

Artificial intelligence (AI) has the potential to transform business models – a new industrial revolution. But without assurance over the strategy, how it is implemented and the subsequent outputs and outcomes, boards are wary of giving the green light to AI. And if they do go ahead, they risk operating in the dark. So how can your business gain the control it needs to unleash the full potential of AI?

Accelerating innovation

How to build trust and confidence in AI



Responsible artificial intelligence study 2017

Introduction

Unleashing the potential

Artificial intelligence needs both more robust governance and a new operating model to realise its full potential.

AI is emerging as the defining technology of our age, with many industries already utilising AI in some form. And as humans and machines collaborate more closely, and innovations come out of the research lab and into the mainstream, AI offers transformational possibilities for consumers, businesses and society as a whole.

More than 60% of the 2,500 consumers and business decision makers in the US, who took part in a PwC survey on attitudes to AI, believe that it can help provide solutions for many of the most important issues facing modern society, ranging from clean energy to cancer and disease¹.

At the heart of the opportunity for businesses is the ability to turn data into intellectual property (IP) – more than 70% of business leaders in our survey believe that AI will be the business advantage of the future. We're currently analysing how AI could enhance the quality, personalisation and value of hundreds of different products and services across eight prominent sectors – we'll be publishing the overall and sector-specific findings over the course of 2017.

^{1,2} 'Bot me: A revolutionary partnership: How AI is pushing man and machine closer together', PwC, April 2017 (<https://www.pwc.com/us/en/press-releases/assets/img/bot-me.pdf>)

Uncertainty holds back innovation

The burning question is how to realise this potential. Netflix offers a textbook example of how a business can use accumulations of data to create a virtuous circle of ‘machine learning advantage’. The disruptive results have changed the way we access media content and led to a complete rethink of business models in the entertainment sector. In the global economy as whole, however, progress on AI has been mixed. Many organisations are still at the beginning of the journey – less than 40% of the business leaders taking part in our latest Global CEO Survey have begun to explore the impact of AI on their future skills needs³, for example.

Other organisations risk finding themselves strategically hamstrung. In part, this stems from the difficulties of choosing from such a bewildering array of technologies, innovations and vendors. Many others are finding it difficult to evaluate, mitigate and manage the ‘black box’ reputational as well as technological risks of using such new and largely untried technology. Key challenges include determining whether the data is valid and what safeguards are needed to ensure that machines carry out human orders as intended. Related ethical considerations range from is it acceptable to influence human choices to do consumers understand enough about how their data is used and who has access to it? While some companies are keen to press ahead with AI, they can often find themselves chasing too many opportunities or failing to adequately assess the full business case and associated risks – the current landscape is littered with too many abandoned proof of concepts.

72%

of business leaders in the US believe AI will be the business advantage of the future²

67%

of CEOs think that AI and automation (including blockchain) will have a negative impact on stakeholder trust in their industry over the next five years⁴

Robust evaluation and execution

These challenges underline the need for a new model of strategic evaluation, governance and delivery – without it, the uncertainties surrounding AI mean that it will either remain stuck in the lab within many organisations or they will find themselves facing unacceptable and potentially damaging risks.

At the heart of this framework is the need for trust and transparency in creating responsible AI. The adoption of AI may be met with scepticism from a variety of stakeholders, both within the organisation and clients, regulators and others outside. It’s therefore important to consider how we can build trust among all the affected stakeholders. The key to this is increasing transparency and awareness around how AI is being used, the jobs it performs, the decisions it makes and the opportunities it brings – we believe this is the essence of ‘responsible AI’.

While no one can guarantee good behaviours from complex autonomous agents, there’s a series of best practices that include designing and monitoring controls, which would minimise risk and encourage responsible adoption of AI.

In this paper, we explore the challenges that such a framework would need to address, the opportunities it would open up and set out how it might work. The objective isn’t to stifle or slow down innovation, but rather to accelerate it by giving boards the assurance and platform for execution they need to deliver the desired outcomes.

^{3,4} ‘20 years inside the mind of a CEO...What’s next?’, 20th Global CEO Survey, PwC, 2017 (ceosurvey.pwc)

Harnessing disruption

How AI is changing the rules of the game

The adoption of AI has profound implications for everyone engaged in business management.

