring not to become overly reliant on external suppliers and putting in place the HR processes to attract, develop, and retain individuals with the relevant technical skill sets. Employees can also benefit—in acquiring the basic skills around automated processes, they can improve the efficiency and quality of their daily work and equip themselves for future opportunities. A key element is to educate technology users, who are often surprisingly averse to using new tools, even when they find their old systems frustrating.

#### *Plan and budget for ongoing maintenance.*

Departments, grounded in the business case for each process, should plan and budget for ongoing maintenance and support of automated processes. Failing to do so creates an unrealistic view of efficiencies and will mean that changes may not be financially sustainable.

---

The opportunity for automation at scale in government is significant and accessible, and private companies are already seeing the results. The same tools these companies use can help governments reduce costs, improve the employee experience, and provide faster and better services. However, governments must take urgent action to instigate change, as they have done for some citizen-facing services. That means developing methodologies that can be hardwired into the automation process and taking a strategic approach to planning, piloting, and scaling.

**Jens Riis Andersen** is a partner in McKinsey's Copenhagen office, **Matthias Daub** is a partner in the Berlin office, and **Andrew Goodman** is a partner in the London office, where **David Taylor** is an associate partner.

Copyright © 2019 McKinsey & Company. All rights reserved.

# Cracking down on government fraud with data analytics

New data tools are giving government agencies the upper hand in taming fraud, waste, and abuse. Lessons from scaled approaches show how to capture the hundreds of billions of dollars at stake.

*by Susan Cunningham, Mark McMillan, Sara O'Rourke, and Eric Schweikert*



© Teekid/Getty Images

**Crime, the old saying goes, does not pay.** Or does it? Within US federal government programs, for example, fraud, waste, and abuse (FWA) are widespread, largely unmeasured, and a growing drain on taxpayers and citizens, reducing the effectiveness of government services. Given how difficult it is to identify and measure fraudulent activity, the precise extent of monetary losses is not known. Yet consider that the US government each year makes more than \$140 billion<sup>1</sup> of improper payments—defined broadly as those funds that go to the wrong recipient, for the incorrect amount, for which documentation is not available, or that the recipient uses improperly. Many of these payments are essentially paperwork errors, but about 40 percent result in a monetary loss to the government. Add in payments that are intentionally misspent or directed to the wrong recipient, and some experts estimate federal government losses from potential fraud at nearly \$150 billion (Exhibit 1).

Such estimates frame the size of the opportunity that better policing could capture for governments everywhere. In the United States, fraud, waste, and abuse against federal agencies takes many forms, including identity theft by criminals who submit fraudulent tax forms to steal someone else’s refund, wasted procurement and healthcare spend due to devious third-party providers, and an inability to identify redundant payments or payments for services that did not occur. More than 70 percent of these losses come from programs that involve payments to citizens or third-party providers (Exhibit 2).

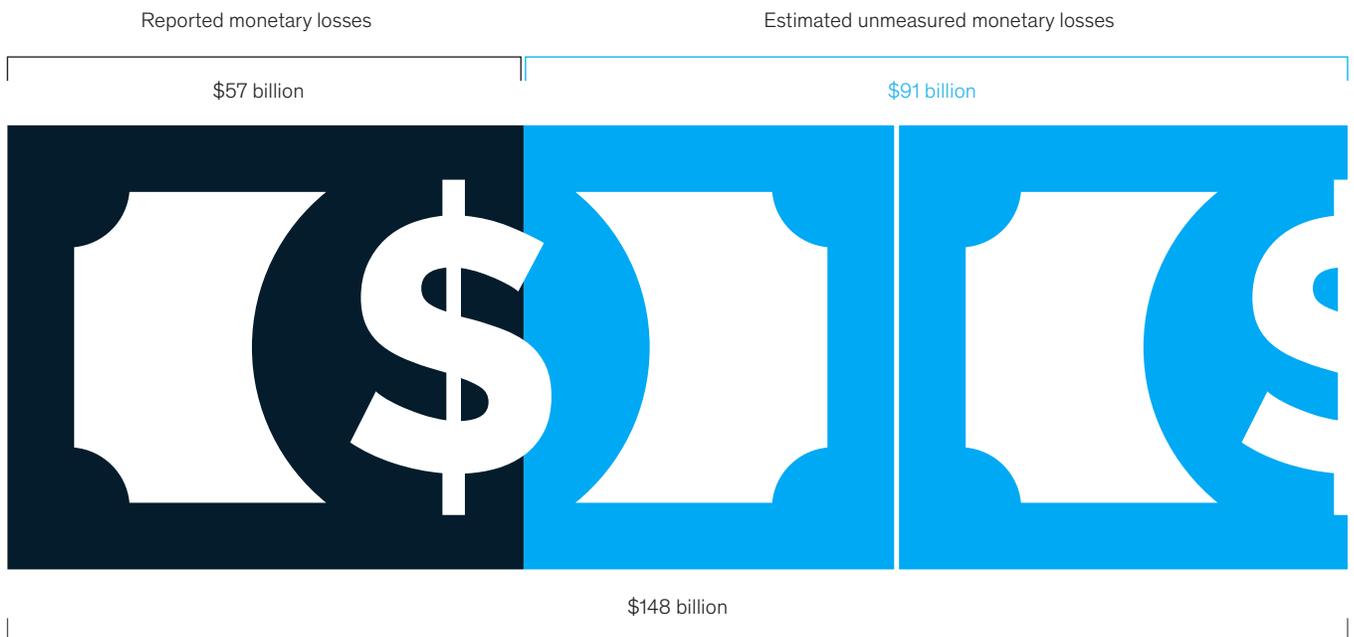
Stemming these losses is important for ensuring a well-functioning, credible government. They erode the government’s efficacy and the public’s trust, wasting money and preventing agencies from properly fulfilling their missions to citizens,

<sup>1</sup> “Government efficiency and effectiveness: Opportunities to address pervasive management risks and challenges while reducing federal costs,” US Government Accounting Office, May 17, 2017, gao.gov.

Exhibit 1

## More than half of monetary losses to fraud, waste, and abuse go undetected.

A breakdown of missing US government funds through fraud, waste, and abuse in 2017, \$ billion

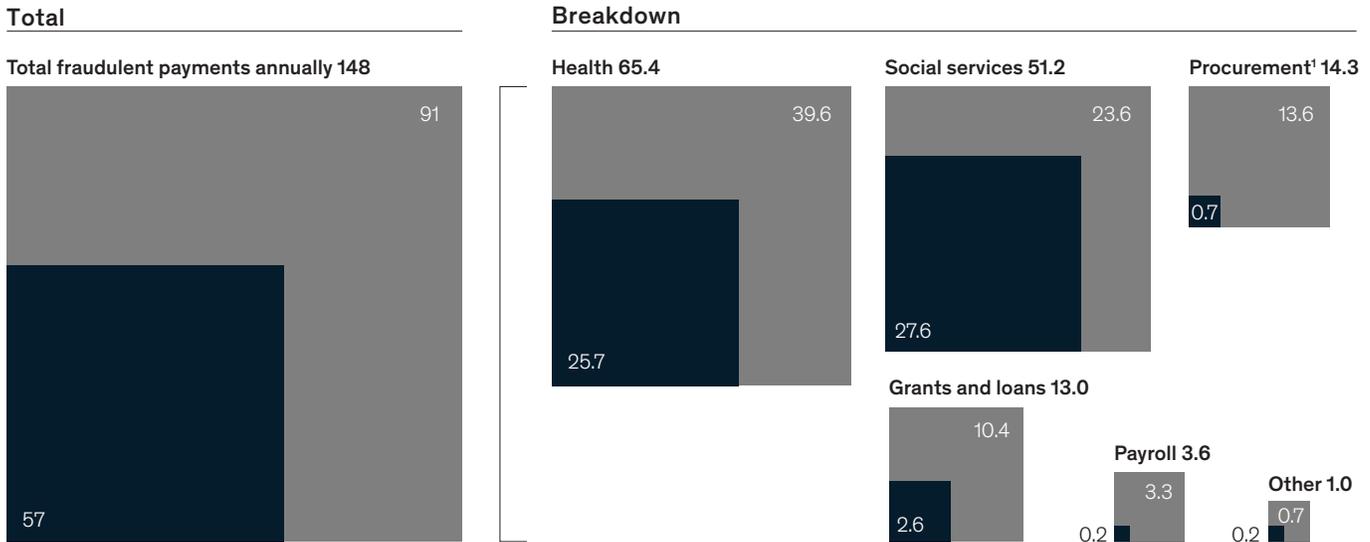


Analytics can both prevent and treat measured loss more efficiently and detect additional unmeasured loss.

Source: The Association of Certified Fraud Examiners annual fraud, waste, and abuse (FWA) report, 2018; independent analysis of government programs; Paymentaccuracy.gov

## Today, we estimate most government programs detect less than half of monetary loss due to fraud, waste, and abuse.

Detected and undetected monetary loss, by US government area, \$ billion



Note: Figures may not sum, because of rounding.

<sup>1</sup> Procurement estimates exclude procurement waste and inefficiencies, which may contribute an additional 15–50% loss in payments to third-party vendors or states, due to duplicate or unneeded services.

Source: The Association of Certified Fraud Examiners annual fraud, waste, and abuse (FWA) report, 2018; independent analysis of government programs; Paymentaccuracy.gov

businesses, and partners in government. Further, the risk of inaction is high: as private organizations tighten their defenses, criminals increasingly turn their focus to target government programs, as evidenced by the surge in identity-theft attempts against the Internal Revenue Service (IRS) in the early 2010s.

The good news is that a number of government agencies have had success focusing on combating the negative effects of fraud, waste, and abuse by using more advanced analytical methods to detect additional unmeasured losses and prevent and treat measured losses more effectively. Organizing to make effective use of these tools and the required skills isn't always easy. However, when the work is done well, we've seen returns on investment ranging from 10:1 to 15:1. These stories provide a road map for successfully improving detection and prevention that is applicable across a wide range of institutions.

### Combating FWA through better analytics

Let's be blunt: in our experience, agencies looking to ferret out FWA often measure what they can but not what matters most. In some cases, such as measuring improper payments, they focus on what's simplest to measure, for instance, administrative errors (such as the lack of proper signatures), which often do not result in monetary loss. In other cases, the fraud that agencies do identify, often after the payment has gone out the door, is difficult for them to recover. In still others, there is a significant amount of fraud that agencies are simply not currently able to measure, in part because it reflects intentional misrepresentation by vendors or benefit recipients, sometimes as the result of organized criminal schemes.

Granted, it is not exactly straightforward to establish a comprehensive approach to tackling FWA that

employs analytical tools and data sources to improve detection, and then to organize and scale the institution to deploy that approach. Data are often unstructured, incomplete, and located in silos across the institution. Staff are likely working in a resource-constrained environment—even while we find that tackling fraud, waste, and abuse generates impressive savings for the US Treasury, those savings rarely result in a bigger agency budget to fund the fight. With competing priorities, many agencies simply fail to drive the organizational changes and sustain the momentum needed to achieve impact.

Yet some have overcome these challenges by tapping into new analytics tools and treatments, establishing the organizational changes needed, and recruiting new skill sets to transform their anti-FWA programs in an environment of constrained resources. Armed with more—and more-timely—data, better analytical tools, and enhanced processing power, agencies have an opportunity to identify new types of fraud, including major schemes and networks. For example, the US Securities and Exchange Commission (SEC) has leveraged new tools to pull in and connect large amounts of disparate internal and external data to identify new patterns of insider trading. The agency has also turned to machine learning and text-analysis techniques to apply analytics to what were once paper-based enforcement and compliance reviews.

