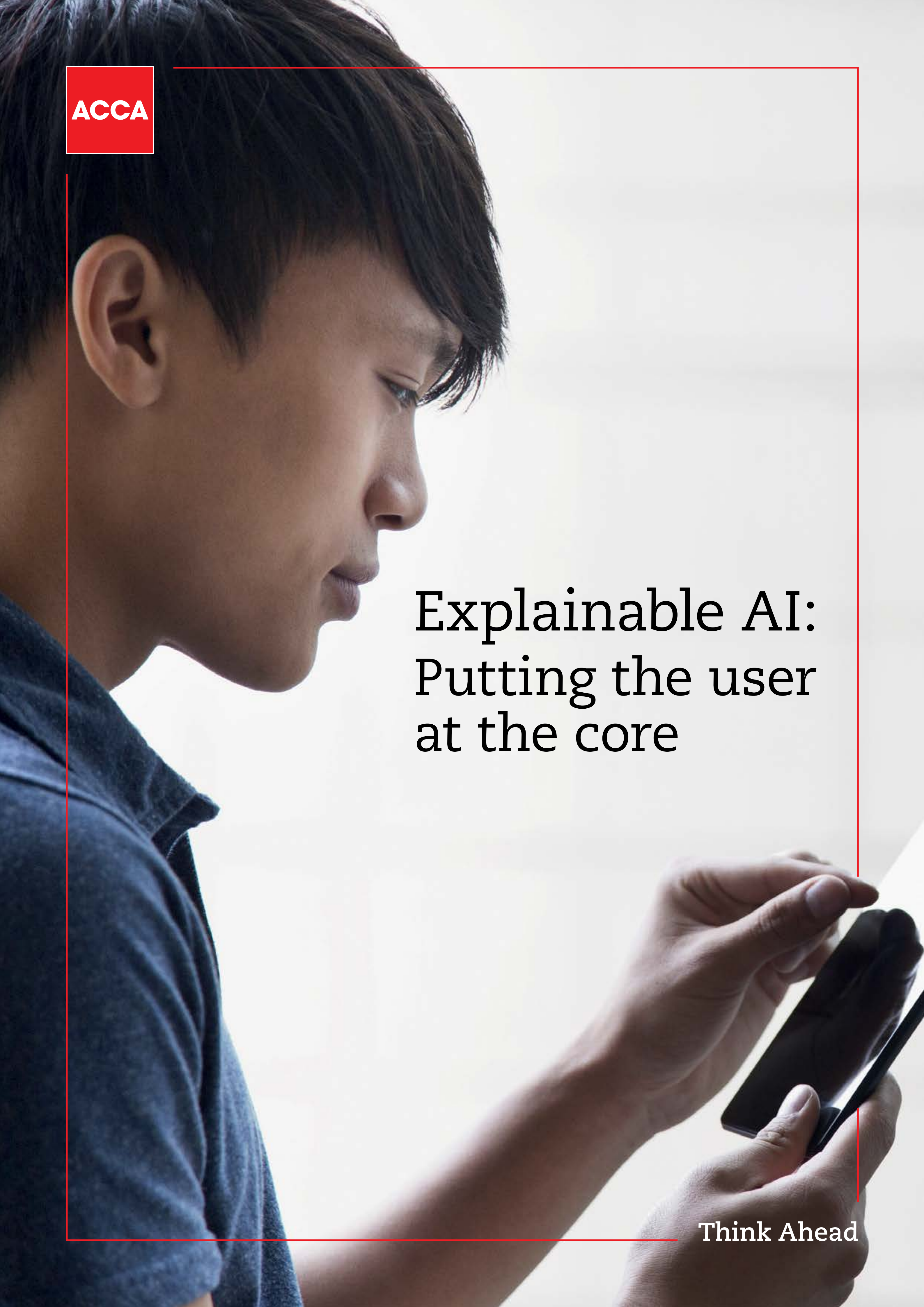


Explainable AI: Putting the user at the core

About ACCA

ACCA (the Association of Chartered Certified Accountants) is the global body for professional accountants, offering business-relevant, first-choice qualifications to people of application, ability and ambition around the world who seek a rewarding career in accountancy, finance and management.

ACCA supports its **219,000** members and **527,000** students (including affiliates) in **179** countries, helping them to develop successful careers in accounting and business, with the skills required by employers. ACCA works through a network of **110** offices and centres and **7,571** Approved Employers worldwide, and **328** approved learning providers who provide high standards of learning and development.

Through its public interest remit, ACCA promotes appropriate regulation of accounting and conducts relevant research to ensure accountancy continues to grow in reputation and influence.

ACCA has introduced major innovations to its flagship qualification to ensure its members and future members continue to be the most valued, up to date and sought-after accountancy professionals globally.

Founded in 1904, ACCA has consistently held unique core values: opportunity, diversity, innovation, integrity and accountability.

More information is here: www.accaglobal.com

Disclaimer:
This document includes references to third-party products/services. This is to give a sense of current market developments and are not endorsements of these offerings.

February 2020

Explainable AI:
Putting the user at the core

About this report

This report shines a light on explainable AI and its implications for accountancy and finance professionals.

FOR FURTHER INFORMATION:

Narayanan Vaidyanathan
Head of Business Insights, ACCA

Contents

Executive summary

Explainable artificial intelligence (XAI) emphasises the role of the algorithm not just in providing an output, but also in sharing with the user the supporting information on how the system reached a particular conclusion. XAI approaches aim to shine a light on the algorithm's inner workings and/or to reveal some insight into the factors that influenced its output. Furthermore, the idea is for this information to be available in a user-readable way, rather than being hidden within code.

> The middle path of augmenting, as opposed to replacing, the human actor works best when the user understands what the AI is doing; this needs explainability.

Historically, the focus of research within AI has been on developing and iteratively improving complex algorithms, with the aim of improving accuracy. Implicitly, therefore, the attention has been on refining the quality of the answer, rather than explaining the answer. But as AI is maturing, the latter is becoming increasingly important for enterprise adoption. This is both for decision making within a business, and post-fact audit of decisions made. Auditable algorithms are essentially ones that are explainable.

The complexity, speed and volume of AI decision-making obscure what is going on in the background, the so-called 'black box' effect, which makes the model difficult to interrogate. Explainability, or any deficit thereof, affects the ability of professional accountants to display scepticism. In a recent survey of members of ACCA and IMA (Institute of Management Accountants), those agreeing with this view, 54%, were more than twice the number who disagreed. It is an area that is relevant to being able to trust the technology, and to being confident that it is being used ethically. XAI can help in this scenario with techniques to improve explainability. It may be helpful to think of it as a design principle as much as a set of tools. This is AI designed to augment the human ability to understand and interrogate the results returned by the model.

The purpose of this report is to address explainability from the perspective of practitioners, ie accountancy and finance professionals. For practitioners, explainability can improve the ability to assess the claims made by vendors for their marketed applications, enhance value captured from AI that is already in use, boost return on investment (ROI) from AI investments; and augment audit and assurance capabilities, where data is managed using AI tools.

Key messages for practitioners

- **Maintain awareness of evolving trends in AI:** 51% of survey respondents were unaware of XAI, which impairs their ability to engage with the technology. To raise awareness, this report sets out some of the key developments in this emerging area.
- **Beware of oversimplified narratives:** in accountancy, AI is neither fully autonomous nor a complete fantasy. The middle path of augmenting, as opposed to replacing, the human actor works best when the user understands what the AI is doing; this needs explainability.
- **Embed explainability into enterprise adoption:** consider the level of explainability needed, and how it can help with model performance, ethical use and legal compliance.

Policymakers, for instance in government or in regulatory bodies, frequently hear the developer/supplier perspective from the AI industry. This report can complement that with a view from the user/demand side, so that policy can incorporate consumer needs.

Key messages for policymakers

- **Explainability empowers consumers and regulators:** improved explainability reduces the deep asymmetry between experts who understand AI, and the wider public. And for regulators, it can help reduce systemic risk if there is a better understanding of factors influencing algorithms that are being increasingly deployed across the marketplace.
- **Emphasise explainability as a design principle:** an environment that balances innovation and regulation can be achieved by supporting industry to continue, indeed redouble, its efforts to include explainability as a core feature in product development.

Introduction

Artificial intelligence (AI) offers the capacity for machines to 'learn' through exposure to examples and data, and to use that learning to drive inferences and decision-making (ACCA 2019).[1] This is a step beyond automation, and additionally involves cognition. Cognition provides a value layer that opens up new insights, while automation provides an efficiency layer to reduce the costs of doing so.

[1] For an introduction to AI for accountants, see ACCA's CPD course *Machine learning: an introduction for finance professionals*, <*https://www.accaglobal.com/sg/en/member/discover/events/global/e-learning/digital-technology/machine-learning.html*>.

> Broadly speaking, to 'explain' an AI algorithm means to be able to shine a light on its inner workings and/or to reveal some insight on what factors influenced its output, and to what extent.

There is also a second important reason why AI is featuring so prominently in our collective consciousness, namely that it is a general purpose technology (GPT). This means that it has the power and relevancy to reimagine, beyond incremental effects, our entire way of living.

That contrasts with, say, shipping containers, which were a clever innovation, but pertained specifically to the transport and logistics industry. The arrival of electricity at the turn of the twentieth century is a better parallel. It is not just a technology – it is an enabler that flows through every aspect of life whether professional or personal. AI will probably have a similar impact.

WHAT IS EXPLAINABILITY?

Broadly speaking, to 'explain' an AI algorithm means to be able to shine a light on its inner workings and/or to reveal some insight on what factors influenced its output, and to what extent, and for this information to be human-readable, ie not hidden within impenetrable lines of code.

Strictly speaking, interpretability is referred to as the ability to see inside a model transparently and understand its working, while explainability relates to situations where the model approach has to be inferred, rather than directly observed, because it is an opaque 'black box'. This report, being aimed at users, will use 'explainability' to mean quite simply an understanding of how/why a model returns the results it does.

Explainability matters for reasons that trace back to why AI is a different kind of technology. That it is 'cognitive' means that it can be non-trivial and easy to get wrong, given the complexities involved, and explainability is a checks-and-balances mechanism.

That it is a GPT means that explaining it cannot be relegated to a secondary or tertiary priority. Doing so can create serious risks for the public interest. Errors could range from honest mistakes to more sinister questions of incentive.

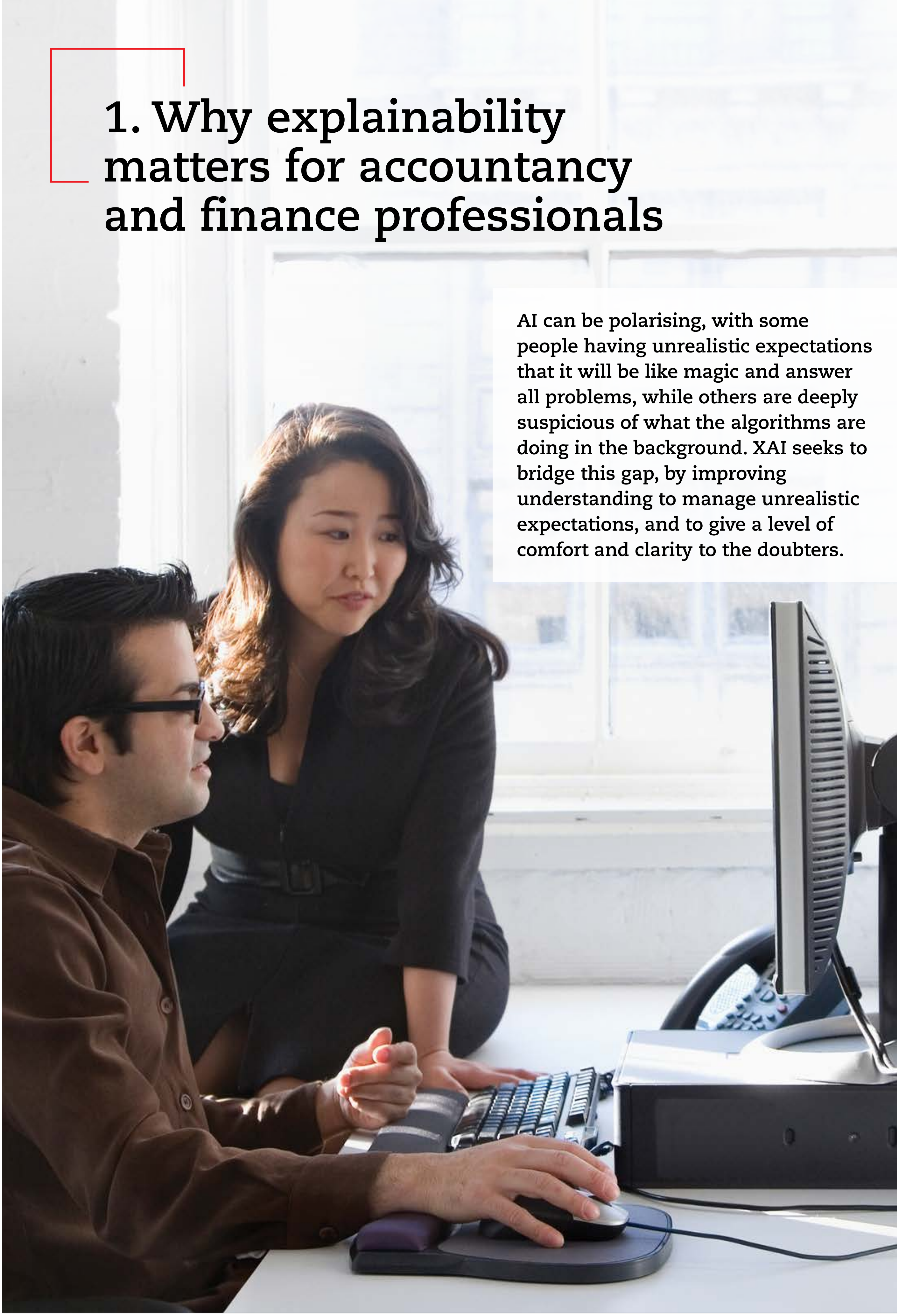
Has the AI performed in a certain way because ulterior motives were at play in its design or use? The public interest for greater explainability is intensified by the extreme asymmetry of understanding between those 'in the know', and the public at large.