The emergence of AI is paving the way for a whole new set of operating and business models. The ability to analyse levels of data that are beyond human comprehension and act on each new set of information allow businesses to personalise experiences, customise products and services, and identify growth opportunities with a speed and precision that's never been possible before.

What is artificial intelligence?

In their book, 'Artificial Intelligence: A Modern Approach', Stuart Russell and Peter Norvig define AI as “*the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment*”.⁵ The most critical difference between AI and general-purpose software is in the phrase “**take action**”. AI enables machines to respond on their own to signals from the world at large, signals that programmers do not directly control and therefore can't anticipate. The fastest-growing category of AI is machine learning, the ability of software to improve its own activity, based on interaction with the world at large.

The spectrum of AI can be divided into three:

- **Assisted Intelligence**, widely available today, improves what people and organisations are already doing.
- **Augmented Intelligence**, emerging today, enables people to do things they couldn't otherwise do.
- **Autonomous Intelligence**, being developed for the future, establishes machines that act on their own.

Some of the ways AI is making its mark

- **Keeping us informed:** 'Personal assistants' such as Alexa and Siri, as well as banking and mobile phone network operator chatbots.
- **Predicting behaviour:** The UK National Health Service are piloting machine learning to predict outpatient non-attendance at a UK hospital, optimising scheduling and rescheduling of appointments while understanding the drivers that can influence patient behaviour and enable actions to be taken to reduce rates of non-attendance.
- **Keeping us well:** AI is being used to aid medical diagnosis – our research shows that a significant proportion of people worldwide are willing to choose certain treatments, tests or services administered by an AI or robot⁶.
- **Keeping us engaged:** Telecoms and media companies have been using machine learning customer analytics to predict and then recommend actions to prevent customer turnover.
- **Anticipating demand:** Retailers are beginning to use deep learning to predict customers' orders a week in advance.

⁵ 'Artificial Intelligence: A Modern Approach', Stuart Russell and Peter Norvig (Pearson, 2009)

⁶ 'What doctor: Why AI and robotics will define the New Health' (<https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/survey-results.html>)

- **Customisation for all:** Robo-advice has made it possible to offer customised investment solutions to a wider range of consumers. Until recently, this level of investment advice was only available to high net worth (HNW) clients.
- **Improving quality:** Manufacturers are using AI to improve quality control, reduce production line downtime and increase the speed and yield of industrial processes.
- **Intelligent processes:** Intelligent process automation is driving huge savings in finance, HR and compliance. Robotic process automation is combined with AI to perform high volume, routine tasks.

Navigating the sheer breadth of algorithms and applications that fall under the banner of AI has become a formidable task in its own right. To date, a lot of the focus has been on automation of tasks that are already carried out⁷. Yet as workers are freed from routine tasks and human and machines begin to collaborate more closely, the real breakthroughs will come from the ability to make more insightful decisions and the emergence of completely new augmented intelligence-led business models. Entertainment is a clear example of a sector that has already undergone significant disruption and change. Driverless cars are one of the many ways that AI is set to transform everyday lives and the businesses that support this.

Commercially-applied AI has expanded in recent years, driven by a combination of computing power, the availability of huge datasets and advances in machine learning (which includes deep learning). While often used for predictive analytics, as well as image and speech classification, machine learning can be combined with elements of ‘traditional’ AI such as natural language processing, strategic planning and logical reasoning to deliver powerful autonomous agents.





So how prevalent is AI? Outside of large tech companies that have been utilising AI in service delivery for a number of years, much of the innovation is still in its infancy and is largely confined to the lab in the form of proof of concepts or R&D. The focus for business now has to be on creating an environment which fosters successful transition into real world value delivery.

⁷ We explore the impact of automation and AI on production and employment in ‘Will robots steal our jobs? The potential impact of automation on the UK and other major economies, March 2017’ (<https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf>)

New approach

As Figure 1 highlights, the adoption of AI demands a new way of thinking about technology, business development and strategic execution, along with the reshaped operating model and decision making processes that underpin this. And this affects the entire business, rather than just technology and innovation teams.