Or consider the US Department of Health and Human Services' Centers for Medicare and Medicaid Services (CMS). CMS built the Fraud Prevention System (FPS), which uses advanced analytics to identify, prevent, and stop payments that match certain suspicious patterns and to raise the priority of other suspect payments for investigation. FPS helped CMS prevent \$527 million in losses to fraud in fiscal year 2016 as part of a suite of program-integrity efforts that saved CMS almost \$18 billion a year at a return on investment of 12.4:1.<sup>2</sup> Or consider the Department of Veterans Affairs (VA), where a focused effort applies analytics to healthcare payments by

scanning for incorrect treatment combinations, unbundling of procedures, duplicate payments, outliers, and other questionable activities in payments to external providers.

We also see global examples of analytics-led efforts to reduce FWA through targeted, low-cost interventions. In Australia, the Department of Health has seen a 19.5 percent reduction in urgent after-hours Medicare billing, which has saved the federal government about \$12 million, by using analytics to identify 1,200 high-claiming doctors and sending them letters to “voluntarily acknowledge” incorrect claiming of urgent items and encouraging them to shift their claiming behavior away from urgent items to standard items. This behavioral-economics approach, seeded in the theory that targeting the provider by assessing their holistic claiming patterns rather than managing individual claims through rules and alerts, is evidence of the effectiveness of prevention-focused, low-cost techniques. Making the most of these interventions depends on accurate advanced analytics, and ensuring that the right clustering techniques are deployed to ensure providers are compared on an “apples to apples” basis.

### **Scaling up and cracking down at the IRS**

In the early years of this decade, the commissioner of the IRS determined that increasing the use of data and analytics would be one of the agency's six key goals. A cornerstone of the effort was the creation of the Office of Compliance Analytics (OCA), an analytics center of excellence reporting directly to the commissioner that integrated state-of-the-art analytics talent deeply into day-to-day operations. The center also employed “test and learn” pilots to quickly gauge the effectiveness of innovations and refine them with less risk.

While the IRS recruited the new unit's leadership externally, the focus was on building sufficient analytics capabilities within the agency to be able to solve some of the agency's toughest problems. Another key design element was tight integration with the operating divisions. Executives from the businesses were detailed to OCA to lead the projects

---

<sup>2</sup> Annual report to Congress on the Medicare and Medicaid Integrity Programs, Centers for Medicare & Medicaid Services, 2016, cms.gov.

# With competing priorities, many agencies simply fail to drive the organizational changes and sustain the momentum needed to achieve impact.

in their home units, ensuring the analytics answers were both relevant to and actionable by the business. Agency leaders took part in selecting OCA's portfolio to create a balance: quick wins that established credibility, longer projects to demonstrate big impact, and selected fire-fighting support to build goodwill with divisions under pressure.

When moving from modeling to action, OCA focused on rigorous testing using randomized, controlled trials to tease out and validate the best combinations of analytics models and operational approaches. For instance, to deter tax preparers from coaching clients to misstate items on tax returns, the OCA team launched a series of different tactics, from “nudge” letters to in-person visits by IRS investigators. By maintaining control groups during implementation, the team continually measured proof of impact. With the help of sophisticated modeling techniques and tight integration with operations, OCA was able to effectively address a diverse range of issues, from identity theft to small-business cash underreporting to abuse of the earned income tax credit. Collectively, OCA's projects delivered more than \$30 billion worth of improved tax compliance annually by its third year of operation.

## Keys to success

Analytics-driven FWA programs are examples of US government efforts to create nimble and effective approaches to keep technologies and workforce skills current.<sup>3</sup> Few agencies have tackled fraud

problems with comprehensive transformation efforts, which in our experience can take years and as much as \$10 million annually to address \$100 million worth of fraud, waste, and abuse. But even gradual programs, executed well, can deliver success and savings for agencies and taxpayers. While many agencies have the pieces needed to succeed, the most difficult step is bringing these together to achieve a sustained and successful effort that can be scaled. Almost all agencies, when they think about launching an analytics program, correctly consider the need to acquire analytics technology, hire and train staff in analytics, and make data more easily available. However, as evidenced by many of the case examples, there are five other factors that, when taken together, form the essential core of enablers needed to drive a successful analytics program:

- **Analytics tools and skills.** There are many advanced analytical techniques that allow agencies to identify FWA, such as artificial intelligence, cluster analysis, outlier analysis, network analysis, machine learning, “fuzzy matching,”<sup>4</sup> and others. For instance, the IRS has used network analysis to use identified types of fraud to uncover groups of related bad actors. Taking advantage of these techniques often involves using an agile approach to build analytics “sandboxes” and expanding or elevating existing analytics expertise. While external parties can help jump-start an advanced-analytics program, agencies should

<sup>3</sup> “President’s management agenda,” President’s Management Council and the Executive Office of the President, March 2018, [whitehouse.gov](http://whitehouse.gov).

<sup>4</sup> “Fuzzy matching” refers to a technique where partial matches are used to link records together. For instance, a social-security number with transposed digits would match seven of nine digits, and in combination with an exact match on address could be sufficient to assume the records should be linked.

focus in parallel on consolidating and elevating existing analytics efforts, and recruiting or raising the skills of individuals who can fill required roles: data engineers, data scientists, test designers, and “translators” who can bridge business and technical staff.

- **Access to broad sets of quality data.** Data are the lifeblood of analytics, and for maximum impact, the relevant data need to be made available in near-real time at the point of decision-making in order to stop fraudulent or abusive payments before they go out the door. Many agencies have siloed data or incompatible data sets across the organization, so a robust data-management capability is necessary to improve data collection, maintenance, storage, and integration. Some agencies have built “data lakes” before really knowing what actions they intend to take—and therefore what analytics will be needed. This can be very costly in terms of time and money, with little to show in regard to impact. We find that an agile approach to data—focusing on only the limited data required to achieve a particular goal—can deliver results much more quickly and cost-effectively. The SEC, for example, has successfully made data management a major priority for the commission, through robust communities of interest regarding analytical approaches, new initiatives to streamline governance of core data sets, and the creation of a data catalog to enable awareness of the use case and access points of data, as well as other initiatives to drive value and reduce cost.
- **Domain expertise.** Many analytics-led efforts fail to show impact because the team is insufficiently familiar with the business processes their model must affect. This often results in the team solving the wrong problem or coming up with a solution that can’t be implemented. Thus, it is important to partner the analytics team with leaders and frontline managers from the operations team on the ground, who will be driving the change. This expertise can come from internal or external sources, depending on the situation. For example, the Federal Emergency Management

Agency built upon insights it gleaned from payment-card providers to shut down an identity-theft ring in its emergency-support program. Rigorous information exchange between the analytics and operations teams strengthens efforts to continuously refine the analytics goal and help work through implementation hurdles.

- **Ability to operationalize insights.** Translating insight into action is another common point of failure for analytics efforts. Many successful programs begin with an end goal in mind and aim the entire effort at achieving that action. Once fraud, waste, or abuse is detected, agencies should develop, test, and refine scalable and cost-efficient interventions. Advanced analytical approaches are likely to unearth large new instances of fraud, waste, and abuse, so agencies need to be prepared to address the problem in volume. If the agency responds with a new action—for example, by preemptively stopping payments suspected of stemming from identity theft—then an iterative test-and-learn approach is crucial to analyzing and scaling a response. The IRS, for instance, was able to stop more than one million identity-theft tax returns a year by shifting the burden for verifying identity from the IRS to the likely thief, relying on models that generated a sufficiently low rate of errors in stopping refunds.
- **Strong executive sponsorship.** While many FWA-analytics programs demonstrate impressive returns on investment, such efforts still require executive sponsorship and interest, given the investment involved. Successful programs also involve setting priorities for hiring and matching analytical talent with staffers close to operations. Ensuring program impact requires senior leadership to actively champion FWA as an institutional priority, create sufficient room in the budget for a sustained effort at scale, define clear lines of authority and responsibility, and rigorously measure and monitor results. For instance, when one state instituted an analytics unit to address FWA in its Medicaid program, it established the unit just two levels below the agency’s director.

## Getting started

Whether an FWA effort is big or small, there are several approaches critical to organizing it and building and managing critical enabling capabilities. First, it's important to set aspirations correctly from the beginning. Agencies start to attack FWA from many different places—they may have the talent, but not the tools; the tools, but not usable data; the data, but not the executive sponsorship needed. Leadership must first identify the kind of impact the agency wants to have, as well as its level of maturity across the important enabling capabilities, and progress with a roughly equal emphasis across all five factors.

A diagnostic that builds an independent fact base on current performance can help. An agency can quickly scan the current environment, assess the available data, evaluate current FWA detection and prevention infrastructure, and better understand the current use of analytics throughout the organization. From there it is possible to help identify and size the potential value of an analytics-led FWA effort and set priorities for specific areas of focus.

When the effort moves into the execution phase, agencies can consider adding experts with real-world, relevant experience to jump-start the effort, possibly private-sector leaders from the finance or high-tech sectors. In addition, we find that a strategy that focuses on delivering—and measuring—quick wins from analytics is important for building momentum and support. Aiming to deliver value in the first three or four months can

help secure funding for future phases of a program. For instance, an agency paying benefits could initially focus on identifying situations where a large number of beneficiaries appear to share a bank account and evaluate it for possible identity theft, or, as CMS does, look for service providers billing for more services than are possible to deliver, such as doctors billing for more than 24 hours in a day. These demonstration cases can illustrate the analytics techniques, innovative treatments, integration with operations, and test-and-learn processes that underpin success in analytics, all while delivering value and building staff capabilities.

---

Governments have a tremendous opportunity for both near- and long-term fraud, waste, and abuse savings. By leveraging advanced-analytics tools, new treatment mind-sets and staff skills, agencies can prevent billions of dollars of losses.