Algorithms can be opaque and XAI can help to keep up with the scale and real-time decision making of AI. This is an emerging field and one that is expected to be a key focus in coming years, in order for AI to achieve mainstream use on a large scale.

Compared with the widespread use of mature everyday technologies, we are still in the early stages of AI adoption. Human systems and structures have an opportunity to use AI in a way that places the public good at the heart of its future development. This requires a mix of technological understanding, strategic decision making, governance mechanisms and agile delivery across multiple domains of subject-matter expertise – all underpinned by the highest standards of ethical behaviour. Explainability will be a central aspect of connecting all these elements.

1. Why explainability matters for accountancy and finance professionals

AI can be polarising, with some people having unrealistic expectations that it will be like magic and answer all problems, while others are deeply suspicious of what the algorithms are doing in the background. XAI seeks to bridge this gap, by improving understanding to manage unrealistic expectations, and to give a level of comfort and clarity to the doubters.



> Increasing awareness can improve the ability of accountancy and finance professionals to ask the right questions about AI products in the market and those in use within their organisations.

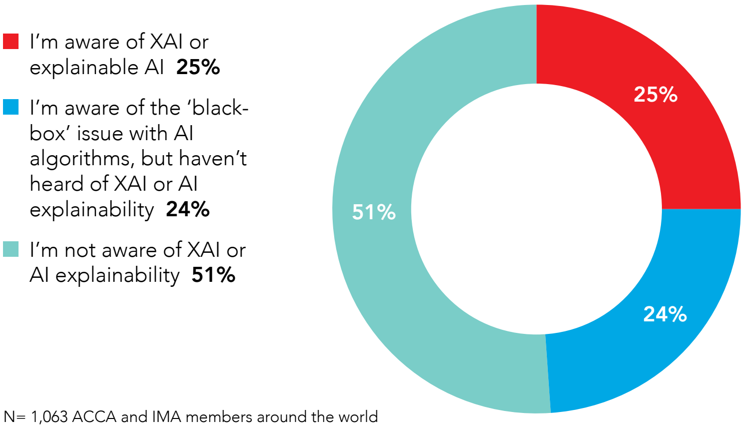
A survey of ACCA members conducted in November 2019 revealed that more than half of respondents were not aware of explainability as a focus of attention within the AI industry (Figure 1.1). Increasing awareness can improve the ability of accountancy and finance professionals to ask the right questions about AI products in the market and those in use within their organisations.

All the factors involving the public interest and the wider case for explainability apply, but it is worth additionally reflecting on why explainability matters for accountancy and finance professionals in particular.

ADOPTION – ENGAGING WITH AI

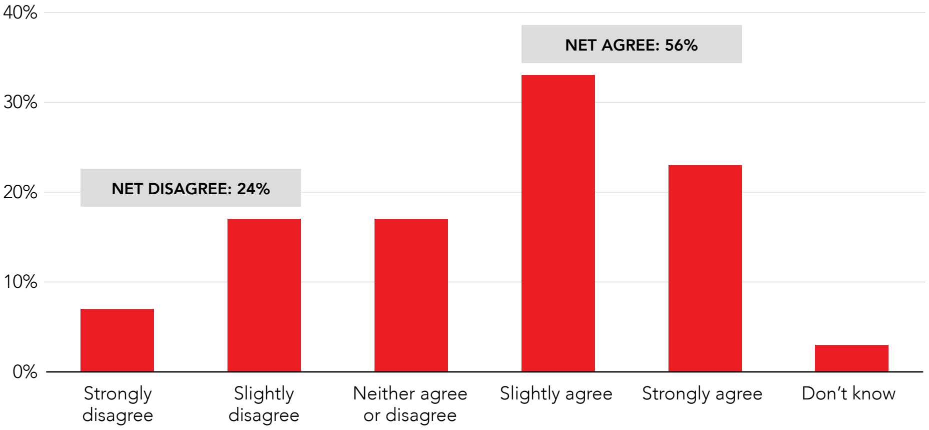
Professional accountants frequently refer to the concept of 'scepticism' as a Pole Star to guide their ability to serve their organisations. Scepticism involves the ability to ask the right questions, to interrogate the responses, to delve deeper into particular areas if needed and to apply judgement in deciding whether one is satisfied with the information as presented. More than twice as many survey respondents agreed than disagreed that explainability does have relevance when trying to display scepticism as a professional accountant (Figure 1.2).

FIGURE 1.1: Awareness of XAI



N= 1,063 ACCA and IMA members around the world

FIGURE 1.2: AI explainability affects the ability of professional accountants to display scepticism



N = 269, 'I'm aware of XAI or explainable AI'

> Rather than merely telling the user which transactions appear suspicious, an explainable approach would illuminate the components affecting the prediction most often or to the greatest extent.

XAI can provide a record/evidence or illustration of the basis on which the algorithm has operated. For AI to be auditable, it needs to incorporate principles of explainability. This provides an important foundation for adoption, whereas an opaque system in which the technology cannot be interrogated limits the ability to use model outputs. That's no longer a realistic position to take.

Moreover, establishing the ROI of adoption will be an important consideration for any organisation. And better explainability drives these returns because users no longer just wait to see what the model says, but have a more precise understanding of how the model can be used to drive specific business outcomes.

IMPACT – USE AT SCALE

The mathematics underlying AI models is theoretically well tested and has been understood for decades, if not longer, and converting it to production-ready models is a core task of data scientists. For accountancy and finance professionals, having an appreciation of the model they are using is essential, but their particular interest is scaling up its use to enterprise level, because this is the point at which the theory becomes reality. Scaling up presents challenges for deriving value from the model owing to the volume and variety of additional data, and the 'noise' that comes with it, to all of which the algorithm is exposed.

Greater explainability could help finance professionals understand where a model might struggle when production is scaled up. A recognised risk with AI algorithms is that of 'over-fitting'. This means that the model works very well with the training data, ie the historical data set chosen to train the algorithm, but then struggles to generalise when applied to wider data sets.

This defeats the purpose. It usually happens when the model takes a very literal view of the historical data. So instead of using the data as a guide to learn from, it practically 'memorises' the data and all its characteristics as they are (verbatim).

Consider a simplified example of a machine learning model for identifying suspicious transactions that need further investigation. During the training phase, the model observed that a high proportion of transactions that turned out to be suspicious occurred outside normal office hours. It therefore attached a high weight to this feature, the timestamp of the transaction, as a predictor for suspicious activity. When the model was applied more widely across all the organisation's transactions, however, the accuracy rate was poor. It identified a large number of 'out-of-hours' transactions as suspicious but most of these turned out to be perfectly legitimate, resulting in a considerable waste of time and resources, as the follow-on investigation of these flagged transactions had to be done manually.

A closer look revealed that the training data comprised transactions involving the core full-time staff of the organisation, but when rolled out across the organisation, the data comprised transactions involving all staff. This included the company's large pool of shift workers, who often worked outside the regular office hours as part of their agreed contracts.

Using the actual values of the timestamp feature from the historical training data set caused the model to misinterpret the link with suspicious transactions. An obvious and better correlation would have been to analyse the transactions' time stamp in relation to the contractual hours of the individual inputting the transaction. As discussed later, this improves model accuracy but increases complexity.

Rather than merely telling the user which transactions appear suspicious, an explainable approach would illuminate the components affecting the prediction most often or to the greatest extent. This could help the user spot when outlier values, such as the timestamp, were over-represented in the flagged transactions. In other words, an XAI approach could efficiently identify the most high-impact features that the algorithm is using to power its deductions and the level of importance (probability) it was attaching to each feature, when deciding whether to flag a feature as suspicious.

While this is a highly simplified illustration, the wider point is that when scaling a model with hundreds of features in a production environment with considerable noise, volume and complexity of inputs, details get lost or misinterpreted. And finding the reasons might feel like looking for a needle in a haystack. Ultimately, this situation creates costs of adoption.

XAI helps to test the model's decisions against finance professionals' domain knowledge and understanding of the process and business model.

TRUST – ETHICS AND COMPLIANCE

As AI enters the mainstream through scaling up, having the necessary governance, risk and control mechanisms becomes extremely important. Greater explainability can help to ensure that one is checking for the right things. Human responsibility doesn't go away, but explainability tools will be the support mechanism to augment the ability of professional accountants to act ethically.

Use at large scale highlights the role of explainability for model effectiveness, while ethics and compliance issues relate to the role of explainability for model trustworthiness. This includes ensuring that the model is fair, and is designed to allow for the rights of users in areas such as data privacy.

> Use at large scale highlights the role of explainability for model effectiveness, while ethics and compliance issues relate to the role of explainability for model trustworthiness.

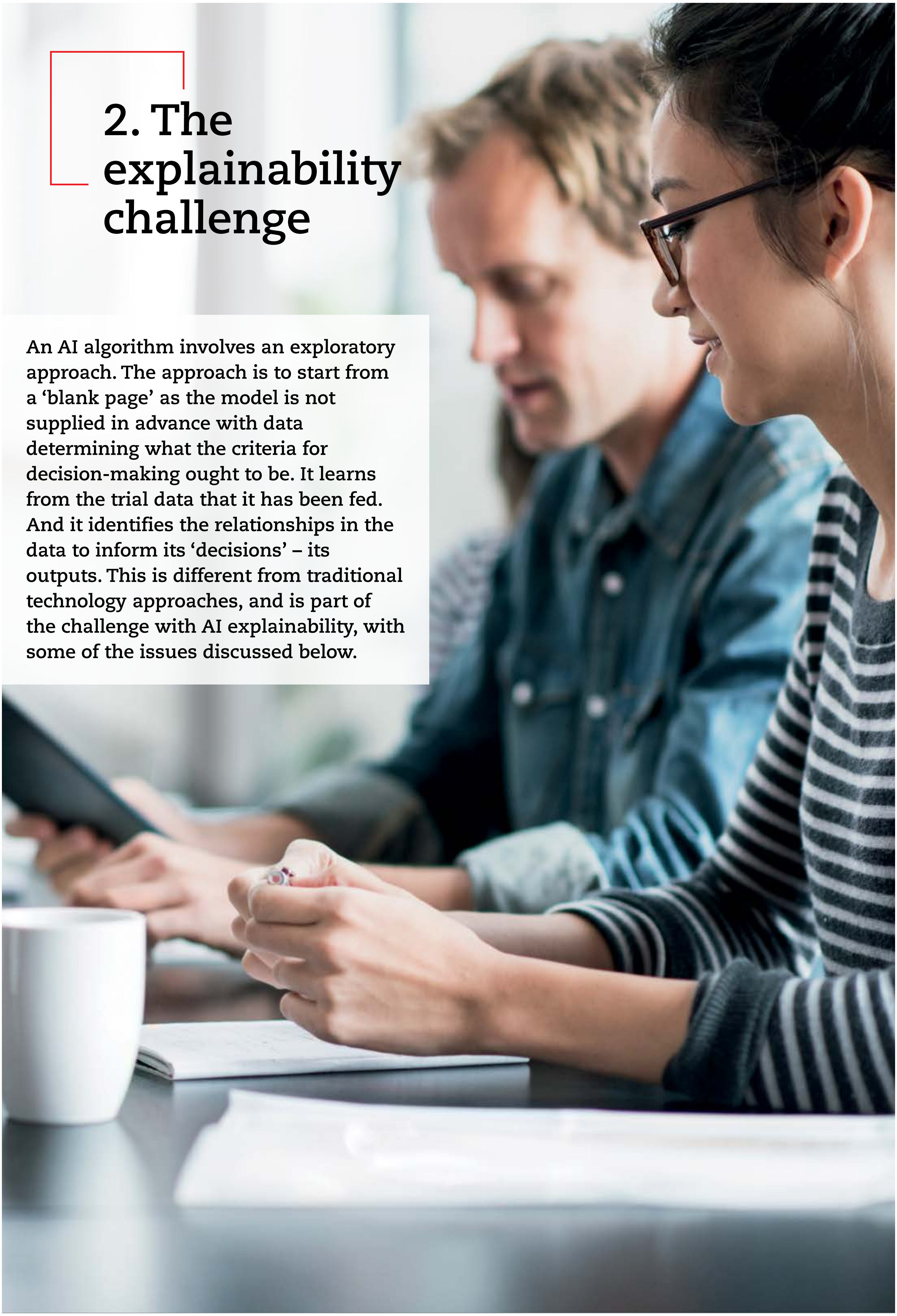
In the EU, for example, the General Data Protection Regulation (GDPR) requires various factors to be taken into consideration when using AI-based systems. For instance, is there sufficient transparency so that the user can understand the purpose of data use for making automated decisions? Have users given meaningful consent and is there a way for them to withdraw this if they wish? And is there sufficient explanation of how the algorithm works in general, and potentially how specific decisions for a particular user were arrived at?