Figure 1: Rethinking the way you do business

	Traditional approach	New approach
Strategy 	Technology for information management Data as business intelligence Deterministic approach	Technology that manages your business Data as your differentiating intellectual property Directional (iterative) approach
Design 	User experience as an application layer Decision making hard coded Information retrieval as fact from database	User experience as the primary application feature Decision process learnt by software Information retrieval most probable correct answer
Development 	Linear technology development Business management teams specify, technology team builds	Iterative technology and business model development Business subject matter experts integrated into technology development teams
Operating Model 	Steady state technology, punctuated with upgrades Technical risks dominated by system downtime and errors Cyber attacks	Dynamic, adaptive models. Continuous test driven development Technical risks include learned and unexpected behaviour Adversarial attacks

Source: PwC



So what are the key considerations for bringing AI into the centre of your business and operating model?

Strategy

1. Aligning with your strategic goals

It's vital to align AI innovation with core strategic objectives and performance indicators, rather than allowing a scattered series of initiatives to operate in isolation. In our experience, a lot of organisations have set various pilots in train. What most aren't doing is taking a fundamental look at how AI could disrupt their particular business and then determining the threats and opportunities this presents.

2. Don't expect magic

AI may be intelligent, but it's still a machine. A common problem is believing the AI will magically learn without human intervention. In reality, you have to put a lot of effort into acquiring and cleansing data, labelling and training both machines and employees⁸.

3. Clear about your partners

Everywhere you look, there are start-ups offering solutions to this and opportunities for that. Partnership with these vendors accelerates innovation, agility and speed to market. But it's clearly important to pick your spot. This includes being clear about the strategic and operational priorities you're looking to address through the choice of partner. It's also important to bear in mind that while vendors may be good at selling the possibilities, they're not always as clear about how to deliver them – the way they look at development risks is certainly very different from what you're used to.

AI may be intelligent, but it's still a machine. A common problem is believing the AI will magically learn without human intervention.

In a high risk and fast-moving vendor landscape, the first consideration is the financial viability of the potential partners – will they still be there when you need them? It's also important to determine how to acquire the necessary data, develop the knowledge needed to deploy your new capabilities and how to integrate new platforms into existing infrastructure. When buying commercially available off-the-shelf software, a proof of concept development phase is often necessary.

4. Opening up to scrutiny

Before you adopt AI, you clearly need to know what it's doing and how. This includes ensuring the software can communicate its decision making process in a way that can be understood and scrutinised by business teams. In particular relation to machine learning, it's important to think about how to ensure the software will deliver the anticipated results. Boards want this assurance before they proceed. Regulators are also likely to expect it. Algorithmic transparency is part of the solution, though this may require a

trade-off between decision making transparency, system performance and functional capabilities.

5. Demonstrating regulatory compliance

Regulators need to move quickly to keep pace with emerging technologies. We may see regulatory constraints that prevent adoption in key regulated industries such as health and financial services. Developments such as the EU's General Data Protection Regulation (GDPR) are heightening the challenges. Staying compliant with relevant regulatory requirements is essential to build trust in your AI platform.

6. Organisational structure

The changes in your business models as part of your overall AI strategy will also need to be reflected in your organisational structure. Your organisation needs a dedicated AI governance structure, this could include a nominated member of the C-suite and a central hub of technical expertise. Embedding data scientists throughout your business either through training or hiring is essential to achieve AI organisational maturity.

⁸ PwC's Dr Anand S Rao explores the 'Five myths and facts about artificial intelligence' in Predictive Analytics and Futurism, Society of Actuaries, December 2016

Design

1. Opening up the black box

AI applications can communicate with customers and make important business decisions. But a lot of this is carried out within a black box, with the lack of transparency creating inherent reputational and financial risks. It's important to ensure that the software is designed in a way that is as transparent and auditable as possible.

Proper governance and protection include the ability to monitor component systems. It would also include the ability to quickly detect, correct and, if not, shut down 'rogue' components without having to take down whole platforms. Related priorities include identifying dependencies and being able to make modifications with minimal disruption if regulations or some other aspect of the operating environment changes.

2. Creating a compelling user experience

Many AI applications deploy highly subjective user experience performance metrics akin to IQ, personality, and predictability. Even though the bulk of development may focus on the analytics, the success of the product will be determined by an emotional response. This subjectivity means that frequent feedback is required between product owners and developers to make sure evolving expectations and functionality are properly managed. Often it makes sense to bring in specialist user interface vendors or use your in-house digital team alongside the core analytics team.