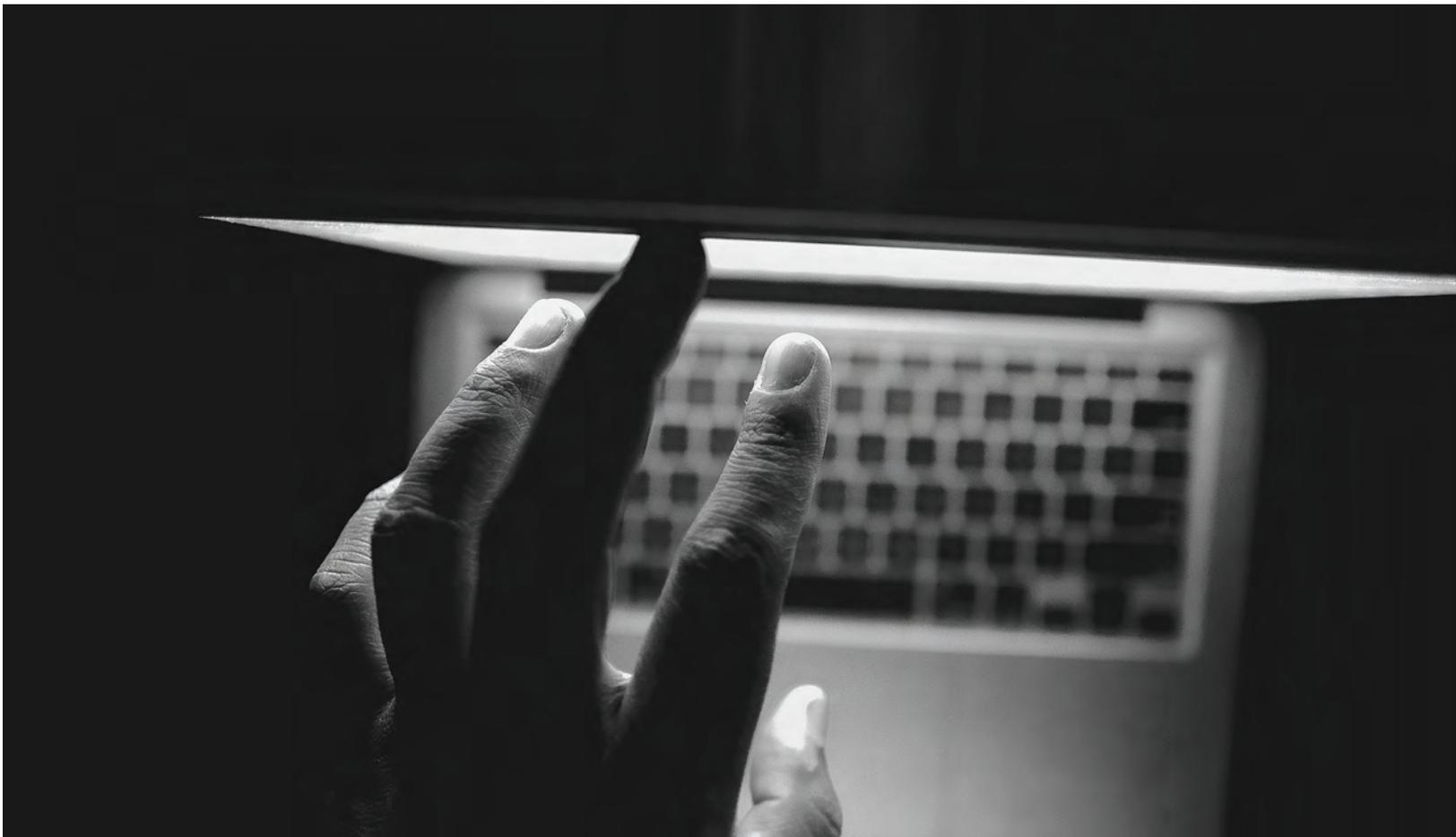
**Susan Cunningham** is a senior expert in McKinsey's Washington, DC, office, where **Mark McMillan** is a partner, **Sara O'Rourke** is an associate partner, and **Eric Schweikert** is a senior expert.

The authors wish to thank Damien Bruce, Sian Griffiths, and Akshay Y. Gupta for their contributions to this article.

Copyright © 2019 McKinsey & Company. All rights reserved.

# Defense of the cyberrealm: How organizations can thwart cyberattacks

Governments and companies have much work to do to protect people, institutions, and even entire cities and countries from potentially devastating, large-scale cyberattacks.



© Towfiqu/Getty Images

**In this episode of the *McKinsey Podcast*,** Simon London speaks with McKinsey senior partner David Chinn and cybersecurity expert Robert Hannigan, formerly the head of GCHQ,<sup>1</sup> about how to address the major gaps and vulnerabilities in the global cybersecurity landscape.

**Simon London:** *Hello, and welcome to this edition of the McKinsey Podcast, with me, Simon London. 2018 was a year of good news and bad news in cybersecurity. The year passed without a major international incident, certainly nothing on the scale of the WannaCry ransomware attack, in 2017. And yet, every few weeks brought news of another big data breach at another big company. So where do we stand going into 2019? Are we winning, in any sense? When and where will the next so-called tier-one attack occur? And, importantly, what is the role of government in helping to ensure national cybersecurity. To find out more, I sat down in London with David Chinn, a McKinsey senior partner who works with public- and private-sector organizations on these issues, and also with Robert Hannigan, who is the former head of GCHQ, the UK government's electronic-surveillance agency. Robert also led the creation of the UK National Cyber Security Centre, or NCSC. Today he's a McKinsey senior adviser. Robert and David, welcome to the podcast.*

**David Chinn:** Thank you, Simon. Glad to be here.

**Robert Hannigan:** Thanks.

**Simon London:** *I think for a layperson, the general question around cybersecurity is, probably, are we winning?*

**Robert Hannigan:** No, I think we are making progress, but I think it would be very rash to say we're winning. If you look at the two big trends, the rise in volume of attacks and the rise in sophistication, they are both alarming. On volume, particularly of crime, there were something like 317 million new pieces of malicious code, or malware, [in 2016]. That's nearly a million a day, so that's pretty alarming.

On the sophistication, we've seen, particularly, states behaving in an aggressive way and using very sophisticated state capabilities and that bleeding into sophisticated criminal groups. It's a rise in the sheer tradecraft of attacks. So no, I don't think we're winning, but I think we're doing the right things to win in the future.

**David Chinn:** I would agree with Robert. We may not have seen a single attack that brought down multiple institutions in the same way that WannaCry did, but look at the list of institutions reporting very sizable breaches of increasingly sensitive data.

Now we've got some more regulation forcing people to be more transparent about the breaches and the length of time that attackers were inside networks before being discovered. And it's not always clear to those attacked what they've lost. I'm broadly pessimistic.

**Simon London:** *When you think about where the next tier-one attack might come, what are some of the vulnerabilities that in business and government people are thinking about, talking about?*

**Robert Hannigan:** I think most of the focus now is on supply-chain and upstream risk, because even the best-defended companies now realize that their vulnerability is either those who are connected to their vendors, their suppliers, even their customers. And, increasingly, government is worrying about the IT infrastructure, so the global supply chain, both hardware and software, and its integrity.

And some of the state attacks we've seen in the last couple of years have been against the backbone of the internet, if you like. Routers, switches, places that give you massive options to do different things with internet traffic (Exhibit 1). It's going deeper and more sophisticated.

**David Chinn:** I think there's different versions of what tier one might feel like. I think that the increasing ability of both criminals and states to attack critical infrastructure [is one of them]. Taking

---

<sup>1</sup> Government Communications Headquarters.

## Companies should assess threats and develop controls to the most critical.

Assets	Threats	Controls
 <b>Data</b>	<ul style="list-style-type: none"> <li>• Data breach</li> <li>• Misuse or manipulation of information</li> <li>• Corruption of data</li> </ul>	<ul style="list-style-type: none"> <li>• Data protection (eg, encryption)</li> <li>• Data-recovery capability</li> <li>• Boundary defense</li> </ul>
 <b>People</b>	<ul style="list-style-type: none"> <li>• Identity theft</li> <li>• “Man in the middle”</li> <li>• Social engineering</li> <li>• Abuse of authorization</li> </ul>	<ul style="list-style-type: none"> <li>• Controlled access</li> <li>• Account monitoring</li> <li>• Security skills and training</li> <li>• Background screening</li> <li>• Awareness and social control</li> </ul>
 <b>Infrastructure</b>	<ul style="list-style-type: none"> <li>• Denial of service</li> <li>• Manipulation of hardware</li> <li>• Botnets</li> <li>• Network intrusion, malware</li> </ul>	<ul style="list-style-type: none"> <li>• Control of privileged access</li> <li>• Monitoring of audit logs</li> <li>• Malware defenses</li> <li>• Network controls (configuration, ports)</li> <li>• Inventory</li> <li>• Secure configuration</li> <li>• Continuous vulnerability assessment</li> </ul>
 <b>Applications</b>	<ul style="list-style-type: none"> <li>• Manipulation of software</li> <li>• Unauthorized installation of software</li> <li>• Misuse of information systems</li> <li>• Denial of service</li> </ul>	<ul style="list-style-type: none"> <li>• Email, web-browser protections</li> <li>• Application-software security</li> <li>• Inventory</li> <li>• Secure configuration</li> <li>• Continuous vulnerability assessment</li> </ul>

Source: European Union Agency for Network and Information Security; The SANS Institute

out power to a city might have relatively limited impact in terms of the actual damage done, but could have a huge impact on the way people feel.

**Robert Hannigan:** There’s a difference between a genuinely catastrophic damaging attack and a politically sensitive attack that spreads fear and terror or a lack of trust in data. It’s fairly easy to imagine things that will lead to public panic.

You’ve seen big public controversies over airlines and banks being unable to function, often not through cyberattacks. But if you were to multiply that and see it as a malicious attack, you could see genuine public disquiet, a lot of political pressure to do something about it.

**Simon London:** *Yes, it’s interesting, because when you talk about critical infrastructure of the modern economy, you often think about things, like, as you say, the internet backbone. It’s those kind of things. Or maybe financial services, the financial system. But just talk a little bit more*

*about the supply chain, for example. That’s one that I think in the broad conversation and the broad business public is less discussed.*

**David Chinn:** If you think about, at the simplest level, how a pint of milk gets onto the supermarket shelf, there are many stages in that, from the farm—by the way, the cows are milked by a machine, which is probably connected to a network—through to the transport network. The cold chain. The monitoring of the cold chain.

You don’t need to disrupt anything except the record that says the milk was kept cold for it no longer to be a product that can be given to the public. The integrity of that data is the essential glue that sticks it all together.

**Robert Hannigan:** If you think of the big ransomware attacks of WannaCry and NotPetya a couple of years ago, one of the lessons from those is that although they almost certainly weren’t targeting big manufacturing enterprises in Europe,

they effectively disabled quite a lot of household-name companies. They simply couldn't do business, couldn't manufacture for, in one case, several weeks. It was a wake-up call to sectors of the economy who thought they weren't a target for cyberattacks because they didn't have great IP or data that was worth stealing.

The Internet of Things is simply connecting more processes and more devices to the internet. And it is quite striking that the level of security built into those is usually very low because they're designed and built and procured on cost (Exhibit 2). There will probably be a role for regulation to improve the standards there.

But it does mean companies are, both through digitization and through the Internet of Things, increasing their attack surface, making it harder for them to understand the perimeters of their own networks, harder to see where their vulnerabilities are. That is a real problem for the next five, ten years.

**Simon London:** *And is this one of the reasons that people are very interested, for example,*

*in blockchain? The application of blockchain in the supply chain.*

**Robert Hannigan:** Yes, I think blockchain holds a massive potential because of the holy grail, really, of having a ledger that is distributed and unchangeable and visible to everybody. That has great benefits in cybersecurity. It's got a bad name because it's used for Bitcoin, and Bitcoin has a bad name, but I think blockchain technology is fantastic.

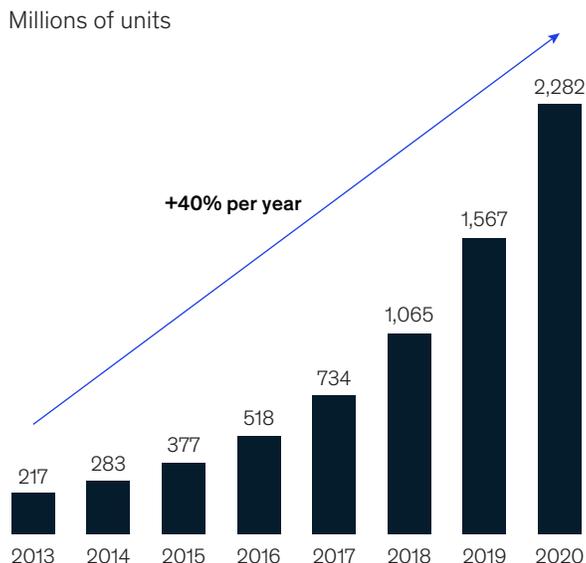
It's not straightforward to apply, and I think there's a lot of talk about it. The application in particular sectors for particular uses is still to be developed, to be honest. But it certainly ought to be a net gain for security, and particularly for data integrity, because one of the big future worries is it's one thing to destroy data or steal it or ransom it. To change it and undermine trust in data, particularly in financial services, could be catastrophic.