Given the large volumes of data involved with machine learning and AI systems' ability to arrive automatically at decisions using the data, these questions quickly get tricky to resolve.

This is an evolving area and some of the questions related to a legal right to explainability may well be tested in court to establish the answers. While the precise legal boundary lines may be up for discussion, explainability is broadly accepted as a principle for long-term sustainable adoption.

2. The explainability challenge

An AI algorithm involves an exploratory approach. The approach is to start from a ‘blank page’ as the model is not supplied in advance with data determining what the criteria for decision-making ought to be. It learns from the trial data that it has been fed. And it identifies the relationships in the data to inform its ‘decisions’ – its outputs. This is different from traditional technology approaches, and is part of the challenge with AI explainability, with some of the issues discussed below.



Accuracy refers to the extent to which the algorithm's predictions or decisions turn out to be correct. The more accurate an algorithm, the less explainable it is, and vice versa.

In general, there is a trade-off between the accuracy and explainability of algorithms. Accuracy refers to the extent to which the algorithm's predictions or decisions turn out to be correct. The more accurate an algorithm, the less explainable it is, and vice versa. This is because more accurate algorithms tend to be more complex with a larger number of variables and multi-layered calculations with complex paths from input to output. This complicates explainability.

The type of model also affects explainability. Decision trees tend to be more explainable as they follow a sequential series of logical 'if this – then that' statements that create a path from input to output. Accuracy may be improved by combining decision trees, creating so-called 'random forests', but these are harder to explain. Also, if visual or text data is being analysed (eg contracts) this can involve deep learning using neural networks, which is opaque and less explainable.

The number of dimensions that the model is optimising across can create explainability challenges. It is not unusual for a machine learning algorithm to have several hundred or more features on which to model its results. So it is impossible for the human brain to visualise or intuitively understand the relationships that might be at play. Often, an accompanying need for speed of decisions further complicates the matter. This is why XAI is needed to augment the human ability to interrogate the outputs effectively.

Also, as time passes and circumstances change there can be so-called model drift, where the initial model is no longer as effective as it was when first created. This may be because of material changes to the data or relationships within it, beyond what can be dynamically updated or learned within the model.

The first order effect from this is, of course, accuracy, which suffers. And in dealing with this a variety of techniques are used, such as updating the training/historic data or other parameters. Making changes improves accuracy, but as noted earlier, increases complexity and makes explainability more challenging.

EXPLAINABILITY IS NOT JUST AN AI ISSUE

These observations can tempt one to think of explainability as an AI-specific issue – a new challenge confronting the accountancy profession. But, in fact, the need for accountancy and finance professionals to be able to explain decisions is as old as the profession itself.

Human decision making is not fully explainable either. It is just more familiar. Business leaders might take a view based on their years of experience or on their personal judgement, and auditors trying to unpick the decisions at a later date may find themselves dealing with partial information and an opaque decision-making process.

Consider the issue of bias, a poster child of the complexities of AI. Bias is an issue in AI decision making because bias is an issue both in individual human beings and in society. ACCA's report on this matter (2017) highlights a range of examples of how bias affects audit quality – yet the report itself is not about AI.

So it is important not to use the fact that the decision comes from an algorithm as a reason to disengage from this issue. This technology is likely to be a reality in the future, and the AI industry's increasing acceptance of the need for greater explainability is a step in the right direction which will aid the ability of accountancy and finance professionals to adopt AI.

3. Approaches to explainability

The purpose of this report is to address explainability from the perspective of the accountancy and finance professional. So the expectation is that the reader will need an appreciation of the thinking about and options for explainability, so that they can interrogate AI applications for their organisational needs.

'Global explainability' is the ability to understand the overall structure or approach of a model, while 'local explainability' is the insight pertinent to understanding how a specific data point has been treated within the model.

The report is not concerned with developing the mathematical underpinnings of the explainability techniques themselves. Consequently, the focus is on understanding explainability conceptually, using some approaches to highlight key points. These are further illustrated with examples from practitioners.

Given the multifaceted nature of the explainability challenge, it is perhaps not surprising that various factors affect our understanding of it. Figure 3.1 provides a categorisation along two dimensions, which can serve as a useful starting point.

WHAT WE ARE TRYING TO EXPLAIN

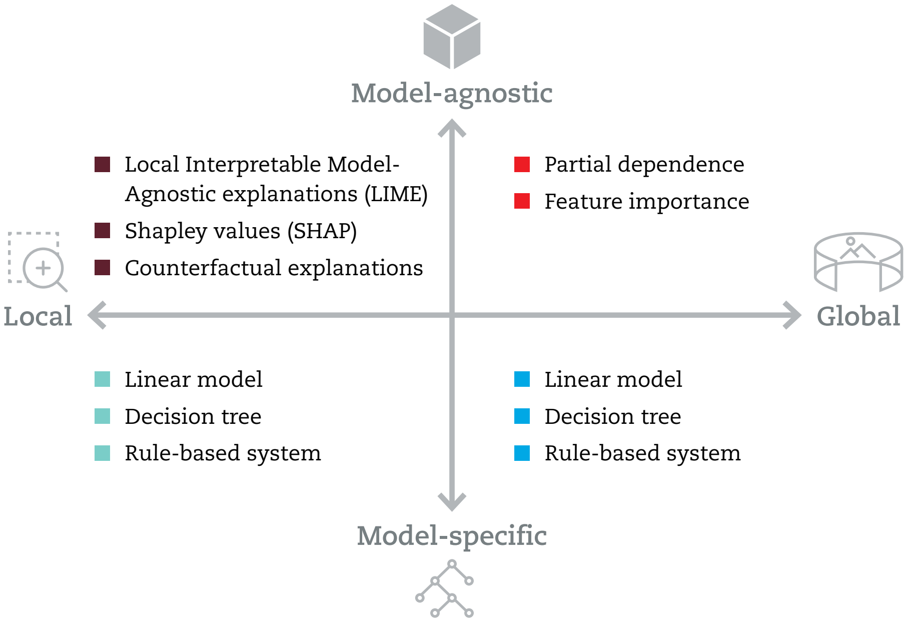
The horizontal axis in Figure 3.1 extends from local to global explainability. 'Global explainability' is the ability to understand the overall structure or approach of a model, while 'local explainability' is the insight pertinent to understanding how a specific data point has been treated within the model.

This distinction is required because one does not automatically follow from the other. Understanding the principles of why the sun rises in the east and sets in the west does not automatically explain why sunrise happened at a certain time on a certain day. In practice, both global and local explainability are relevant and important although, for given situations, certain stakeholders may value one more than the other.

Consider a credit model used by a bank to identify which applicants to approve for a housing loan. To inform its profitability outlook and risk appetite, the bank will want to understand the impact of interest rates or predicted provisions for bad loans. And it may be viewing this question as part of a broader strategy question, which places home lending within the wider landscape of the opportunity cost of being in this business – instead of focusing on, say, lending to small businesses.

The bank's management may be interested in understanding the overall approach and structure of the model and how it might create better outcomes over time for the bank. An understanding of how various parameters (or 'features') are prioritised by the model and how they relate to the bank's lending strategy matters.

FIGURE 3.1: Approaches to explainability



Source: A. Koshiyama, UCL; with thanks to Janet Adams, TSB Bank for making this chart available to ACCA, alongside wider insights in relation to explainability.

> Some of the underlying ideas for improving explainability are surprisingly familiar, even if they are applied in quite complex or sophisticated ways. The use of 'perturbations' for example, is conceptually like sensitivity analysis.

For example, the model approving a larger proportion of higher loan-to-value, riskier, home loans may be acceptable as part of a strategy for identifying cash-poor customers who may be high-potential entrepreneurs in the early stages of their business. Global explainability can help to understand the overall approach of the model and relate it to the bank's priorities.

Now consider an applicant rejected for a home loan by the bank (using this model), and who is understandably disappointed. Furthermore, the applicant may well disagree with the decision and wish to challenge it, or at least to understand why this has happened. This question matters to them because a declined application harms their credit record and affects their chances of obtaining a home loan from a different lender. This applicant doesn't care about the overall structure of the model. They care about one thing – why or how their individual decision was arrived at.

Even for the bank, because these are usually regulated industries, there may be obligations such as 'Treating Customers Fairly' that require them to explain how the customer's case was handled. Local explainability can be crucial here for identifying specific factors on the critical path followed in deciding the applicant's case.

HOW WE ARE TRYING TO EXPLAIN IT

The vertical axis on Figure 3.1 refers to specific and agnostic approaches to explainability. As the names suggest, model-specific (or intrinsic) approaches analyse a particular model and its workings, and use this to devise explanations informed by that model.

On the other hand, agnostic (or post hoc) approaches, can be applied to any model. They sit on top of a given model, like a sort of translation layer, to derive explanations from that model. Often, for more complex models, it may be difficult to look directly inside the workings of the model, and a 'model agnostic' tool can help here.

Some of the underlying ideas for improving explainability are surprisingly familiar, even if they are applied in quite complex or sophisticated ways. The use of 'perturbations' for example, is conceptually like sensitivity analysis. By measuring the extent to which changes in an input affects the model's output, a sense of the relative importance of that input can be established. Taken further, a 'counterfactual' seeks the level of perturbation at which the change in the input actually changes the decision at the output stage. While these analyses relate to the sensitivity of a feature in relation to a specific data point, the logic can also be extended to assess sensitivity to a feature globally across the model, a technique used in Partial Dependence Plots.

In order to get a slightly more intuitive sense of all this, it may help to consider a few techniques that are well recognised in the market.

A good starting point is the decision tree. This is perhaps one of the most well-recognised logical constructs, and is based on the simple principle of mapping out all the paths that lead from input to output, usually through a sequence of 'if this,-then that' commands. For businesses, this is a popular basis for building models, because it is intuitively easy to understand the pathways within the model. The ease of looking inside a model of this type is why it lends itself to both local and global explainability.

Decision trees are not the most complex model type available; others, such as neural networks, are more sophisticated. But many accountancy-relevant uses involve relatively structured data for which decision trees work well. Also, they can be strengthened through 'ensemble' techniques, where multiple models can be combined to improve the overall accuracy. Decision trees combined in this way, called 'random forests', can benefit from the visibility and transparency that trees provide, while improving results.

A related idea here is that of 'gradient boosting', where a systematic step-by-step process is followed to add successive trees. This allows for an inductive approach where one can see what changes are happening as each additional tree element is added, to ensure the best possible combination at the end.

> The explainability advantage is that users can look at the second model, which has translated the behaviour of the original model to a form that can be analysed more meaningfully.

Decision trees lend themselves to model-specific approaches to explainability. Their relative transparency and interpretability allows for the user to see what's happening within the model and to use that to inform their explanation of it.

On the other hand, a Local Interpretable Model-Agnostic Explanation (LIME) provides a mechanism that users can deploy on any kind of model. The aim is to derive a greater level of explainability than the user might have been able to achieve just by examining the model itself. This could be because the model is quite opaque (a 'black box') and, unlike, say, decision trees, it is difficult to get a sense of the pathway from input to output.