AI may excel and often surpass humans at particular tasks or in certain subject domains, but is generally incapable of extending these skills or knowledge to other problems. This is not obvious to people who have to interact with AI, especially for the first time, and can cause frustration and confusion.

The most effective controls are built into the design and implementation phase, enabling you to catch issues before they become a problem and also identify opportunities for improvement.

Branding and persona development ('functionality framing') are therefore key design considerations. Get it right and very basic software can appear human. Get it wrong and users will give up.

Some of the analysis performed by AI will inevitably be probabilistic based on incomplete information. It's therefore important that you recognise the limitations and explain this to customers. Examples might include how you present recommendations on investments from robo-advisors.

3. Embedding the control framework

The most effective controls are built during the design and implementation phase, enabling you to catch issues before they become a problem and also identify opportunities for improvement.

An important question is who designs and monitors the controls? Both the breadth of application and the need to monitor outcomes requires engagement from across the organisation. Control design requires significant input from business domain experts. Specialist safety engineering input is likely to be required for physical applications.

A key part of implementation is breaking the controls down into layers ('hierarchical approach').

At a minimum, there would be a hard control layer setting out 'red lines' and what to do if they're breached. Examples might include a maximum transaction value for a financial market trading algorithm. In more complex applications such as conversational agents, you could introduce a 'behaviour inhibitor' that overrides the core algorithm when there is a risk of errors such as regulatory violation or inappropriate language.

These core controls can be augmented by 'challenger models', which are used as a baseline to monitor the fitness and accuracy of the AI techniques or look for unwanted bias or deviations as the models learn from new data. Moreover, this approach can be integrated with continuous development to improve existing models or identify superior models for system upgrades.

Development

1. Rethinking programme management

Applying conventional planning, design and building to such data-dependent developments is destined to fail. Innovating and proving the concept through iterative development is needed to handle the complexity of the problems encountered and requires a high level of engagement from the product owners.

2. Managing data dependency

AI functionality is heavily data-dependent for machine learning model training and is likely to need a store of information known as a 'knowledge base'. This often means initial design specifications and expectations are set beyond the limits of what can be supported by the data, no matter how 'intelligent' the software. A key requirement of data dependent projects is a discovery phase to outline data quantity, quality and the limits this places on the resulting models and functionality. This is one reason why AI software implementations require significant design iteration during the development phase.

3. Taking the time to test and train

For machine learning in particular, it's important for the development team to apply best-practices to tuning and cross-validation methods to avoid overfitting and other common problems. To get a clear picture of the use case and user experience, programmes should bring in input from beyond the software design team, who may inevitably be too close to the project to look at it objectively enough. The monitoring should include testing to correct for functional blind spots.

One way to augment testing and cut down on the risks is to pilot new AI-based applications on a small scale first and encourage a thorough review by analysts and non-technical users in a 'business-as-usual' context. Expert judgement and additional contextual information allow further validation, impact assessment and tuning before launching AI initiatives on a larger scale.

In many ways, AI software development can be more akin to video game development than automation or web development, especially when it interacts directly with humans. Testing should reflect this through intense user experience evaluation and phased 'beta testing' on unseen audiences prior to release.

Finally, AI development may require a number of attempts to get it right. It's therefore important to ensure you have implemented the right level of programme assurance and quality controls, which can provide early indication of when data, technology or model training methods are not sufficient to support the business case.

4. Setting confidence thresholds

A balance between automation and human validation/verification is crucial. Defining the right thresholds and confidence levels at which to trigger human intervention can be challenging, however. Too cautious and the AI provides limited value. Too relaxed and the AI assumes more risk than can be contained. Continual monitoring of business performance is essential to confirm the technology is operating within the expected parameters.

Conversational AI agents engage in subjective communication with humans, so it's important to ensure the confidence thresholds are set up to conform to social norms and user expectations.

To get a clear picture of the use case and user experience it's important to bring in input from beyond the software design team...

Operating AI

1. Curbing unintentional bias

As more information becomes available and your model matures, it's important to guard against unintended bias against particular groups. Transparency is vital to be able to spot these biases. For systems that learn through customer interactions, periodic functional monitoring, perhaps based on a set of standardised interactions, is recommended to catch any adverse 'training drift'.