**Simon London:** *Or, indeed, milk, which is what gave me the thought. It's a very, very simple example, but it underlines how much of the economy runs on trust in that data.*

Exhibit 2

## Many professional building managers are not addressing Internet of Things security threats.

**Installed base of connected smart-building solutions is expanding rapidly...**



**...but most building managers have not addressed cybersecurity risks**

% of building faculty managers (n = 224)

**78%** Do not feel knowledgeable about cybersecurity issues related to building-automation systems

**46%** Do not monitor their building-automation systems for cyberattacks

**45%** Have not involved their company's IT department in implementing security measures

**63%** Have not developed a response plan in event of cyberattack

**71%** Have not taken measures to improve cybersecurity

Source: Gartner; IBM: smart-building facility-manager survey in *Building Operating Management*, January 2015

**Robert Hannigan:** We're just seeing criminals moving in this direction and looking at ways of looking at the corruption of data to, for example, affect stock prices. There's a huge potential there to use the changes to data, or to put out false data, to affect the value of a company.

**David Chinn:** Fake news is a great example. They haven't affected the integrity of the core data. They're just simply putting out noise. In the reports on the attacks of the integrity of the electoral system in the United States, in a system which is highly distributed, where different standards and technologies are used across the United States, there was clear evidence of attempts to penetrate electoral registers. You imagine changing the electoral register so that people of a certain party simply didn't appear. In the hustle and bustle of Election Day, they probably wouldn't get to place their votes. That could dramatically undermine trust in democracy.

**Simon London:** *Robert, we're lucky to have you on the podcast today. Why don't you talk a little bit about what is the role of government in all of this?*

**Robert Hannigan:** It's a challenge that every government is grappling with in different ways and has been over the last ten years. There are a couple of things that make cyber particularly difficult. One is cyberdefense undercuts the assumption that government can do defense for everybody.

David has spent a lot of his time dealing with government defense in a traditional sense. And you, as a citizen, expect government to defend you using the armed forces. It's unrealistic to expect government to do cyberdefense in the same way for the whole economy, because of the scale of it, and because most of what you're dealing with is outside government. Quite apart from the fact that the skills and resources just aren't there in government to do it on that scale. So that's one problem.

The other problem is that cyber is crosscutting in every sense. It is in a new domain, so it's a bit like discovering water or air. Every department, every part of the economy, is dependent on this, increasingly, as we digitize more and become more

dependent. You can't really point to a single bit of government and say, "You're responsible for cyber." That was the tendency in the early days.

The answer has to be to find a way of organizing government that gives sufficient speed and command and control to deal with the pace at which digital networks work and cyberattacks work but that actually drags out the whole of government to be good at cybersecurity, because if any one bit is bad at it, the whole system suffers.

**David Chinn:** Robert, it's interesting what you say because in a sense, government has three challenges. First, it is an actor in cyberspace in service of national interest, usually in secret. Second, it has to protect itself from cyberattack. And third, it has to create, at the minimum, an environment which protects the citizens and businesses of the country.

My observation would be that, at least reportedly, the UK is very good in the first. Your old institution is a world-class actor in the national interest in cyberspace. The second is quite hard, defending government, because there's so much of it.

The technical skills of government, government IT, are continually in the newspapers and in the public accounts committee as being something that we struggle to do well. Simple things, like putting a working computer on everybody's desk, let alone defending those networks.

**Robert Hannigan:** Most governments, including the UK, have focused their attention on protecting government networks, sometimes interpreted slightly more broadly to take in some critical bits of national infrastructure that really, really matter, but to encourage the rest of the economy to get better. So we spent ten, 15 years, in a sense, preaching at companies to get them to raise their standards.

There was quite a critical shift, certainly in the UK, about three or four years ago, where we decided that a security model that depended on everybody and every company doing the right thing all the time was almost bound to fail. The whole system was not designed with security in mind, so the

people who invented the internet and then the web that sits on it didn't have security at the front of mind, and so we are retrofitting that, and have been over the last 15 years.

Things like scanning websites for vulnerabilities, which is, again, being done across government, you could do nationally, and you could make that available nationally. One of the problems, I think, is that because the internet wasn't designed with security in mind, security is seen as something you need to add on rather than something that's built in.

We need to reach a point where security is designed in and is there by default, particularly with the Internet of Things. That may require some regulation and certainly will require bits of the economy, including insurance, to start to drag up standards.

**David Chinn:** Do you think government's been remiss on regulation? My observation would be that GDPR (Exhibit 3), which is not a cyberregulation, but that puts significant penalties on institutions for allowing private information to be misused, which includes being stolen, is having quite a big impact already in terms of reporting and transparency, which is then going to inevitably lead to more investment and more focus by organizations on protecting that data. Do you think government missed the boat a little bit on regulation?

**Robert Hannigan:** I think government, certainly in this country, has been reluctant to regulate, for all sorts of reasons. In cyber, there's a particular reason why regulation can be difficult, because it can end up being very prescriptive and very tick box, and it doesn't take account of the speed at which technology is changing and the particular networks that a company may have. We preferred an advisory, "Here are objectives you should meet"—a risk-based approach, I suppose.

**Simon London:** *Best practices and these kind of things.*

**Robert Hannigan:** Yes. Then there is a good case for saying we need a tougher approach on regulation. I think the EU is moving in that direction. I think GDPR

has been a net benefit, because essentially there are two sides to most cyberattacks. There's "Did you do the right things to prevent it, and then how did you handle it afterward?"

So GDPR has been particularly strong on the second bit. First of all, it's removed the debate in companies about whether they reveal the attack and how long, because they have to. That's good. It's raised awareness in boardrooms and so, to some degree, panic in boardrooms.

But I think the best regulation probably is in the states. It's interesting to see that California is introducing some hardware-IT supply-chain regulation, which will have a big impact, I think, given that so much of it is designed there, even if it's mostly made in China. There is a place for regulation, and we probably should have done more of it. The difficulty is lack of skills, again. I think most governments don't have sufficient skills.

**Simon London:** *Ah, well, that was going to be my next question. Yes. To your point, David, I mean government IT doesn't have a massively positive reputation in the world at large. Sometimes unfairly. But yes, do governments have the technical skills in cyber to protect their own networks?*

**David Chinn:** The interesting thing about cyber is that the source of innovation in attacks is mostly coming from inside governments. Many governments have very highly skilled people who when their knowledge leaks into the public domain gets adopted quickly by criminals. We have the equivalent of government weapons proliferation into cyberspace.

If you follow the cyberindustry, where there's a huge number of start-ups, effectively, each year's retiring crop of government hackers is bringing new innovation from inside the secret domains of government in an appropriately, hopefully appropriately, modified way to the benefit of those who are under attack, often from other governments. One can't say that there are no skills in government. The best skills are probably in government.

## The General Data Protection Regulation sets out guiding principles for data protection.

Principle	Explanation
<b>Lawfulness</b>	Data should be processed only when there is a lawful basis for such processing (eg, consent, contract, legal obligation)
<b>Fairness</b>	The organization processing the data should provide data subjects with sufficient information about the processing and the means to exercise their rights
<b>Transparency</b>	The information provided to data subjects should be in a concise and easy-to-understand format (eg, the purpose of consent should not be buried in a lengthy document of terms and conditions)
<b>Purpose limitation</b>	Personal data may be collected only for a specific, explicit, and legitimate purpose and should not be further processed
<b>Data minimization</b>	The processing of personal data should be adequate, relevant, and limited to what is necessary in relation to the purposes for which those data are used
<b>Accuracy</b>	Data should be accurate and kept up to date
<b>Storage limitation</b>	Data should not be held in a format that permits personal identification any longer than necessary
<b>Security</b>	Data should be processed in a manner that ensures security and protection against unlawful processing, accidental loss, damage, and destruction
<b>Accountability</b>	The data controller is responsible for demonstrating compliance

Source: Regulation (EU) 2016/679 of the Council of the European Union, European Commission, and European Parliament

**Robert Hannigan:** That’s true, but I wouldn’t underestimate the creativity and innovation of criminal groups. They are genuinely creative. They are talking to each other about, “How could we do this in a better way? How could we defraud this particular bank? What technique is going to work best? What’s the best way of delivering it?”

They are doing what so many traditional companies are trying to do, which is pull in skills from around the internet. Not necessarily colocated. They’ve clocked something about how to harness young innovative skills and do creative things. We have quite a bit to learn from them, I think. I agree that governments have been very good in quite a small and narrow way, but the criminal world is also pretty innovative.

**David Chinn:** I think this is similar, certainly in the UK, to the crisis in STEM education. If people don’t study

STEM subjects, we’re just not going to have the inflow into the economy, whether it be for government or private industry.

I’ve been particularly impressed by the way that Israel has effectively said that this is a national defensive-capability issue, but it’s a national industrial-growth issue. The country decided they wanted to have one of the world’s leading cyberindustry platforms and that to do that they had to make a massive investment in skills.

They started with after-school activities in the most deprived areas, because they recognized that if you start young enough in a country where almost every home has a computer, even those with very low means, who think that having a computer is important, that you can build those skills, in a sense, in parallel to formal education.

Many people who are extremely talented in the cyberdomain actually don't do particularly well at school. It's an outlet for those people, and I think it's been very, very successful. It's created a great pipeline of talent into government and private industry.

**Simon London:** *I think about another interesting question for government is how you manage this tension between the need for transparency and bringing the whole economy with you, and yet at the same time there is an element of secrecy, acting in the national interest and so on. How do you manage that tension in practice?*

**Robert Hannigan:** I think the key insight of the last ten years has been that you can't do cybersecurity in secret. You can't do it behind a wall in the intelligence agencies. For the obvious reason that the attacks are out there in open source in the economy, on the internet. It's all visible. Well, most of it is visible.

It makes no sense to try to do it in the way that you've tackled traditional security threats, which may be very, very secret and coming from very sophisticated governments. There is a side of that that is true for cyber, but most of it is not. Most of what people experience in cyber, whether companies or individuals, is crime. Some of it's state-backed crime, but still crime. And it simply doesn't work to be referring constantly to a secret world that can't really communicate.

The obvious development here has been to create a national cybersecurity center that was outside the secret world, but under the aegis and under the control of GCHQ, which is where the skills sat. And to have a blend of both. In the headquarters, you've

got access to secret systems for some people, but the key point is that you have openness to industry, and you have industry people sitting alongside government experts.

It goes back to our discussion of regulation. What you need in cyber, you can't simply have cyberregulators who do it for everybody, because so much is domain-specific. You need to understand the energy sector to regulate or advise on how to do cybersecurity of energy, or for any other sector. It's different. Therefore, the idea is to have experts from those particular sectors sitting literally alongside a deep cyberexpert.

**Simon London:** *To your point, David, it sounds like a lot of companies are struggling with this same cultural pull between the secrecy but the need to share information really to be effective, or to be more effective and to collaborate with your peers and share information.*

**David Chinn:** Yes, and I think we'll see the information commissioner shaping the environment around transparency quite actively in the very near future.

**Simon London:** *This is your point around regulation?*

**David Chinn:** Yes. I think that will really change people's understanding of how much they can legitimately keep secret.