LIME works by creating a second model that is a simplified approximation of the original model. This approximation generally takes the form of a more understandable model structure such as a decision tree or one with linear relationships between the variables. This is a local explainability technique – LIME is not designed to explain the overall workings of the model. Instead, it looks at the target variable or end result that the model is designed to achieve. It then uses the approach of perturbations to develop sophisticated copies of the outcomes that the model generates. So LIME notes that the model is sensitive to parameters in a certain way, and to a certain extent, and it reproduces (as closely as possible) that outcome, through a second model which is more transparent than the original.

The explainability advantage is that users can look at the second model, which has translated the behaviour of the original model to a form that can be analysed more meaningfully. And this approach can be applied regardless of the structure of the original model – making it agnostic or model-neutral.

Clearly, being able to translate any model in this way is not a trivial exercise. But LIME is helped by a critical mass of adoption in the market and is aided by the availability of open-source resources that have created a community-based approach to improving and refining it.

For those interested in delving deeper, a host of explainability techniques are available. SHAP (SHapley Additive exPlanations) is another model-agnostic technique that calculates the contribution of particular features in influencing a data point. In other words, one could 'deconstruct' the conclusion for a particular data point (eg its predicted value) by each individual feature. This would tell the extent to which a particular feature contributes to the predicted end result for the target variable.

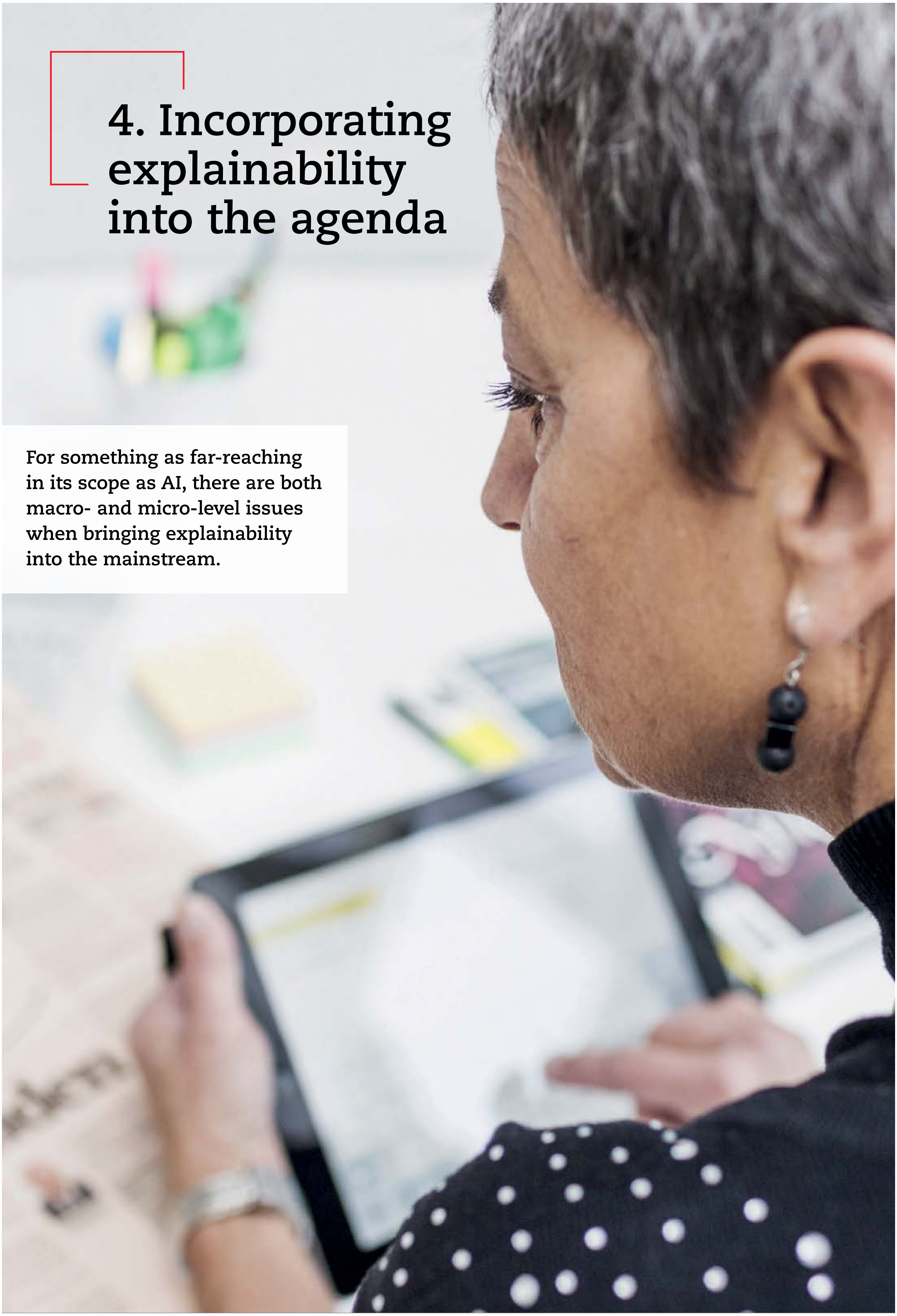
There are a range of off-the-shelf XAI tools which are growing in popularity. Some examples of these include Google's What-if tool (WIT), supplied as part of Google Cloud; SKATER and ELI5 packages for Python; and Interpretable Machine Learning for R.

On the other hand, specialist providers such as Chatterbox Labs provide their own patented explainable AI applications, which the company advises can work with any AI engine. According to Dr Stuart Battersby, chief technology officer at Chatterbox Labs:

> “ *'We built our patented Explainable AI software to work with any AI engine (including Google Cloud, Microsoft Azure, IBM Watson, AWS Sagemaker and in-house built AI engines) because we recognise that very few enterprises have a coherent and unified AI strategy. By layering our Explainable AI on top of existing AI assets, business users can audit, trace and explain AI outcomes irrespective of the underlying AI models used, and there is no requirement to change, modify or disrupt their existing AI investments'.*

4. Incorporating explainability into the agenda

For something as far-reaching in its scope as AI, there are both macro- and micro-level issues when bringing explainability into the mainstream.



For policymakers, whether in government, regulatory bodies or elsewhere, addressing wider macro-level concerns is a natural starting point.

For policymakers, whether in government, regulatory bodies or elsewhere, addressing wider macro-level concerns is a natural starting point. There is a public interest consideration in ensuring that citizens are appropriately protected. The form of this, whether as a right to explanation or greater transparency, is already being discussed in many jurisdictions, and will be continually assessed as developments in AI unfold.

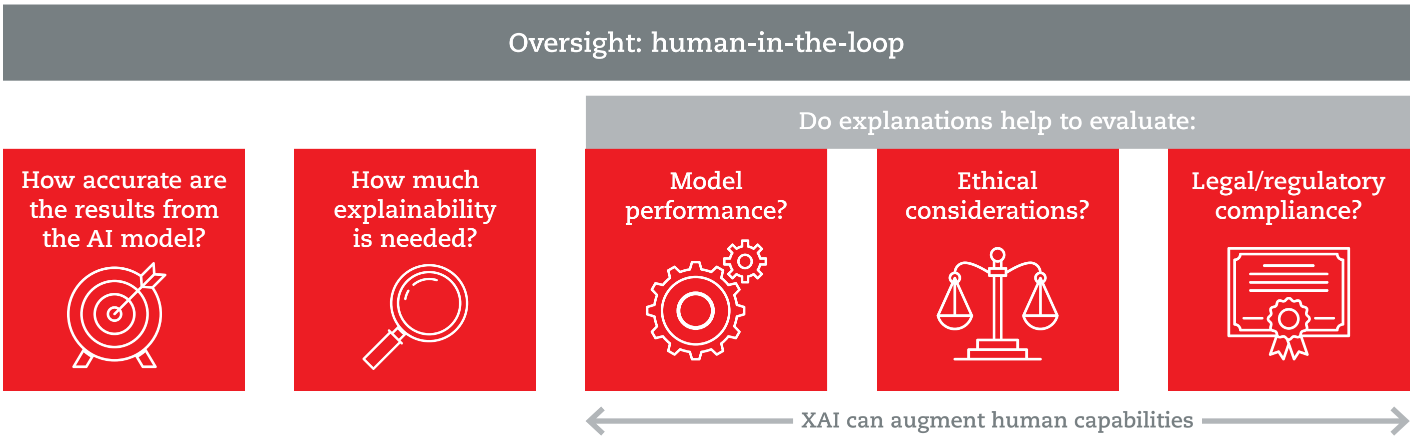
The potentially widespread future adoption of AI means that its involvement may not even be immediately visible to the end user of every application. So there could be systemic risk if policymakers are slow to emphasise to developers the importance of explainability as a design principle and their expectations that it will be incorporated into the systems that affect the lives of individuals in society.

As the history of the 2008–9 financial crisis demonstrated, it can be difficult to make the link between sophisticated financial products traded in wholesale markets in an international financial trading centre; with indiscriminate home lending occurring in small towns in another country thousands of miles away.

Policymakers will lag behind those at the forefront of innovation, which is neither surprising nor a problem in itself. In fact, many regulators (through mechanisms such as the 'sandbox') recognise this and try to get an early impression of how ideas are developing, where possible. The goal, ultimately, is for them to provide an enabling environment – one that encourages the right behaviours from participants and provides a forum for discussion where different perspectives can be debated, with a balance between innovation and regulation.

While the above macro considerations are needed for creating such an enabling environment, accountancy and finance professionals will frequently find themselves dealing with more micro-level concerns. As they adopt AI tools in their organisations, understanding how to think about explainability and incorporate this into the agenda will matter. Below, we set out a starting point for engaging with these concerns (summarised in Figure 4.1).

FIGURE 4.1: Embedding AI explainability into enterprise adoption


Oversight: human-in-the-loop
Do explanations help to evaluate:
How accurate are the results from the AI model?
How much explainability is needed?
Model performance?
Ethical considerations?
Legal/regulatory compliance?
XAI can augment human capabilities

> Not all situations need the same level of explanation. Greater explainability comes at a cost, and the returns need to be commensurate with that.

i) How accurate are the results from the AI model?

a) *What is the business question that the model needs to answer?*
This sounds obvious but, given the detail and complexity of machine learning models, this can be misunderstood. An example would be focusing on an 'adjacent' question that is confused with what one is trying to understand. Are we really trying to predict where late payments will come from? Or is the issue that our customer base is highly concentrated, making us heavily exposed to even a very small number of customers not paying on time? Should the real question be: how do we predict customer characteristics relevant to diversifying our income stream? Getting this right matters a lot in machine learning because there are many parameters to optimise, and fine-tuning the wrong variables can completely change the model's workings and output.

b) *How accurate is the model for answering this business question?*
This is important baseline information that sets the foundation for the 'accuracy versus explainability' balance. It allows for an understanding of the ROI associated with this model. The section on benefits case of explainability for finance professionals (impact – use at large scale) highlighted how not knowing what to look for when interrogating model performance has real consequences for increased costs, in that instance, of chasing false alarms. Costs of inaccuracy add up significantly when scaling-up across an organisation.

ii) How much explainability is needed?

Not all situations need the same level of explanation. Greater explainability comes at a cost, and the returns need to be commensurate with that. So it is about an appropriate level of explainability for a situation, rather than setting fixed thresholds.

A recommendation for a television show is a low-stakes decision and simply providing the algorithm's output might be enough – cost effective for the organisation and acceptable to the customer. If it was not what the customer was looking for, they'll endure some lost time before they abandon the show.

On the other hand, quantifying a business risk in a regulated industry can require much higher levels of explanation to protect against legal liability and reputational loss. This comparison is extreme but, even within a business, some use cases have higher stakes than others. This could be linked to how time-sensitive the decision is, whether it is reversible, upfront sunk costs versus opportunity for spreading these over time, and the nature of the data (eg client details), etc.

iii) What do explanations help to evaluate?

a) *Model performance*
The aim of step (i) was to evaluate how good the model is at answering the business question – accuracy. This step is about understanding why the model is as good (or bad) as it is at achieving accurate results. The model's boundaries for value extraction and risk protection can then be better understood.

This approach to explainability will be informed by whether global and/or local explainability will be most appropriate. And the type of model being used will determine whether model-specific or agnostic approaches would be most effective.

b) *Ethical considerations*
This is a broad area, but one of the most common concerns is of bias in the data, and the consequent bias in decision making. There will be many in an organisation, whether in technology or business units, with strong incentives to drive greater adoption.