2. Guarding against attacks

Machine learning (especially deep learning) models can be duped by malicious inputs known as 'adversarial attacks'. It is possible to find input data combinations that can trigger perverse outputs from machine learning models, in effect 'hacking' them. This can be mitigated by simulating adversarial attacks on your own models and retraining models to recognise such attacks. Specialist software can be developed to 'immunise' your models against such attacks. This should be considered in the design phase.

Machine learning (especially deep learning) models can be duped by malicious inputs known as 'adversarial attacks'.

3. Recognising the role of data as your key intellectual property

AI is only as effective as the data it learns from. Maintaining high quality data and continuously evaluating the effectiveness of the model will be key to a successful AI platform. As data and technology applications move on to the cloud, commercial advantage will be driven by the magnitude and scale of the 'IP' you hold.

Partnership with a vendor may inevitably involve data exchange – i.e. intentionally or unintentionally passing on valuable IP. It's therefore important to understand the value of the data you're sharing and closely monitor and manage its supply and use.

4. Looking out for systemic risks

The flash crash that hit financial markets in 2010 demonstrates what can happen when AI interact happen when multiple AIs interacts in unintended ways and this isn't sufficiently monitored. Safeguards should include scenario planning, understanding of your own vulnerabilities and how to respond quickly.

In control of your AI

AI should be managed with the same discipline as any other technology enabled transformation. How can your business ensure it's in control?

So what are the starting points for assured and controlled, responsible AI? While this is an open-ended process, we believe that there are five key foundations that need to be in place before you move forward:

1. Ensure clarity over AI strategy

While there is a constant need to evaluate and adapt, it's still vital to be clear about your direction of travel.

- Are you ready for the disruption from AI?
- Have you considered the societal and ethical implications?
- What business outcomes are you looking to achieve (e.g. product customisation or back office efficiency)?

2. Transparency by design

Adoption of AI will be an emotive subject both within your organisation, with your customers and in society as well, so it's important to consider how you will build stakeholder trust in the solution.

It's important to build the controls framework into the solution up-front rather than being designed and applied once systems are developed and in operation. This includes a mechanism to monitor outcomes and compliance.

3. Build your AI organisation in advance

There are many possible models for developing an organisation-wide AI capability ranging from a centre of excellence and a dedicated board member to a 'develop-and-steer'

strategy. Whatever approach you adopt, it's important to ensure cross-organisational communication, collaboration and centralised co-ordination of AI initiatives.

4. Build data management into AI

If data is the new IP, it's important to put in place mechanisms to source, cleanse and control key data inputs and ensure data and AI management are integrated.

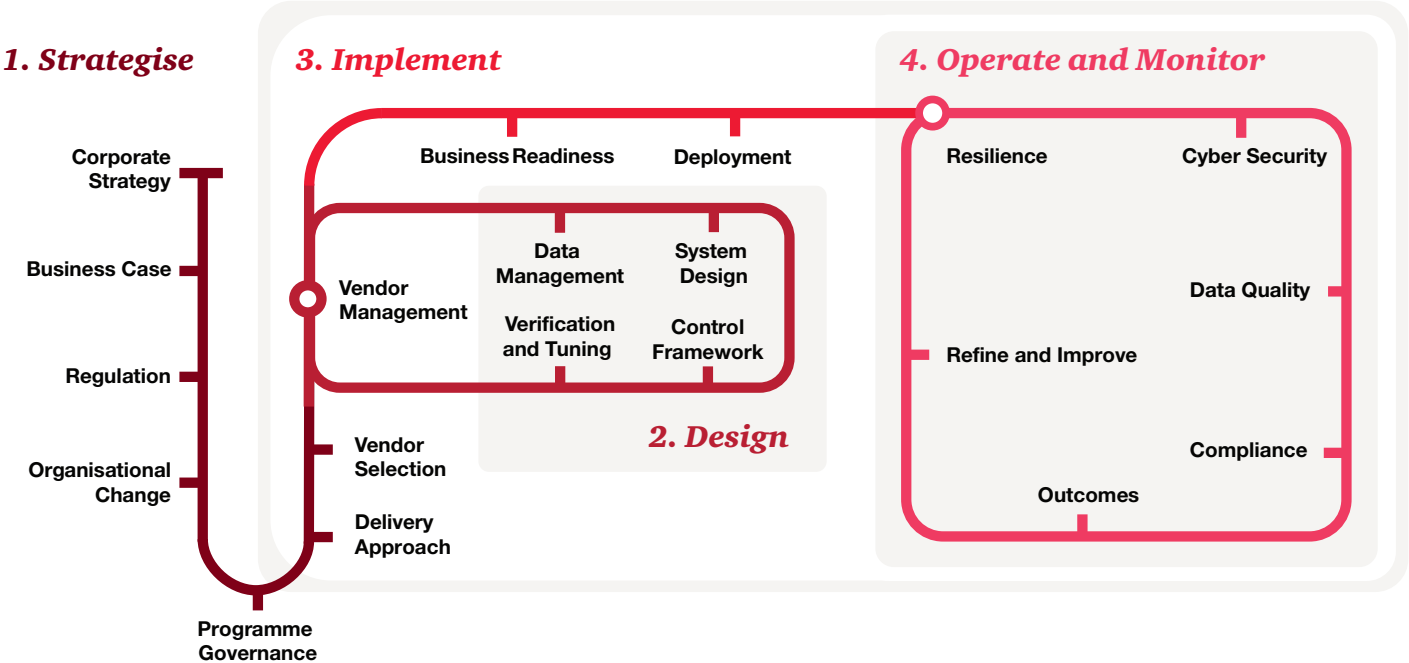
5. Integrate Assurance into your AI operating model