**Simon London:** *Can we just internationalize the conversation a little bit? If you look across the international context, what are other governments who are doing this well and innovatively, and who we can all learn from?*

## Many people who are extremely talented in the cyberdomain actually don't do particularly well at school.

**Robert Hannigan:** I would say Singapore and Israel are doing it very well, in slightly different models. Australia has chosen a model that's similar to the UK model (Exhibit 4). Having it all in one place effectively. Certainly, the operational side of cyber.

Most governments are organizing and constantly tweaking the system. There are very different models, and in Europe, perhaps in Germany especially, the cyber agencies are purely civilian. And then there is a secret-world element of cyber, and I think they're also looking at how to bring those two together in a way that works for them, given the different constitutional setup.

The military in many countries has a primacy in cyber, and certainly in Germany it has been given a strong lead in cyberdefense. That brings both opportunities, because the military always has a

lot of resources and it's very good at organizing stuff. But also challenges, because it's not used to dealing with defending banks and the economy, and it's a culture shock for it. Leaders don't necessarily feel that's part of their remit. There are difficulties in the military.

The US, everybody looks to, but I think it's so large, with its multiplicity of agencies, that it's struggling. It has fantastic capabilities, obviously. The private sector is probably better organized, particularly in financial services, than anywhere in the world. But you often get the criticism or complaint from the private sector that the links to government are not quite right yet.

That I think reflects partly the fact that it's still evolving; the Department of Homeland Security, that was given this leadership under the

Exhibit 4

## The National Cyber Security Centre leads the UK government's cybersecurity work.

### Responsibilities:

---



Protect the UK's critical services from cyberattack.

---



Manage major cybersecurity incidents.

---



Improve the underlying security of the UK internet through technological improvement and advice to citizens and organizations.

### Sample functions:

---



Develops knowledge and distills insight on cybersecurity into practical guidance for public consumption.

---



Responds to cybersecurity incidents to reduce the harm they cause to people and organizations.

---



Applies industry and academic expertise to build capability in the cybersecurity system.

---



Secures public- and private-sector networks.

---



Provides a single point of contact for government agencies, departments, and organizations of all sizes.

---



Collaborates with law-enforcement, defense, intelligence, and security agencies and international partners.

Source: National Cyber Security Centre, [ncsc.gov.uk](https://www.ncsc.gov.uk)

Bush administration, is still developing. It's not straightforward, particularly on that scale. I don't think anybody has a perfect answer.

**David Chinn:** I think the military is a very interesting subset of government because I don't think there was even one model in the military. Some countries are creating cybercommands. Others are building cyber in all of their commands. Others are concentrating in their intelligence services, and then combining those in different ways. And that's also changing over time.

**Simon London:** *It sounds like we're in an era of institutional innovation, in many ways—to some degree, institutional improvisation to try and figure out what models work in what context.*

**Robert Hannigan:** Absolutely. I think the military's a very good example, particularly outside the US. The US is ahead of anybody, I think, in developing cyberskills in the military at scale. On the broader point about civilian structures and civilian/military, I think the one thing that is probably key is that many of the questions are the same, starting with, "What does government actually want to achieve?" And not being overambitious in what government can achieve, and what's the appropriate role of government, is a good starting point. And trying to define what people expect from their government. Things like a single source of advice, incident response, protection of certain networks. I think that is a conversation that just about every government is having in different ways.

**David Chinn:** But I think there's a paradox here, because if you were to interview the chairman or chief executive of any large corporation and ask them what's their top three risks, cyber would be on that top three, for every single one. And for many of them, it would be number one. Yet, if we look at what governments are doing, this is the one area of national security, of crime prevention and prosecution of critical national infrastructure, that governments have, to a large extent, abdicated their responsibility. Great, some small steps. And sorry, I don't mean to be critical of what was a big small step. But exalting the private sector to do better feels like a very different role that government takes in almost every aspect of life that would feature for most people in their

top three risks. I think there's a lot more to do, but unfortunately we may have to wait for a genuine event—people talk of the cyber 9/11—to create a big change in focus, understanding, spending, and so on.

**Simon London:** *Let me just put that back to you. What should be done?*

**David Chinn:** What would your list be, Robert?

**Robert Hannigan:** Your criticism is very fair. I mean I think the government has moved from an absolutely sort of hands-off position to say, "Well, we'll look after our networks, but everybody else should get better." And sort of slightly hectoring them when they're not good enough. To saying, "Yes, there are things that we could do at national scale."

The problem, I suppose, at the risk of sort of making excuses, is that the nature of cyberspace, however defined, makes politicians feel quite impotent, because it cuts across jurisdictions. They can pass laws in their own parliament that really have zero effect. They can regulate their own companies, but not necessarily others. That is a real problem.

For cybercrime, for example, most of it is based in countries which are either endemically corrupt or unwilling to do anything about it for geopolitical reasons. What do you do about that? I mean there's a much bigger context here of international relations, and we are a million miles from getting any kind of international agreements on the security and safety of cyberspace.

**Simon London:** *David, you were the rousing voice of critique just now. What should be done?*

**David Chinn:** First, a sophisticated debate around the legislative and regulatory environment. The use of product liability has been very effective in other sectors for changing the game for the manufacturers. A robust thinking about product liabilities, extension to the technology arena, would frankly have quite a chastening effect on industry.

**Simon London:** *In other words, selling a product that has technology embedded that is deemed to be insecure could be breaking the law.*

# We have a whole new issue emerging with quantum computing, and people have not quite woken up, including the regulators, to the fact that current encryption will cease to be useful once quantum arrives.

**David Chinn:** Well, not necessarily breaking the law, but would expose you to civil action that could have severe financial consequences. Effectively, it would create a market mechanism for valuing more secure products. Second, there is room for some better and some more regulation. For example, if you want to sell anything to the UK government, you have to meet a minimum standard called Cyber Essentials. This is not the most sophisticated, but, as we've discussed, most of the attacks are not the most sophisticated attacks.

These kind of standards are very helpful because they're easily adopted by people for their own supply chains. I think a promulgation of standards, ideally with some degree of harmonization. And it's very interesting, in the US the national standards organization, NIST,<sup>2</sup> has created a number of models, which have got global acceptance. Once an authority puts it out there in a world where there's a lot of uncertainty, there's a lot of demand for good standards.

The traditional tools of government around legislation, regulation, standards setting, and so on could be used quite a lot more, without throttling innovation. Industry always says, "You're going to throttle innovation." What they mean is it's going to cost them more. But the cost to society of insecurity is high and is going to get higher.

**Simon London:** *One of my takeaways from this conversation, tell me if this is right or wrong, is*

*that there will be one or more significant tier-one, we might call them attacks, on critical national infrastructure. We're recording this in London, but it may not be based in the UK. But that will come. We know where it will come. And that will probably shift the debate into a higher gear. That probably will shift the international debate about what is to be done and, in some ways, get this taken more seriously, perhaps at government policy and regulatory level. Is that a correct takeaway?*

**Robert Hannigan:** I think for most people, most of what they would experience, and most companies, is still crime. So that's the volume, but everybody understandably gets excited about the catastrophic attack and that there is a range of possibilities for and the insurance industry worries a lot about systemic failure. So systemic failure of cloud providers, for example. Systemic failure of some major financial institutions, two or three of which would bring down the system or could bring down the system. So those are the kind of real tier one. But there may be some political tier-one problems and attacks that will have the kind of effect that David was talking about earlier, of panic and political pressure.

**Simon London:** *Trust.*

**Robert Hannigan:** Yes, either trust or an attack that leads to loss of life. It might not be massive loss of life, but it would put huge pressure, as terrorism does, on politicians to react.

---

<sup>2</sup> National Institute of Standards and Technology.

**Simon London:** *So what's that Churchill phrase, this is not the beginning of the end. This is the end of the beginning?*

**Robert Hannigan:** Well, I don't think it's even really the end of the beginning. I think we're still at very early stages of this technology. For most people, it's 15, 20 years old. Even if you look back to the ARPANET,<sup>3</sup> it's, what, 40, 50 years old? That's not long, and it's developing incredibly fast.

We are about to add a massive amount of new processing power, and therefore new data to the system, mostly through the Internet of Things. We have a whole new issue emerging with quantum computing, and people have not quite woken up, including the regulators, to the fact that current encryption will cease to be useful once quantum arrives.

We need now to be building in quantum-safe encryption standards, which are available through NIST and through others. But if we don't do that, everything, every company's records, every bit of financial data, every transaction is going to be readable from the moment that quantum computing really arrives at scale. It's a wonderful innovation, and it has obviously lots of possibilities on the other side of the equation, but it is one that we need to start thinking about in regulatory terms now.

**Simon London:** *All right. Well, I think that's all we have time for. Robert and David, thank you so much, and thanks, as always, to you, our listeners, for tuning in. To learn more about our work on cybersecurity, technology, and related matters, please go to [McKinsey.com](https://www.mckinsey.com).*

---

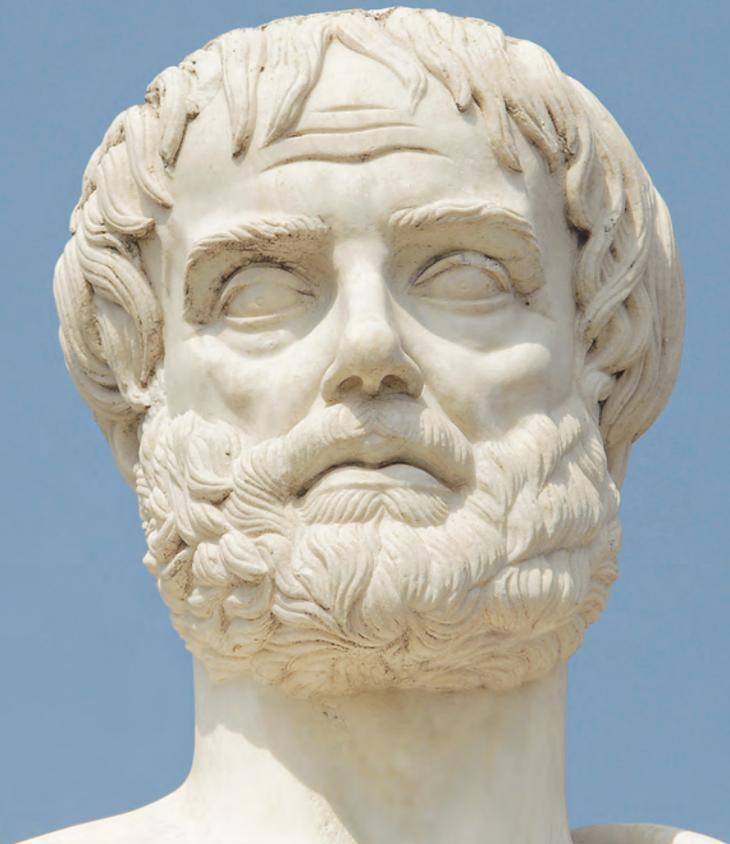
<sup>3</sup> Advanced Research Projects Agency Network.

**David Chinn** is a senior partner in McKinsey's London office, and **Robert Hannigan**, the former head of GCHQ, is a senior adviser to McKinsey. **Simon London**, a member of McKinsey Publishing, is based in McKinsey's Silicon Valley office.

Copyright © 2019 McKinsey & Company. All rights reserved.

# The ethics of artificial intelligence

Executives and companies can enjoy the benefits of artificial intelligence while also being aware of potential drawbacks and taking careful steps to mitigate their effects.



© Westend61/Getty Images

**In this episode of the *McKinsey Podcast*,** Simon London speaks with McKinsey Global Institute partner Michael Chui and partner Chris Wigley about the ethical implications of artificial intelligence (AI) and how executives can engage more thoughtfully on the use of AI and its potential repercussions.