Explainability approaches to AI allow humans to augment their ability to interrogate and be sceptical – they are not a substitute for this activity.

That's a perfectly acceptable position, but an organisation needs to encourage this alongside an objective mindset, with appropriate levels of challenge.

Without that, adoption may happen but it will not be sustainable. Reputational damage can occur with one misstep, and trust built over years can evaporate permanently in an instant. Professional accountants have an ethical and public interest responsibility. It is crucial that they bring this to bear in looking objectively at the use of algorithms. And constructive challenge is predicated on being informed – which is assisted by greater explainability.

c) *Legal/regulatory compliance*
Disclosure and regulatory reporting obligations may require providing evidence to third parties. These could be end-users or customers, for example where their personally identifiable information is being used. Or it might be bodies such as financial regulators, tax authorities or others with oversight responsibilities.

Local explainability might be relevant for end-users while regulators might, in addition, require global explanations for clarity about the approach underlying the decisions made.

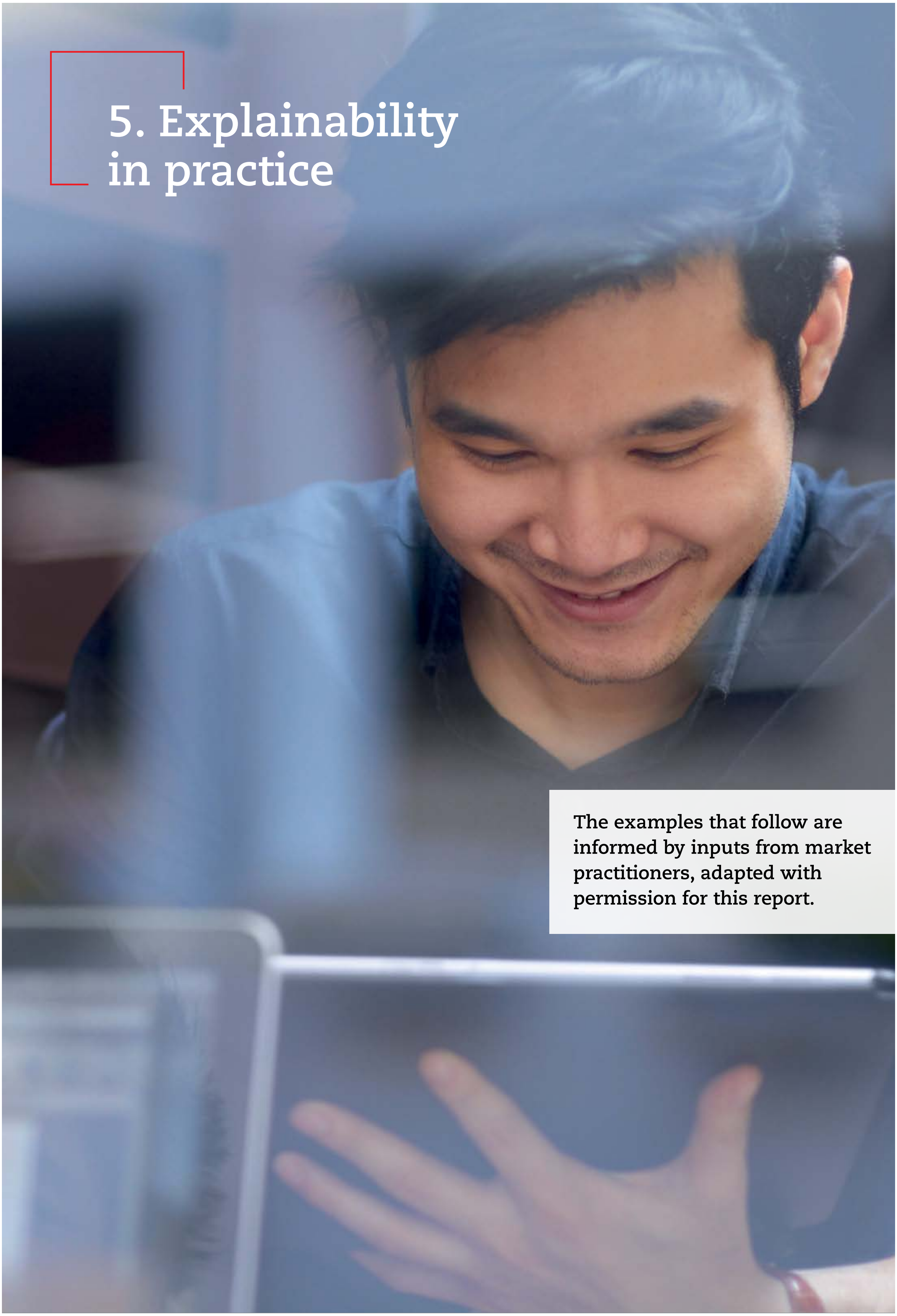
An auditable algorithm with clear trails of events, decisions and timelines –ie an explainable one – is much more likely to meet regulatory requirements than one that is opaque.

iv) *Oversight: the human in the loop*
The whole point of XAI is to bring into focus design principles that improve the ability for human judgement to be exercised appropriately. It would be a supreme irony if, in order to solve the problem of not knowing what the 'black box' was doing, people decided to blindly trust whatever the translation model provided by LIME (see above, section 3) was telling them.

Legal systems and cultural norms still place decision accountability ultimately at the door of the human actor. As things stand it is not possible to outsource responsibility to the algorithm or prosecute it in place of a person. Understanding the extent and reasons for manual overrides during implementation and how oversight/governance mechanisms are incorporated into an organisation are relevant here.

Explainability approaches to AI allow humans to augment their ability to interrogate and be sceptical – they are not a substitute for this activity.

5. Explainability in practice



The examples that follow are informed by inputs from market practitioners, adapted with permission for this report.

Mr Ke Jin (Collin), Deloitte China

Mr Jin is the national audit and assurance innovation leader, and managing partner of the Innovation & Digital Development Center of Deloitte China. With extensive experience in AI, robotics, big data analytics, and cloud, he has directly led and driven Deloitte China's development, marketing and application of a number of innovation and digital products. These include Deloitte's intelligent finance robot, ie robotic process automation (RPA) and robotic cognitive automation (RCA), the 'Spotlight' global audit analytics big data platform, the intelligent finance fraud monitoring platform, the bank loans risk analysis system, bank transaction analytics tools, the intelligent document review platform, the 'wise leasing' application, and the unmanned aerial vehicle (UAV) stock-taking application. He graduated from the London School of Economics and Political Science in 2006 and joined Deloitte China the same year.

For large banking institutes, a credit review involves significant manual effort across multiple departments. And, importantly, the review is expected to be predictive for potential credit risk. Deloitte China has designed and built an XAI-based platform – iCredit – to collect and analyse the borrower's finance data, enterprise quality, industry development trends, etc. to give meaningful support to a more informed credit review process. The platform leverages emerging technologies such as machine learning, natural language processing (NLP), and optical character recognition (OCR).

The challenge here is the rising risk from fraudulent details in loan applications. Credit assessment depends greatly on industry expertise, and historically there has not been a rigorous and consistent mechanism for capturing this. Legacy third-party credit-rating agencies do exist, but there is variability in level of use of these and reliance on them by the market. Top banks, for example, have their own credit frameworks and procedures, which rely heavily on manual work and expert experience.

In the absence of reliable platforms or applications for credit assessment, the entire process depends significantly on post-loan reporting by account managers in the bank. And information is often contained in large amounts of unstructured document data that is not easy to extract and use.

So iCredit has been very attractive to banks, given Deloitte China's knowledge of and insights to the market, the diversified data sources used, and the newest technologies that this application has been built upon.

Detailed research was conducted to understand the possible means used to commit fraud, to inform a comprehensive analysis of this problem. As a result, a standardised workflow for credit assessment was designed that was based on industry expertise.

This XAI-based platform for credit assessment acts as a reliable information source to support decision making, using machine learning models for data calculation and risk prediction. It integrates multiple data sources and incorporates consistent, centralised data cleansing, storage and management. OCR and NLP are used to read documents and extract data.

The tool has enabled all credit assessment work streams and teams to be online with a standardised workflow, saving hundreds of hours in this work each year. It has enabled a high-speed digital risk-assessment model and provided greater flexibility and adaptability for all departments. This is accompanied with proven, accurate, loan-risk identification using an intelligent model alongside a comprehensive loan dashboard.

Deloitte China is currently in the process of promoting and deploying iCredit to the largest state-owned and commercial banks in China.

iCredit uses a 'gradient boosting' framework as the core model for these supervised learning problems (it uses XGboost libraries for programming). Training data, such as financial statements and company operational status, are used to predict a target variable, namely, the 'credit score'. Since there are a number of indicators with a material impact on the credit rating result, a framework of parallel tree boosting is used to identify the most important factors in a fast and accurate way.

EXPLAINABILITY FOR iCREDIT IS PRODUCED IN SEVERAL WAYS

The iCredit model is based on industry expertise and thus all factors are created and selected by the joint work of Deloitte's credit experts and machine learning experts. The platform has a distinct feature in that it provides the end-users with the flexibility to choose the applicable pool of indicators by themselves. In that way, they can manage what kind of data set is involved, or which year of data is calculated. Users are able to see what data has been factored in the model and which factors contribute to the credit rating. Also some of the key information is made available in a so-called 'white-box' and displayed to users so they can judge the potential risks.

iCredit can be a very helpful tool to assist the finance and credit professionals working in a bank or in a firm auditing the approach of a bank. This is principally because the platform generates a credit rating for reference rather than a simple answer to a loan approval decision. So the finance professionals can go through the details of the credit rating such as the specific items identified, in order to inform their understanding of risk process and what is happening in the background. Some details of this are provided below.

Step 1 – Which index?

Users can select or change the risk factors by themselves. Risk assessors ask for not only a risk grade but also the index or factors (examples of which are shown in Figure 5.1) that determined which grade the borrower received, which then fed into the overall credit rating.

Step 2 – Industry benchmarking

An individual grade is not enough in itself, and users also need to know, eg why the grade in a specified index is low. Detailed, reliable and sophisticated industry benchmarking is provided to help users learn more about the context of the debt level. This includes, for example, looking at the industry leverage status, to estimate whether one is looking at an outlier.

Step 3 – Data effectiveness

Several years of data are prepared for the calculation but it is left to the users to decide on which data/data set is used in the model. This is because, after detailed exploration, it became clear that not every piece of data is effective in every situation. So it is best to allow the user to make the decision. Deloitte's job is make sure the data is valid, eg letting users know how many years of data are being used, or which year of data is being used, to make sure out-of-date impacts are eliminated.