Assurance over AI isn't a one-off. You should assess the risks and opportunities as your AI platforms evolve.

Assurance over AI also involves more than just embedding new technology into operational processes. It requires business-wide evaluation to gauge outcomes, identify emerging risks and look out for opportunities.

Drawing on our wide-ranging research and work with clients, our responsible AI framework (Figure 2) is designed to provide transparency over the viability of the AI implementation project and confidence that the controls are in place to ensure that the business outcomes meet expectations.

Figure 2: The PwC Responsible AI Framework



Source: PwC

Conclusion

The way forward

If information is power, then AI is its zenith. Yet like all power it needs to be applied with insight, sensitivity and responsibility.

The opportunity is that AI will be a force for good, empowering people to achieve more and helping to resolve many of the problems faced in today's world. The big risk is that AI is allowed to operate beyond the boundaries of reasonable control. And for you as a business, this is not only ethically and reputationally unacceptable, but will also cause boards to question, delay and even shelve innovations.

Within PwC, we recognise the importance of assurance and control in unleashing the potential of AI. We've developed our responsible AI framework to strengthen clients' confidence in how to effectively deploy AI solutions and have trust in their outputs. The driving force isn't a suspicion of innovation, but rather a zeal for it. We want consumers, businesses and civil society to capitalise on the opportunities by understanding the limitations, vulnerabilities and potential pitfalls of AI, and applying informed governance to their mitigation and management.

Our responsible AI framework provides a practical mechanism for bringing these priorities together and ensuring effective monitoring and stewardship of AI outcomes. We believe that these foundations will enable your business to accelerate innovation and realise your AI vision.

Want to find out more then contact us

Authors



Chris Oxborough

Partner – Technology Risk
(Emerging and Disruptive Technology)

E: chris.oxborough@pwc.com



Euan Cameron

Partner – UK AI Leader

E: euan.cameron@pwc.com



Laurence Egglesfield

Director – Technology Risk
(Emerging and Disruptive Technology)

E: laurence.j.egglesfield@pwc.com



Aldous Birchall

Senior Manager – AI & Machine
Learning Lead (Financial Services)

E: aldous.birchall@pwc.com

Key subject matter experts

Dr. Anand Rao

Partner – Innovation Lead,
Data & Analytics

E: anand.s.rao@pwc.com

Rob McCargow

Senior Manager – Programme Leader,
Artificial Intelligence

E: rob.mccargow@pwc.com

Alan Morrison

Senior Manager – Sr. Research Fellow,
Emerging Technology

E: alan.s.morrison@pwc.com

Andy Townsend

Associate – Technology Risk
(Emerging and Disruptive Technology)

E: andrew.townsend@pwc.com

This publication has been prepared for general guidance on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, PwC does not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.

© 2017 PwC. All rights reserved. “PwC” refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

170503-110931-AT-OS