**Simon London:** *Hello, and welcome to this edition of the McKinsey Podcast, with me, Simon London. Today we're going to be talking about the ethics of artificial intelligence. At the highest level, is it ethical to use AI to enable, say, mass surveillance or autonomous weapons? On the flip side, how can AI be used for good, to tackle pressing societal challenges? And in day-to-day business, how can companies deploy AI in ways that ensure fairness, transparency, and safety? To discuss these issues, I sat down with Michael Chui and Chris Wigley. Michael is a partner with the McKinsey Global Institute and has led multiple research projects on the impact of AI on business and society. Chris is both a McKinsey partner and chief operating officer at QuantumBlack, a London-based analytics company that uses AI extensively in his work with clients. Chris and Michael, welcome to the podcast.*

**Chris Wigley:** Great to be here.

**Michael Chui:** Terrific to join you.

**Simon London:** *This is a big, hairy topic. Why don't we start with the broadest of broad-brush questions which is, "Are we right to be concerned?" Is the ethics of AI something—whether you're a general manager or a member of the public—that we should be concerned about?*

**Chris Wigley:** Yes, I think the simple answer to this is that the concerns are justified. We are right to worry about the ethical implications of AI. Equally, I think we need to celebrate some of the benefits of AI. The high-level question is, "How do we get the balance right between those benefits and the risks that go along with them?"

On the benefit side, we can already see hundreds of millions, even billions of people using and benefiting from AI today. It's important we don't

forget that. Across all of their daily use in search and things like maps, health technology, assistants like Siri and Alexa, we're all benefiting a lot from the convenience and the enhanced decision-making powers that AI brings us.

But on the flip side, there are justifiable concerns around jobs that arise from automation of roles that AI enables, from topics like autonomous weapons, the impact that some AI-enabled spaces and forums can have on the democratic process, and even things emerging like deep fakes, which is video created via AI which looks and sounds like your president or a presidential candidate or a prime minister or some kind of public figure saying things that they have never said. All of those are risks we need to manage. But at the same time we need to think about how we can enable those benefits to come through.

**Michael Chui:** To add to what Chris was saying, you can think about ethics in two ways. One is this is an incredibly powerful tool. It's a general-purpose technology—people have called it—and one question is, "For what purposes do you want to use it?" Do you want to use it for good or for ill? There's a question about what the ethics of that are. But again, you can use this tool for doing good things, for improving people's health. You can also use it to hurt people in various ways. That's one level of questions.

I think there's a separate level of questions which are equally important. Once you've decided perhaps I'm going to use it for a good purpose, I'm going to try to improve people's health, the other ethical question is, "In the execution of trying to use it for good, are you also doing the right ethical things?"

Sometimes you could have unintended consequences. You can inadvertently introduce bias in various ways despite your intention to use it for good. You need to think about both levels of ethical questions.

**Simon London:** *Michael, I know you just completed some research into the use of AI for good. Give us an overview. What did you find when you looked at that?*

**Michael Chui:** One of the things that we were looking at was how could you direct this incredibly powerful set of tools to improving social good. We looked at 160 different individual potential cases of AI to improve social good, everything from improving healthcare and public health around the world to improving disaster recovery (Exhibit 1). Looking at the ability to improve financial inclusion, all of these things.

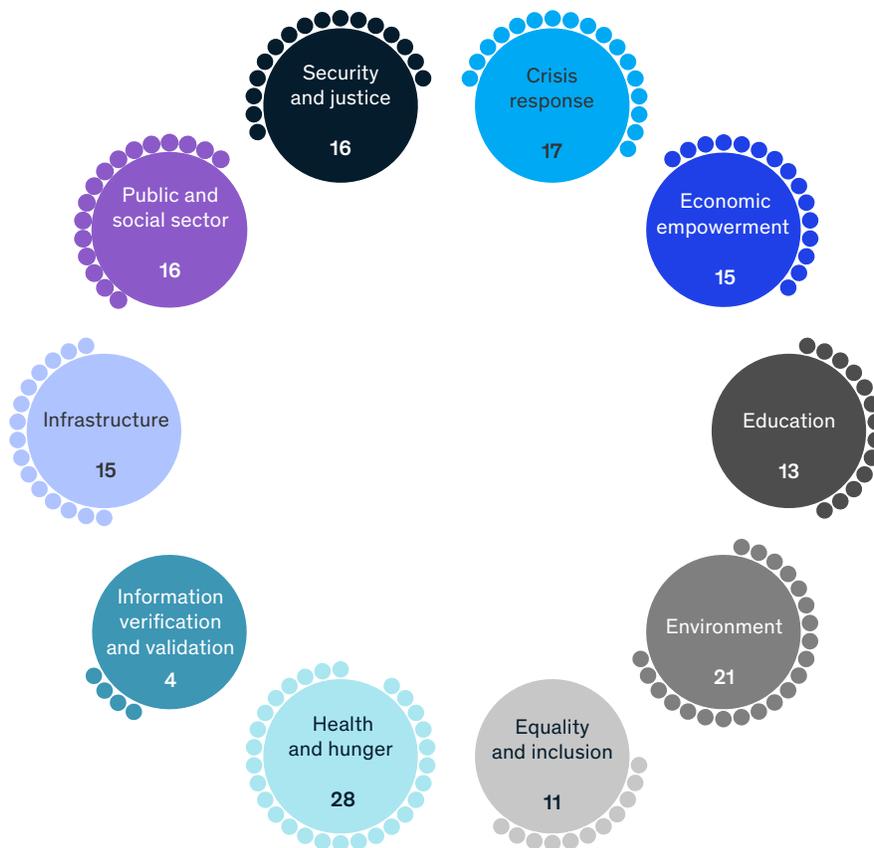
For pretty much every one of the UN's Sustainable Development Goals, there are a set of use cases where AI can actually help improve some of our progress towards reaching those Sustainable Development Goals.

**Simon London:** Give us some examples. What are a couple of things? Bring it to life.

Exhibit 1

## Artificial intelligence (AI) has broad potential across a range of social domains.

AI use cases per domain, number



Note: Our library of about 160 use cases is not comprehensive and will continue to evolve. This listing of the number of cases per domain should thus not be read as exhaustive.

Source: McKinsey Global Institute analysis

**Michael Chui:** Some of the things that AI is particularly good at—or the new generations of AI are particularly good at—are analyzing images, for instance. That has broad applicability. Take, for example, diagnosing skin cancer. One thing you could imagine doing is taking a mobile phone and uploading an image and training an AI system to say, “Is this likely to be skin cancer or not?” There aren’t dermatologists everywhere in the world where you might want to diagnose skin cancer. So being able to do that, and again, the technology is not perfect yet, but can we just improve our accessibility to healthcare through this technology?

On a very different scale, we have huge amounts of satellite imagery. The entire world’s land mass is imaged in some cases several times a day. In a disaster situation, it can be very difficult in the search for humans, to be able to identify which buildings are still there, which healthcare facilities are still intact, where are there passable roads, where aren’t there passable roads.

We’ve seen the ability to use artificial-intelligence technology, particularly deep learning, be able to very quickly, much more quickly than a smaller set of human beings, identify these features on satellite imagery, and then be able to divert or allocate resources, emergency resources, whether it’s healthcare workers, whether it’s infrastructure construction workers, to better allocate those resources more quickly in a disaster situation.

**Simon London:** *So disaster response, broadly speaking—there’s a whole set of cases around that.*

**Michael Chui:** Absolutely. It’s a place where speed is of the essence. When these automated machines using AI are able to accelerate our ability to deploy resources, it can be incredibly impactful.

**Chris Wigley:** One of the things that I find most exciting about this is linking that to our day-to-day work as well. So we’ve had a QuantumBlack team, for example, working with a city over the last few months recovering from a major gas explosion on

the outskirts of that city. That’s really helped to accelerate the recovery of that infrastructure for the city, helped the families who are affected by that, helped the infrastructure like schools and so on, using a mix of the kinds of imagery techniques that Michael’s spoken about.

Also there’s the commuting patterns—the communications data that you can aggregate to look at how people travel around the city and so on to optimize the work of those teams who are doing the disaster recovery.

We’ve also deployed these kinds of machine-learning techniques to look at things like, “What are the root causes of people getting addicted to opioids? And what might be some of the most effective treatments?” to things like the spread of disease in epidemiology, looking at the spread of diseases like measles in Croatia.

Those are all things that we’ve been a part of in the last 12 months, often on a pro bono basis, bringing these technologies to life to really solve concrete societal problems.

**Simon London:** *The other thing that strikes me in the research is that very often you are dealing with more vulnerable populations when you’re dealing with some of these societal-good issues. So yes, there are many ways in which you can point AI at these societal issues, but the risks in implementation are potentially higher because the people involved are in some sense vulnerable.*

**Michael Chui:** I think we find that to be the case. Sometimes AI can improve social good by identifying vulnerable populations. But in some cases that might hurt the people that you’re trying to help the most. Because when you’re identifying vulnerable populations, then sometimes bad things can happen to them, whether it’s discrimination or acts of malicious intent.

To that second level that we talked about before, how you actually implement AI within a specific use case also brings to mind a set of ethical questions

about how that should be done. That's as true in for-profit cases as it is for not-profit cases. That's as true in commercial cases as it is in AI for social good.

**Simon London:** *Let's dive deeper on those risks then, whether you're in a for-profit or a not-for-profit environment. What are the main risks and ethical issues related to the deployment, AI in action?*

**Chris Wigley:** One of the first we should touch on is around bias and fairness. We find it helpful to think about this in three levels, the first being bias itself. We might think about this where a data set that we're drawing on to build a model doesn't reflect the population that the model will be applied to or used for.

There have been various controversies around facial-recognition software not working as well for women, for people of color, because it's been trained on a biased data set which has too many white guys in it. There are various projects afoot to try and address that kind of issue. That's the first level, which is bias. Does the data set reflect the population that you're trying to model?

You then get into fairness which is a second level. Saying, "Look, even if the data set that we're drawing on to build this model accurately reflects history, what if that history was by its nature unfair?" An example domain here is around predictive policing. Even if the data set accurately reflects a historical reality or a population, are the decisions that we make on top of that fair?

Then the final one is [about whether the use of data is] unethical. Are there data sets and models that we could build and deploy which could just be turned to not just unfair but unethical ends? We've seen debates on this between often the very switched-on employees of some of the big tech firms and some of the work that those tech firms are looking at doing.

Different groups' definitions of unethical will be different. But thinking about it at those three levels of, one: bias. Does the data reflect the population?

Two: fairness. Even if it does, does that mean that we should continue that in perpetuity? And three: unethical. "Are there things that these technologies can do which we should just never do?" is a helpful way of separating some of those issues.

**Michael Chui:** I think Chris brings up a really important point. We often hear about this term algorithmic bias. That suggests that the software engineer embeds their latent biases or blatant biases into the rules of the computer program. While that is something to guard against, the more insidious and perhaps more common for this type of technology is the biases that might be latent within the data sets as Chris was mentioning.

Some of that comes about sometimes because it's the behavior of people who are biased and therefore you see it. Arrest records being biased against certain racial groups would be an example. Sometimes it just comes about because of the way that we've collected the data.

That type of subtlety is really important. It's not just about making sure that the software engineer isn't biased. You really need to understand the data deeply if you're going to understand whether there's bias there.