Step 4 – Included/Not Included

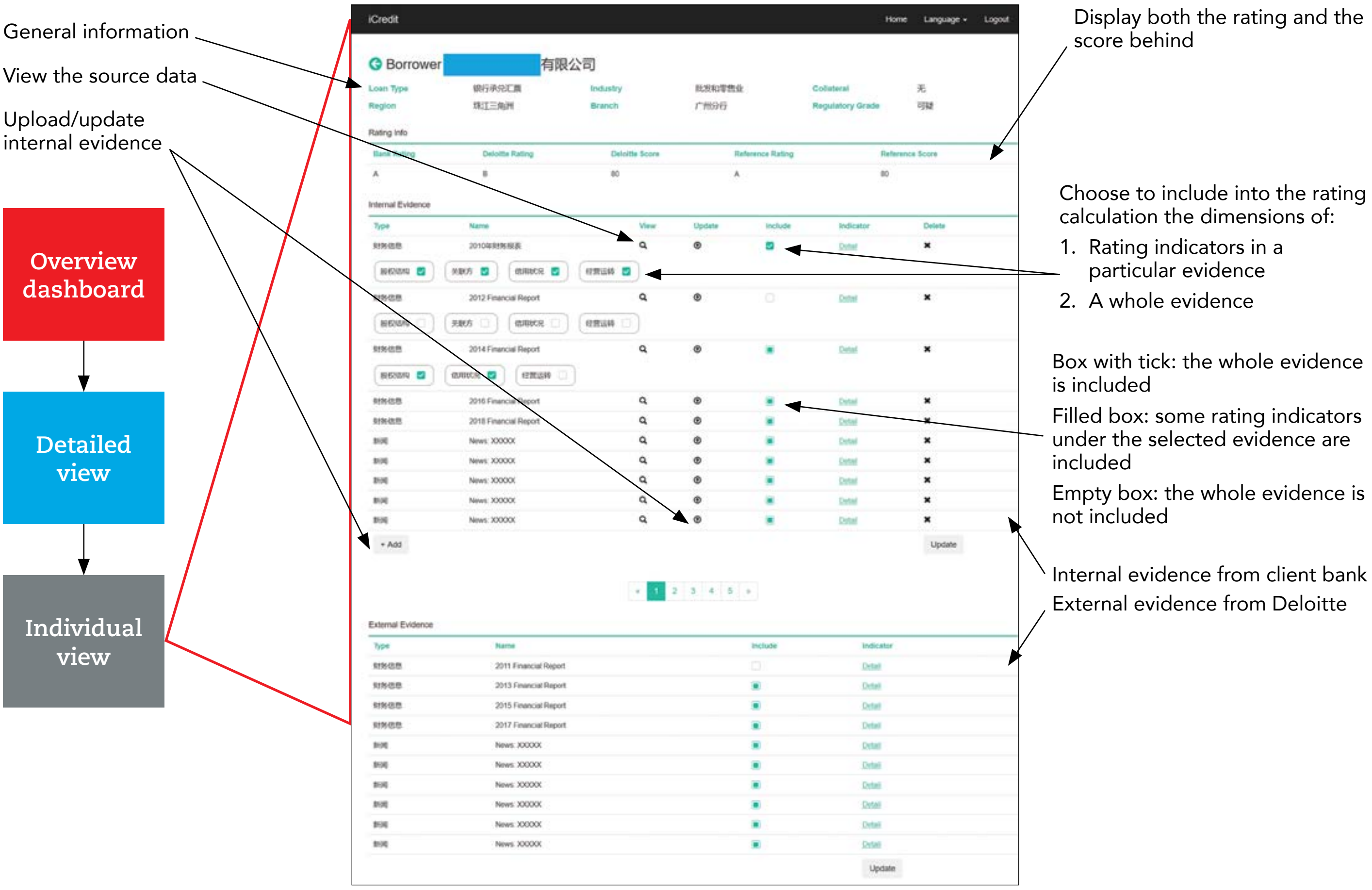
The model is flexible enough to add or eliminate factors or data and it is agile enough to refresh its results as needed. This is so that, although the model can give a final result, the user can dynamically adjust the data involved, the index, and the weights.

This AI-driven credit assessment generates persuasive and explainable results, which can be reviewed and investigated. Figure 5.2 gives a view of the dashboard that users work with as they use the model.

FIGURE 5.1: Risk factors for credit assessment

Enterprise quality	Operating conditions	Finance status	Development prospect
Equity structure	Business operations	Debt paying ability	Economic climate
Related parties	Main business	Profitability	Industrial policy
Credit status	Marketing capacity	Operating capacity	Industry competition
Public opinion	Competitiveness	Fraud histories	Firm growth
Social image	Sustainability	Industry status	National strategy
Legal disputes	Regulatory compliance	Cash flow	Global situation

FIGURE 5.2: Dashboard for the iCredit account credit rating


General information
View the source data
Upload/update internal evidence
Overview dashboard
Detailed view
Individual view
Display both the rating and the score behind
Choose to include into the rating calculation the dimensions of:
1. Rating indicators in a particular evidence
2. A whole evidence
Box with tick: the whole evidence is included
Filled box: some rating indicators under the selected evidence are included
Empty box: the whole evidence is not included
Internal evidence from client bank
External evidence from Deloitte

iCredit Home Language ▾ Logout

Borrower [redacted]有限公司

Loan Type	银行承兑汇票	Industry	批发和零售业	Collateral	无
Region	珠江三角洲	Branch	广州分行	Regulatory Grade	可疑

Rating Info

Bank Rating	Deloitte Rating	Deloitte Score	Reference Rating	Reference Score
A	B	80	A	80

Internal Evidence

Type	Name	View	Update	Include	Indicator	Delete
财务信息	2010年财务报表	🔍	⊙	☑	Detail	✖
股权结构 ☑	关联方 ☑	信用状况 ☑	经营运转 ☑			
财务信息	2012 Financial Report	🔍	⊙	□	Detail	✖
股权结构 □	关联方 □	信用状况 □	经营运转 □			
财务信息	2014 Financial Report	🔍	⊙	■	Detail	✖
股权结构 ☑	信用状况 ☑	经营运转 □				
财务信息	2016 Financial Report	🔍	⊙	■	Detail	✖
财务信息	2018 Financial Report	🔍	⊙	■	Detail	✖
新闻	News: XXXXX	🔍	⊙	■	Detail	✖
新闻	News: XXXXX	🔍	⊙	■	Detail	✖
新闻	News: XXXXX	🔍	⊙	■	Detail	✖
新闻	News: XXXXX	🔍	⊙	■	Detail	✖
新闻	News: XXXXX	🔍	⊙	■	Detail	✖

+ Add Update

« 1 2 3 4 5 »

External Evidence

Type	Name	Include	Indicator
财务信息	2011 Financial Report	□	Detail
财务信息	2013 Financial Report	■	Detail
财务信息	2015 Financial Report	■	Detail
财务信息	2017 Financial Report	■	Detail
新闻	News: XXXXX	■	Detail
新闻	News: XXXXX	■	Detail
新闻	News: XXXXX	■	Detail
新闻	News: XXXXX	■	Detail
新闻	News: XXXXX	■	Detail
新闻	News: XXXXX	■	Detail

Update

Stuart Cobbe, Adviser, MindBridge.ai

Stuart is an adviser at MindBridge.ai with a particular interest in the direction of product development within audit, as well as financial statements and tax compliance processes for small, medium, and large enterprises alike. He is a pioneer in the use of data for professional services. In his previous role at a top 20 accountancy and advisory firm, Stuart led a team involved with full stack application development, data manipulation, visualisation, and application of AI to audit processes.

Currently, AI is being used to improve the rate at which auditors can detect fraud and unusual transactions, as well as to provide a greater understanding of clients' data. A good example at MindBridge was the use of AI to identify a situation where a sales invoice had sales tax attached to it when it should not have done, the total value of which was material. This allowed the end client to reclaim a substantial amount of income tax from the tax authorities.

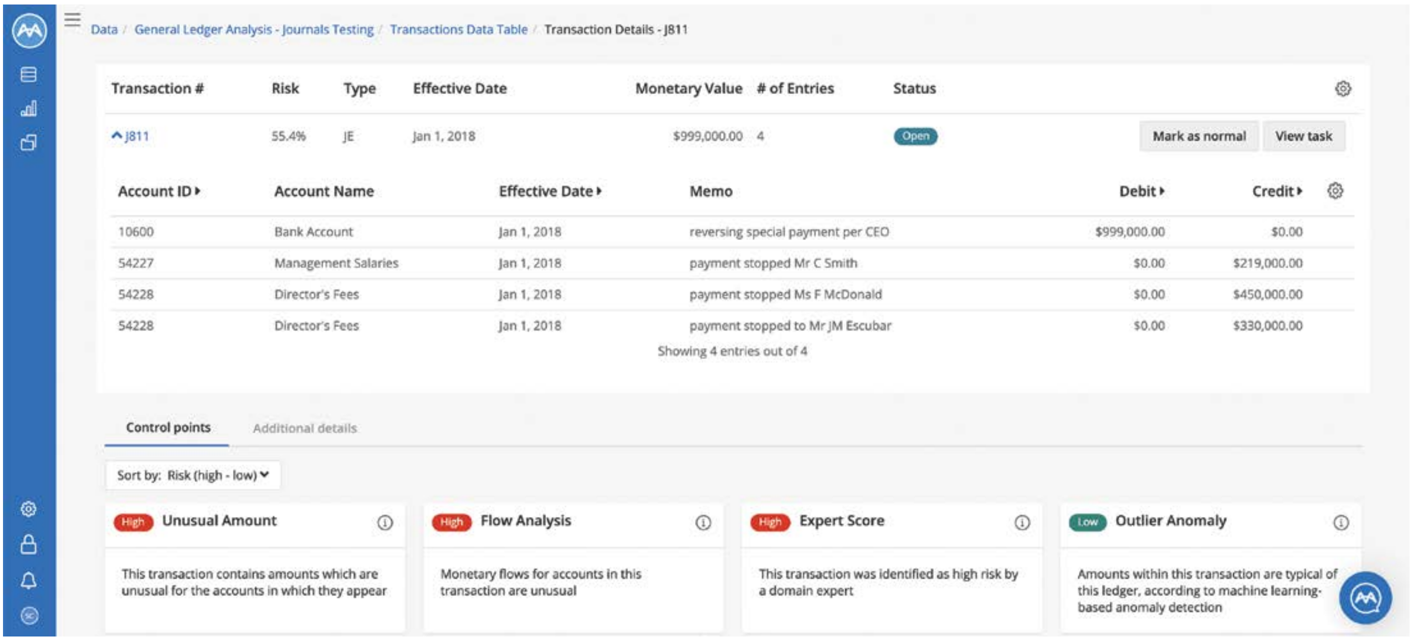
Being able to interrogate the model effectively was important here because, at first glance, this transaction appeared to be completely normal, and so it was not immediately obvious why it had been flagged. It was only on viewing the risk factors, together with a number of other, similar transactions, that the error was spotted. It is this ability to look at a transaction, armed with more information than ever before, that allows the error to be identified.

The MindBridge methodology for explainability is to highlight the attributes of certain transactions that cause the AI to identify them as risky. These might be simple attributes of the data, or they might be more complex. In the tax case described above, both 'outlier anomaly detection' and 'rare flows' indicators were triggered, which indicated that there was something unusual about the nominal codes that this transaction was involving. On investigation, it was the sales tax code, in particular, that was causing this.

The detection methodologies used rely on some concept of 'rarity' in the data, which means that judgement (ie a human) is necessary in order to decide whether this rarity is due just to an unusual business transaction, or is some form of error. Comparison with other, similar, transactions must be done, and combined with the auditor's knowledge of what is normal for this business and for transactions of this type.

FIGURE 5.3: The highest-risk transaction in a sample dataset

This transaction has an unusual monetary flow, meaning it is rare to see a debit to the bank account and a credit to management salaries or directors' fees.


Data / General Ledger Analysis - Journals Testing / Transactions Data Table / Transaction Details - J811

Transaction #	Risk	Type	Effective Date	Monetary Value	# of Entries	Status
J811	55.4%	JE	Jan 1, 2018	$999,000.00	4	Open

Account ID	Account Name	Effective Date	Memo	Debit	Credit
10600	Bank Account	Jan 1, 2018	reversing special payment per CEO	$999,000.00	$0.00
54227	Management Salaries	Jan 1, 2018	payment stopped Mr C Smith	$0.00	$219,000.00
54228	Director's Fees	Jan 1, 2018	payment stopped Ms F McDonald	$0.00	$450,000.00
54228	Director's Fees	Jan 1, 2018	payment stopped to Mr JM Escubar	$0.00	$330,000.00

Showing 4 entries out of 4

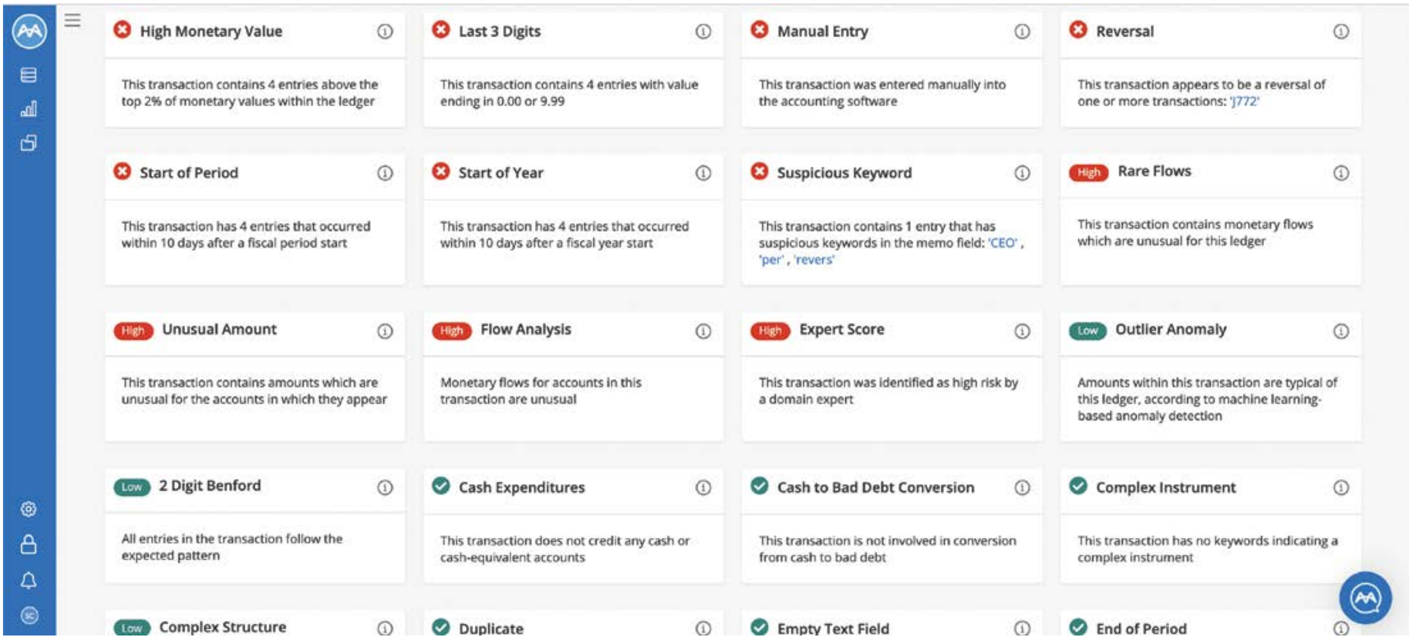
High Unusual Amount: This transaction contains amounts which are unusual for the accounts in which they appear