**Simon London:** *Yes, I think there's that famous example of potholes in Boston I think it was using the accelerometers in smart phones to identify when people are driving, do they go over potholes. The problem with that at the time that this data was collected is that a lot of the more disadvantaged populations didn't have smart phones. So there was more data on potholes in rich neighborhoods.<sup>1</sup>*

**Chris Wigley:** There's a bunch of other risks that we also need to take into account. If the bias and fairness gives us an ethical basis for thinking about this, we also face very practical challenges and risks in this technology. So, for example, at QuantumBlack, we do a lot of work in the pharmaceutical industry. We've worked on topics like patient safety in clinical trials. Once we're building these technologies into the workflows of

---

<sup>1</sup> The Street Bump program is not in active use by the city of Boston.

people who are making decisions in clinical trials about patient safety, we have to be really, really thoughtful about the resilience of those models in operation, how those models inform the decision-making of human beings but don't replace it, so we keep a human in the loop, how we ensure that the data sources that feed into that model continue to reflect the reality on the ground, and that those models get retrained over time and so on.

In those kinds of safety-critical or security-critical applications, this becomes absolutely essential. We might add to this areas like critical infrastructure, like electricity networks and smart grids, airplanes. There are all sorts of areas where there is a vital need to ensure the operational resilience of these kinds of technologies as well.

**Michael Chui:** This topic of the safety of AI is a very hot one right now, particularly as you're starting to see it applied in places like self-driving cars. You're seeing it in healthcare, where the potential impact on a person's safety is very large.

In some cases we have a history of understanding how to try to ensure higher levels of safety in those fields. Now we need to apply them to these AI technologies because many of the engineers in these fields don't understand that technology yet, although they're growing in that area. That's an important place to look in terms of the intersection of safety and AI.

**Chris Wigley:** And the way that some people have phrased that, which I like is, "What is the building code equivalent for AI?" I was renovating an apartment last year. The guy comes around from the local council and says, "Well, if you want to put a glass pane in here, because it's next to a kitchen, it has to be 45-minutes fire resistant." That's evolved through 150, 200 years of various governments trying to do the right thing and ensure that people are building buildings which are safe for human beings to inhabit and minimize things like fire risk.

We're still right at the beginning of that learning curve with AI. But it's really important that we start to shape out some of those building code equivalents for bias, for fairness, for explainability, for some of the other topics that we'll touch on.

**Simon London:** *Chris, you just mentioned explainability. Just riff on that a little bit more. What's the set of issues there?*

**Chris Wigley:** Historically some of the most advanced machine learning and deep-learning models have been what we might call a black box. We know what the inputs into them are. We know that they usefully solve an output question like a classification question. Here's an image of a banana or of a tree.

But we don't know what is happening on the inside of those models. When you get into highly regulated environments like the pharmaceutical industry and also the banking industry and others, understanding how those models are making those decisions, which features are most important, becomes very important.

To take an example from the banking industry, in the UK the banks have recently been fined over 30 billion pounds, and that's billion with a B for mis-selling of [payment] protection insurance. When we're talking to some of the banking leaders here, they say, "Well, you know, as far as we understand it, AI is very good at responding to incentives." We know that some of the historic problems were around sales teams that were given overly aggressive incentives. What if we incentivize the AI in the wrong way? How do we know what the AI is doing? How can we have that conversation with the regulator?

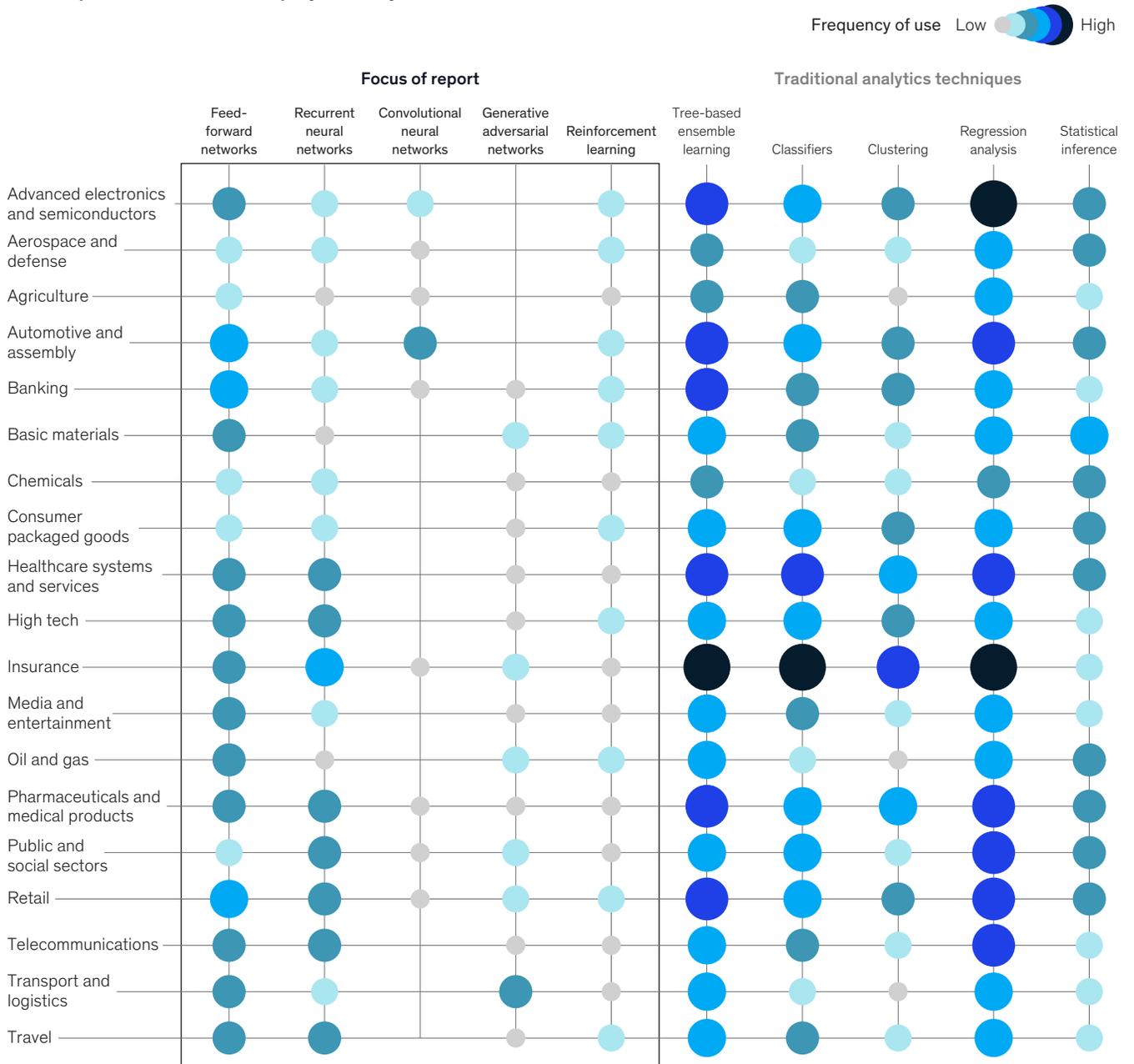
We've been doing a lot of work recently around, "How can we use AI to explain what AI is doing?" The way that that works in practice we've just done a test of this with a big bank in Europe in a safe area. This is how the relationship managers talk to their corporate clients. What are they talking to them about?

The first model is a deep-learning model, which we call a propensity model. What is the propensity of a customer to do something, to buy a product, to stop using the service? We then have a second machine-learning model, which is querying the first model millions of times to try and unearth why it's made that decision. (See Exhibit 2 for more on deep-learning models.)

Exhibit 2

**Advanced deep learning artificial intelligence techniques can be applied across industries, alongside more traditional analytics.**

Technique relevance<sup>1</sup> heatmap by industry



Note: List of techniques is not exhaustive.

<sup>1</sup> Relevance refers to frequency of use in our use case library, with the most frequently found cases marked as high relevance and the least frequently found as low relevance. Absence of circles indicates no or statistically insignificant number of use cases.

Source: McKinsey Global Institute analysis

It's deriving what the features are that are most important. Is it because of the size of the company? Is it because of the products they already hold? Is it because of any of hundreds of other features? We then have a third machine-learning model, which is then translating the insights of the second model back into plain English for human beings to understand. If I'm the relationship manager in that situation, I don't need to understand all of that complexity. But suddenly I get three or four bullet points written in plain English that say, "Not just here is the recommendation of what to do, but also here's why." It's likely because of the size of that company, of the length of the relationship we've had with that customer, whatever it is, that actually A) explains what's going on in the model and B) allows them to have a much richer conversation with their customer.

Just to close that loop, the relationship manager can then feed back into the model, "Yes, this was right. This was a useful conversation, or no, it wasn't." So we continue to learn. Using AI to explain AI starts to help us to deal with some of these issues around the lack of transparency that we've had historically.

**Michael Chui:** You could think about the ethical problem being, "What if we have a system that seems to work better than another one, but it's so complex that we can't explain why it works?" These deep-learning systems have millions of simulated neurons. Again trying to explain how that works is really, really difficult.

In some cases, as Chris was saying, the regulator requires you to explain what happened. Take, for example, the intersection with safety. If a self-driving car makes a left turn instead of hitting the brakes and it causes property damage or hurts somebody, a regulator might say, "Well, why did it do that?"

And it does call into question, "How do you provide a license?" In some cases what you want to do is examine the system and be able to understand and

somehow guarantee that the technical system is working well. Others have said, "You should just give a self-driving car a driving test and then figure out." Some of these questions are very real ones as we try to understand how to use and regulate these systems.

**Chris Wigley:** And there's a very interesting trade-off often between performance and transparency. Maybe at some point in the future there won't be a trade-off, but at the moment there is. So we might say for a bank that's thinking about giving someone a consumer loan, we could have a black-box model, which gets us a certain level of accuracy, let's say 96, 97 percent accuracy of prediction whether this person will repay. But we don't know why. And so therefore we struggle to explain either to that person or to a regulator why we have or haven't given that person a loan.

But there's maybe a different type of model which is more explainable which gets us to 92, 93 percent level of accuracy. We're prepared to trade off that performance in order to have the transparency.

If we put that in human terms, let's say we're going in for treatment. And there is a model that can accurately predict whether either a tumor is cancerous or another medical condition is right or wrong. To some extent, as a human being, if we're reassured that this model is right and has been proven to be right in thousands of cases, we actually don't care why it knows as long as it's making a good prediction that a surgeon can act on that will improve our health.

We're constantly trying to make these trade-offs between the situations where explainability is important and the situations where performance and accuracy are more important.

**Michael Chui:** Then for explainability it's partly an ethical question. Sometimes it has to do with just achieving the benefits. We've looked at some companies where they've made the trade-off that Chris suggested, where they've gone to a slightly

less performant system because they knew the explainability was important in order for people to accept the system and therefore actually start to use it.

Change management is one of the biggest problems in AI and other technologies to achieve benefits (Exhibit 3). And so explainability can make a difference. But as Chris also said, “That can change over time.” For instance, I use a car with [anti-lock] braking systems. And the truth is I don’t know how that works. And maybe earlier on in that history people were worried: “You’re going to let the car brake for itself.”

But now we’ve achieved a level of comfort because we’ve discovered this stuff works almost all the time. If we start to see that comfort change in an individual basis as well.

**Simon London:** *I’m going to ask an almost embarrassingly nerdy management question now. Stepping away from the technology, what’s*

*our advice to clients about how to address some of these issues? Because some of this feels like it’s around risk management. As you think about deploying AI, how do you manage these ethical risks, compliant risks, you could phrase it any number of different ways. What’s the generalizable advice?*

**Michael Chui:** Let me start with one piece of advice, which is as much as we expect executives to start to learn about every part of their business and maybe you’re going to be a general manager, you’re going to need to know something about supply chain, HR strategy, operations, sales and marketing. It is becoming incumbent on every executive to learn more about technology now.

To the extent to which they need to learn about AI, they’re going to need to learn more about what it means to deploy AI in an effective way. We can bring some of the historical practices—you mentioned risk management. Understanding risk is something that we’ve learned how to do in other fields.

Exhibit 3

## **Eighteen bottlenecks could limit artificial intelligence’s (AI’s) benefit to society.**

### **Four categories of limitations to AI use**

<b>Critical barriers for most domains</b>	<b>Critical barriers for select cases<sup>1</sup></b>	<b>Contextual challenges</b>	<b>Potential bottlenecks</b>
● Data accessibility	● Data volume	● Data availability	● Access to software libraries and other tools
● Data quality	● Data labeling	● Data integration	● Organizations able to scale AI deployment
● High-level AI-expertise availability	● AI-practitioner talent availability	● Access to technology	
● High-level AI-expertise accessibility	● AI-practitioner talent accessibility	● Privacy concerns	
● Regulatory limitations	● Access to computing capacity	● Organizational receptiveness	
● Organizational-deployment efficiency			

Note: This list of bottlenecks was derived from interviews with social-sector experts and AI researchers and tested against our use cases.

<sup>1</sup> Bottlenecks that are critical for some domains as a whole or for individual use cases within those domains.

Source: McKinsey Global Institute analysis

We can bring some of those tools to bear here when we couple that with the technical knowledge as well. One thing we know about risk management: understand what all the risks are. I think bringing that framework to the idea of AI and its ethics carries over pretty well.

**Simon London:** *Right. So it's not just understanding the technology, but it's also at a certain level understanding the ethics of the technology. At least get in your head what are the ethical or the regulatory or the risk implications of deploying the technology.*

**Michael Chui:** That's exactly right. Take, for example, bias. In many legal traditions around the world, understanding that there are a set of protected classes or a set of characteristics around which we don't want to actually use technology or other systems in order to discriminate.

That understanding allows you to say, "Okay, we need to test our AI system to make sure it's not creating disparate impact for these populations of people." That's a concept that we can take over. We might need to use other techniques in order to test our systems. But that's something we can bring over from our management practices previously.

**Chris Wigley:** As a leader thinking about how to manage the risks in this area, dedicating a bit of head space to thinking about it is a really important first step. The second element of this is bring someone in who really understands it. In 2015, so three years ago now, we hired someone into QuantumBlack who is our chief trust officer.

No one at the time really knew what that title meant. But we knew that we had to have someone who was thinking about this full time as their job because trust is existential to us. What is the equivalent if you're a leader leading an organization? What are the big questions for you in this area? How can you bring people into

the organization or dedicate someone in the organization who has that kind of mind-set or capabilities to really think about this full time?

**Michael Chui:** To build on that, I think you need to have the right leaders in place. As a leadership team, you need to understand this. But the other important thing is to cascade this through the rest of the organization, understanding that change management is important as well.

Take the initiatives people had to do in order to comply with GDPR. That's something that again I'm not saying that if you're GDPR compliant, you're ethical, but think about all the processes that you had to cascade not only for the leaders to understand but all of your people and your processes to make sure that they incorporate an understanding of GDPR.

I think the same thing is true in terms of AI and ethics as well. You think about everyone needs to understand a little bit about AI, and they have to understand, "How can we deploy this technology in a way that's ethical, in a way that's compliant with regulations?" That's true for the entire organization. It might start at the top, but it needs to cascade through the rest of the organization.

**Chris Wigley:** We also have to factor in the risk of not innovating in this space, the risk of not embracing these technologies, which is huge. I think there's this relationship between risk and innovation that is really important and a relationship between ethics and innovation. We need an ethical framework and an ethical set of practices that can enable innovation. If we get that relationship right, it should become a flywheel of positive impact where we have an ethical framework which enables us to innovate, which enables us to keep informing our ethical framework, which enables us to keep innovating. That positive momentum is the flip side of this. There's a risk of not doing this as much as there are many risks in how we do it.

**Simon London:** *Let's talk a little bit more about this issue of algorithmic bias, whether it's in the data set or actually in the system design. Again very practically, how do you guard against it?*

**Chris Wigley:** We really see the answer to the bias question as being one of diversity. We can think about that in four areas. One is diversity of background of the people on a team. There's this whole phenomenon around group think that people have blamed for all sorts of disasters. We see that as being very real. We have 61 different nationalities across QuantumBlack. We have as many or more academic backgrounds. Our youngest person is in their early 20s. Our oldest person in the company is in their late 60s. All of those elements of diversity of background come through very strongly.

We were at one point over 50 percent women in our technical roles. We've dropped a bit below that as we've scaled. But we're keen to get back. Diversity of people is one big area.

The second is diversity of data. We touched on this topic of bias in the data sets not reflecting the populations that the model is looking at. We can start to understand and address those issues of data bias through diversity of data sets, triangulating one data set against another, augmenting one data set with another, continuing

to add more and more different data perspectives onto the question that we're addressing.

The third element of diversity is diversity of modeling. We very rarely just build a single model to address a question or to capture an opportunity. We're almost always developing what we call ensemble models that might be a combination of different modeling techniques that complement each other and get us to an aggregate answer that is better than any of the individual models.

The final element of diversity we think about is diversity of mind-set. That can be diversity along dimensions like the Myers-Briggs Type Indicator or all of these other types of personality tests. But we also, as a leadership team, challenge ourselves in much simpler terms around diversity. We sometimes nominate who's going to play the Eeyore role and who's going to play the Tigger role when we're discussing a decision. Framing it even in those simple Winnie the Pooh terms can help us to bring that diversity into the conversation. Diversity of background, diversity of data, diversity of modeling techniques, and diversity of mind-sets. We find all of those massively important to counter bias.

**Michael Chui:** So adding to the diversity points that Chris made, there are some process things that are important to do as well. One thing you can

**We're almost always developing what we call ensemble models that might be a combination of different modeling techniques that complement each other.**

do as you start to validate the models that you've created is have them externally validated. Have someone else who has a different set of incentives check to make sure that in fact you've understood whether there's bias there and understood whether there's unintended bias there.

Some of the other things that you want to do is test the model either yourself or externally for specific types of bias. Depending on where you are, there might be classes of individuals or populations that you are not permitted to have disparate impact on. One of the important things to understand there is not only is race or sex or one of these protected characteristics.

**Simon London:** *And a protected characteristic is a very specific legal category, right? And it will vary by jurisdiction?*

**Michael Chui:** I'm not a lawyer. But, yes, depending on which jurisdiction you're in, in some cases, the law states, "You may not discriminate or have disparate impact against certain people with a certain characteristic." In order to ensure that you're not discriminating or having disparate impact is not only that you don't have gender as one of the fields in your database.

Because sometimes what happens is you have these, to get geeky, these co-correlates, these other things which are highly correlated

with an indicator of a protected class. And so understanding that and being able to test for disparate impact is a core competency to make sure that you're managing for biases.

**Chris Wigley:** One of the big issues, once the model is up and running, is, "How can we ensure that while we've tested it as it's being developed, that it maintains in operation both accuracy and not being biased." We're in the reasonably early stages of this as an industry on ensuring resilience and ethical performance in production.

But some simple steps like, for example, having a process check to say, "When was the last time that this model was validated?" It sounds super simple. If you don't do that, people have very busy lives, and they can just get overlooked. Building in those simple process steps all the way through to the more complicated technology-driven elements of this.

We can actually have a second model checking the first model to see if it's suffering from model drift, for example. And then translate that into a very simple kind of red, amber, green dashboard of a model in performance. But a lot of this still relies on having switched-on human beings who maybe get alerted or helped by technology, but who engage their brain on the topic of, "Are these models, once they're up and running, actually still performant?"

**You may not discriminate or have disparate impact against certain people with a certain characteristic.**

All sorts of things can trip them up. A data source gets combined upstream and suddenly the data feed that's coming into the model is different from how it used to be. The underlying population in a given area may change as people move around. The technologies themselves change very rapidly. And so that question of how do we create resilient AI, which is stable and robust in production, is absolutely critical, particularly as we introduce AI into more and more critical safety and security and infrastructure systems.

**Michael Chui:** And the need to update models is a more general problem than just making sure that you don't have bias. It's made even more interesting when there are adversarial cases. When in fact just to say, for instance, you have a system that's designed to detect fraud. People who are fraudulent obviously, don't want to get detected. So they might change their behavior understanding that the model is starting to detect certain things.

And so again, you really need to understand when you need to update the model whether it's to make sure that you're not introducing bias or just in general to make sure that it's performing.

**Chris Wigley:** There's an interesting situation in the UK where the UK government has set up a new independent body called the Centre for Data Ethics and Innovation that is really working on balancing these things out. How can you maximize the benefits of AI to society within an ethical framework?

And the Centre for Data Ethics and Innovation, or CDEI, is not itself a regulatory body but is advising the various regulatory bodies in the UK like the FCA, which regulates the financial industry and so on. I suspect we'll start to see more and more thinking at a government and inter-government level on these topics. It'll be a very interesting area over the next couple of years.

**Simon London:** *So AI policy broadly speaking is coming into focus and coming to the fore and becoming much more important over time.*

**Michael Chui:** It is indeed becoming more important. But I also think that it's interesting within individual regulatory jurisdictions, whether it's in healthcare or in aviation, whether it's what happens on roads, the degree to which our existing practices can be brought to bear.

So again as I said, are driving tests the way that we'll be able to tell whether autonomous vehicles should be allowed on the roads? There are things around medical licensure and how is that implicated in terms of the AI systems that we might want to bring to bear. Understanding that tradition and seeing what can be applied to AI already is really important.

**Simon London:** *So what is the standard to which we hold AI? And how does that compare to the standard to which we hold humans?*

**Michael Chui:** Indeed.

**Chris Wigley:** Absolutely. In the context of something like autonomous vehicles, that's a really interesting question. Because we know that a human population of a certain size that drives a certain amount is likely to have a certain number of accidents a year. Is the right level for allowing autonomous vehicles when it's better than that level or when it's better than that level by a factor of ten?

Or do we only allow it when we get to a perfect level? And is that ever possible? I don't think that anyone knows the answer to that question at the moment. But I think that as we start to flesh out these kinds of ethics frameworks around machine learning and AI and so on, we need to deploy them to answer questions like that in a way which various stakeholders in society really buy into.

A lot of the answers to fleshing out these ethical questions have to come from engaging with stakeholder groups and engaging with society more broadly, which is in and of itself an entire process and entire skill set that we need more of as we do more AI policy making.

**Simon London:** *Well, thank you, Chris. And thank you, Michael, for a fascinating discussion.*

**Michael Chui:** Thank you.

**Chris Wigley:** It's been great.

**Michael Chui** is a partner of the McKinsey Global Institute and is based in McKinsey's San Francisco office. **Chris Wigley** is a partner in the London office. **Simon London**, a member of McKinsey Publishing, is based in McKinsey's Silicon Valley office.

The authors wish to thank Rhea Naidoo and Rolf Riemenschnitter for their contributions to this article.

Copyright © 2019 McKinsey & Company. All rights reserved.





Copyright © McKinsey & Company. All rights reserved.

[www.mckinsey.com](http://www.mckinsey.com)