High Flow Analysis: Monetary flows for accounts in this transaction are unusual

High Expert Score: This transaction was identified as high risk by a domain expert

Low Outlier Anomaly: Amounts within this transaction are typical of this ledger, according to machine learning-based anomaly detection

FIGURE 5.4: List of risk indicators, which explain the conclusions to which Ai Auditor has arrived



High Monetary Value
This transaction contains 4 entries above the top 2% of monetary values within the ledger

Last 3 Digits
This transaction contains 4 entries with value ending in 0.00 or 9.99

Manual Entry
This transaction was entered manually into the accounting software

Reversal
This transaction appears to be a reversal of one or more transactions: 'J772'

Start of Period
This transaction has 4 entries that occurred within 10 days after a fiscal period start

Start of Year
This transaction has 4 entries that occurred within 10 days after a fiscal year start

Suspicious Keyword
This transaction contains 1 entry that has suspicious keywords in the memo field: 'CEO' , 'per' , 'revers'

High **Rare Flows**
This transaction contains monetary flows which are unusual for this ledger

High **Unusual Amount**
This transaction contains amounts which are unusual for the accounts in which they appear

High **Flow Analysis**
Monetary flows for accounts in this transaction are unusual

High **Expert Score**
This transaction was identified as high risk by a domain expert

Low **Outlier Anomaly**
Amounts within this transaction are typical of this ledger, according to machine learning-based anomaly detection

Low **2 Digit Benford**
All entries in the transaction follow the expected pattern

Cash Expenditures
This transaction does not credit any cash or cash-equivalent accounts

Cash to Bad Debt Conversion
This transaction is not involved in conversion from cash to bad debt

Complex Instrument
This transaction has no keywords indicating a complex instrument

Low **Complex Structure**

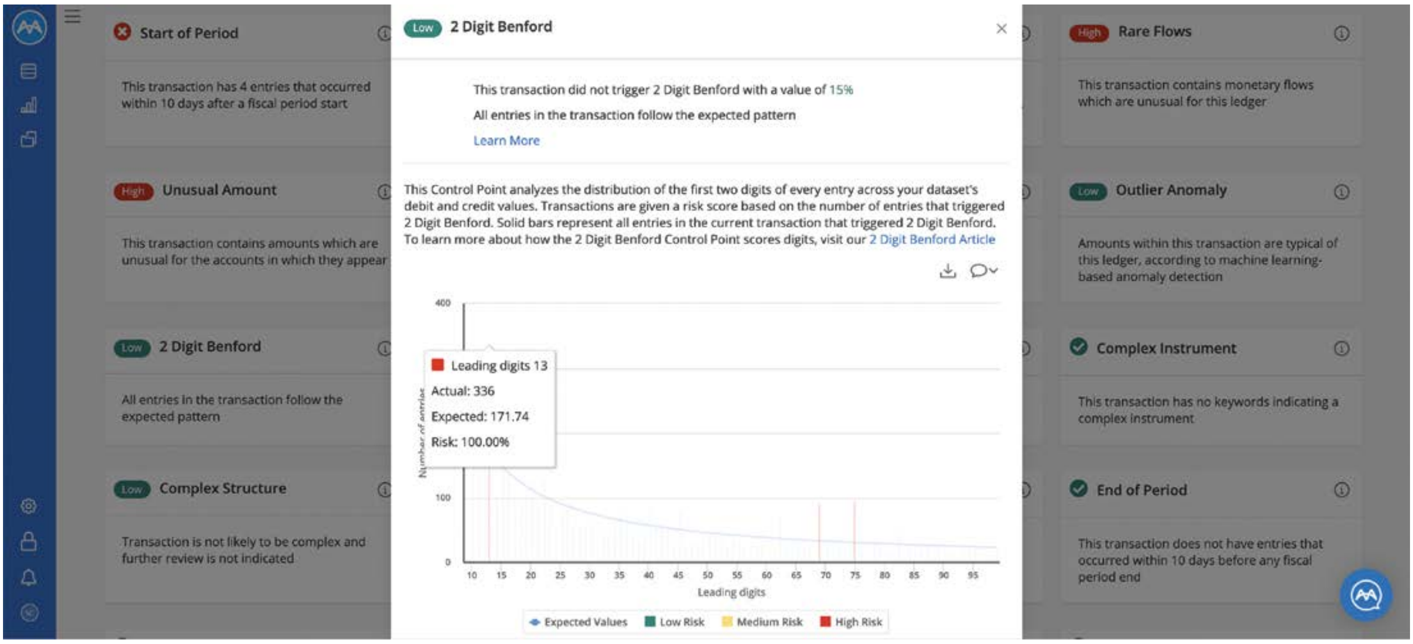
Duplicate

Empty Text Field

End of Period

FIGURE 5.5: MindBridge.ai risk indicator

Below is a view of MindBridge's Benford's Law risk indicator, an example of where Ai Auditor provides further context on where this transaction sits relative to its peers.



Start of Period
This transaction has 4 entries that occurred within 10 days after a fiscal period start

High **Unusual Amount**
This transaction contains amounts which are unusual for the accounts in which they appear

Low **2 Digit Benford**
All entries in the transaction follow the expected pattern

Low **Complex Structure**
Transaction is not likely to be complex and further review is not indicated

Low **2 Digit Benford**

This transaction did not trigger 2 Digit Benford with a value of 15%

All entries in the transaction follow the expected pattern

Learn More

This Control Point analyzes the distribution of the first two digits of every entry across your dataset's debit and credit values. Transactions are given a risk score based on the number of entries that triggered 2 Digit Benford. Solid bars represent all entries in the current transaction that triggered 2 Digit Benford. To learn more about how the 2 Digit Benford Control Point scores digits, visit our 2 Digit Benford Article

🟥 Leading digits 13
Actual: 336
Expected: 171.74
Risk: 100.00%

High **Rare Flows**
This transaction contains monetary flows which are unusual for this ledger

Low **Outlier Anomaly**
Amounts within this transaction are typical of this ledger, according to machine learning-based anomaly detection

Complex Instrument
This transaction has no keywords indicating a complex instrument

End of Period
This transaction does not have entries that occurred within 10 days before any fiscal period end

Jayendran GS, Founder, Prudent.ai

Jay is a chartered accountant and data scientist. He has been working at the intersection of accountancy and analytics for the last 12 years. Before starting his company, he was a director of analytics with EY in India. He is passionate about AI and cloud technology, and using them to build solutions for problems in the accounting and finance world. In his capacity as the founder CEO of Prudent.ai, an AI platform for auditing, he works with leading accountancy professionals around the world.

The accounting industry has been marked out as one of the high-potential areas for AI adoption in multiple articles and surveys. This is mainly because of the manual and repeatable nature of transactions and largely structured data.

On the flipside, there is a significant regulatory oversight and need for compliance to rules, standards and laws to keep in mind. These challenges are quite like those for other expert-driven fields, such as law and medicine. There are also industry and business attributes to be considered in accounting decisions, eg revenue recognition for advertising.

These factors make it necessary to have human-in-loop applications to balance the judgement risks against the efficiency gained from machine driven-decisions.

Higher speed, greater coverage, and lowered costs are the factors that drive adoption of AI audit tools. As finance leaders and auditors come to place higher reliance on these tools, there is a need to provide greater insight into the AI workings along with the output. In this context, key areas of focus include:

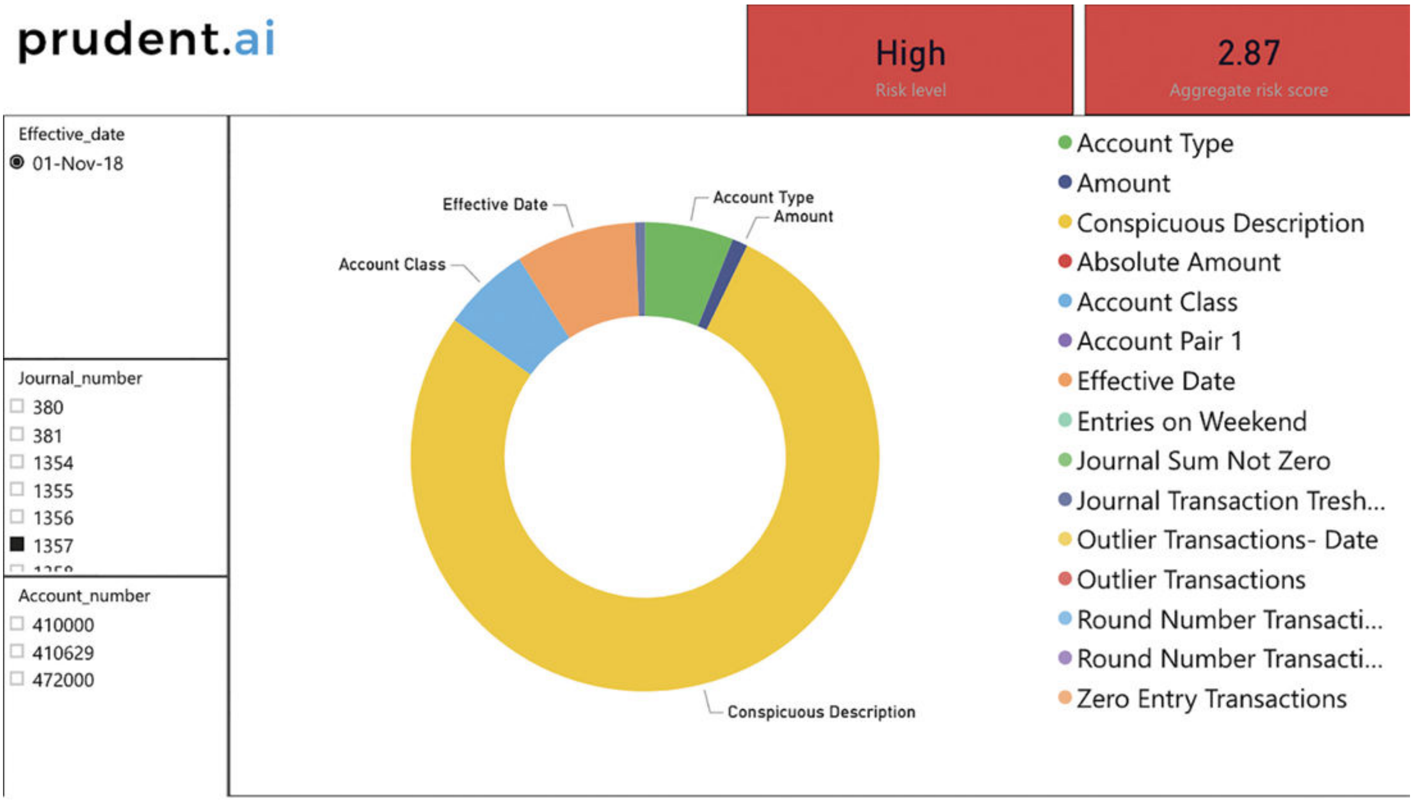
- an overall view of influencing factors and their impact on final risk scores
- transaction-specific explanations to review consistency and build trust in the algorithm
- data points that allow users to provide feedback that can customise and tune the algorithms to match their judgement
- reducing algorithmic false positives to a minimum through iterations.

Prudent.ai uses SHAP techniques (see section 3 above) to provide explainability of its AI models in two respects.

i) Transaction-specific aspects

For flagged transactions, SHAP techniques are used to explain the factors that contributed to the flagging of a risk. For example, parameters for a flagged transaction could be a high transaction value, reversal of a large account balance, or being passed 15 days after book close. This is an important input for the auditor reviewing the flagged transaction and a critical working paper element.

FIGURE 5.6: Transaction specific aspects



prudent.ai

High
Risk level

2.87
Aggregate risk score

Effective_date
● 01-Nov-18

Journal_number
□ 380
□ 381
□ 1354
□ 1355
□ 1356
■ 1357

Account_number
□ 410000
□ 410629
□ 472000

> The need to move from black-box to explainable AI applications is a key factor in improving trust and confidence in AI adoption in the accounting and finance domains.

ii) Overall AI model

Equally important is the general explanation of the AI model for the parameters that contribute to risk flagging, their order of importance and the way they influence (positively or negatively) the risk score of transactions (Figure 5.7). These would help the auditor understand the areas covered by the AI tool and allows them to build audit programmes around it.

Looking beyond the above illustrations, there are various scenarios where explainability could be increasingly important.

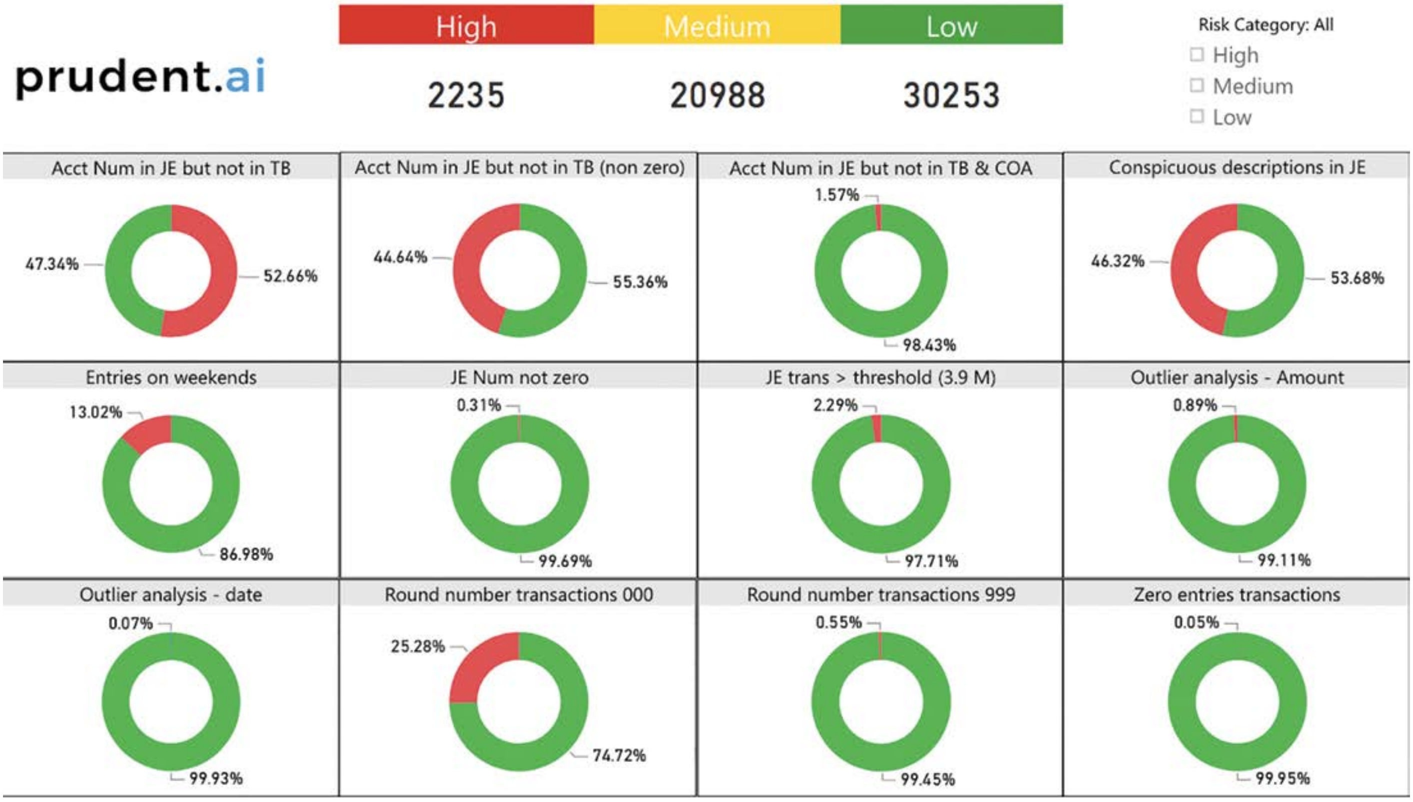
IFRS 9, for example, is an important regulation in the accounting and XAI environment. The expected credit loss model recommended by IFRS 9 will require companies to consider multiple, probability-weighted scenarios and macroeconomic factors.

This not only creates a need for accounting and finance teams to explain credit portfolio valuation models but has also given rise to Model Risk Management as a separate sub-area in the finance world. For the accounting industry this also provides a framework to review AI models used across other business areas.

These business areas could be in any sector or industry. For example, modern hotel aggregator chains use data science models to determine their pricing. The price for any guest at any point in time is determined on the basis of multiple factors by an AI model that takes into account both historical data and external factors such as the prices of other hotels in the area, holiday dates, etc. Accountants who want to use the revenue numbers reliably would require a review of the pricing model with an XAI tool similar to review of IT controls around the system.

The need to move from black-box to explainable AI applications is a key factor in improving trust and confidence in AI adoption in the accounting and finance domains. XAI tools have a critical role to play in building trust and confidence in AI-based applications. These early stage applications will get better over time to improve the value they bring, and resilience against adversarial attacks.

FIGURE 5.7: Assessing risk categories



Opinion piece by Johnnie Ball, chief data officer, Fluidly

Fluidly has been named one of 2019's European Fintech50 companies, was listed in WIRED's Top 100 Hottest European start-ups, and won 'Innovation of the Year' and 'Forecasting, Planning & Analysis Software of the Year' in the Accounting Excellence awards. In June 2019, it was awarded £5m as part of the Royal Bank of Scotland (RBS) Alternative Remedies Package, alongside four other UK FinTech companies.

Fluidly is building a new software category – Intelligent Cash-flow – to automate the forecasting and management of finances for businesses using AI. Fluidly uses cloud accounting, bank transaction data and credit data to predict and optimise the financial future for SMEs.

AI-powered approaches are as accurate as, and far more efficient than, traditional methods, but can be more opaque. As the industry increasingly relies upon more sophisticated prediction systems it is important that finance professionals can interpret the results of these tools. We will require a new perspective on professional scrutiny – the core of which will be the explainability of AI.

The Fluidly Intelligent Cashflow platform features an automated cashflow forecast which produces account line, customer and supplier level predictions for 12-months into the future. To do this, the system ingests vast quantities of historic transaction data and searches for meaningful patterns.

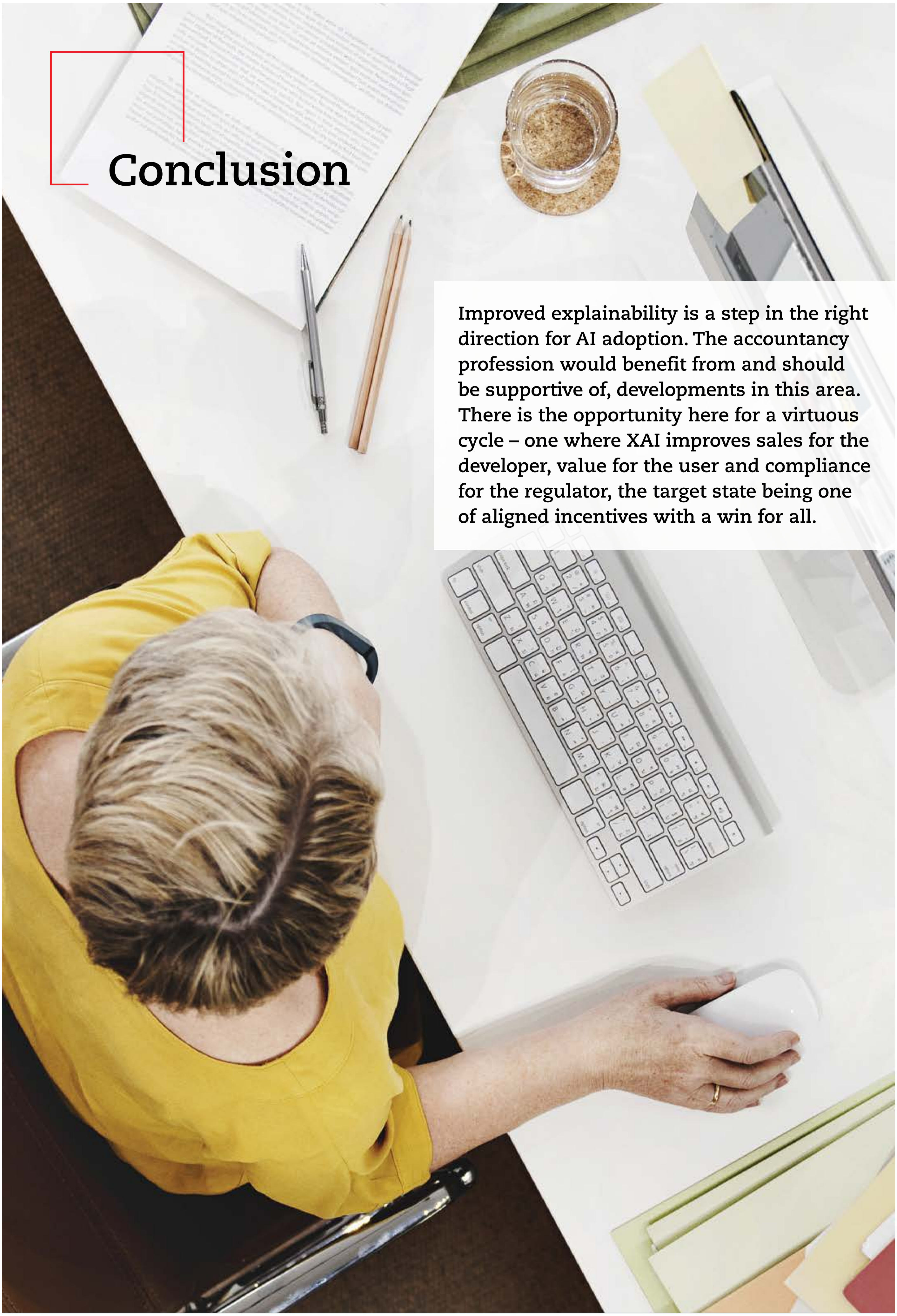
Many types of models are applied to the identified patterns and the one with the optimum prediction is chosen to forecast that particular subset of data. Aggregation across thousands of these predictions gives a baseline bank balance forecast in seconds, which can then be adjusted and refined by the user.

Unlike conventional forecasts, which tend to use averages and simplified trending assumptions, the chosen model may be an algorithm or statistical method that is unfamiliar to the user and therefore has output that is difficult to explain. Increasing forecast accuracy using more sophisticated models comes at the cost of easy interpretability – but both are critical user requirements.

For a software product such as Fluidly, an appropriate explanation to users requires an automatic translation of the importance measure to a representation that is understandable to humans – across multiple models and thousands of inputs, this is extremely challenging.

Fluidly takes inspiration from techniques such as LIME and actively researches explainable AI while emphasising the balance between academic research and user experience. It is important for the company to establish trust in its algorithms, and one way of achieving this is through explanations of the factors used in making predictions.

Fluidly also increasingly guides users' attention to uncertainty in the AI forecast – which is a baseline prediction but should be supplemented by human intelligence. Encouraging and augmenting professional scepticism of the AI system, Fluidly believes, will lead to the most efficient, powerful and accurate forecasts. ■



Conclusion

Improved explainability is a step in the right direction for AI adoption. The accountancy profession would benefit from and should be supportive of, developments in this area. There is the opportunity here for a virtuous cycle – one where XAI improves sales for the developer, value for the user and compliance for the regulator, the target state being one of aligned incentives with a win for all.

References

ACCA (2019), *Machine Learning: More Science than Fiction* <https://www.accaglobal.com/sg/en/member/discover/events/global/e-learning/digital-technology/machine-learning.html>, accessed 22 January 2020.

ACCA (2017), *Banishing Bias: Audit, Objectivity and the Value of Professional Scepticism* <https://www.accaglobal.com/content/dam/ACCA_Global/Technical/audit/pi-banishing-bias-prof-scepticism.pdf>, accessed 22 January 2020.

PI-EMTECH-XAI

ACCA The Adelphi 1/11 John Adam Street London WC2N 6AU United Kingdom / +44 (0)20 7059 5000 / www.accaglobal.com