Deloitte.

CENTRE *for* REGULATORY STRATEGY EMEA

with Swiss Foreword

# Building trustworthy AI

**A comprehensive approach to conduct, data protection and ethics**

175 MAKING AN IMPACT THAT MATTERS since 1845

# Swiss Foreword

The Swiss regulatory approach to conduct and data protection requirements in the context of AI is comparable to those in the UK and the EU as outlined in this report. In short term, the Swiss approach is "technology-neutral" and does not see an immediate need for regulatory action. AI should thus be regulated under the existing Swiss regulatory framework allowing space and opportunity for significant market innovation. Apart from the risks outlined in this report, AI at the same time offers vast opportunities to transform value creation for Swiss market participants.

This section provides a broad overview of the Swiss point of view regarding the potential risks arising from the use of AI by financial services providers. Particular attention should be drawn to the following risks*:

Data Protection: Personal data could be misused by AI systems, resulting in the violation of fundamental personal rights of individuals, in particular in the insurance business.

Conduct: Undetected data manipulations or algorithm errors/decisions could result in breaches of client conduct rules and organisational professional conduct requirements as well as in incremental operational risk.

Fraud: AI systems could be manipulated and trigger fraudulent reporting (financial or regulatory related) or eventually misappropriation of assets.

## Data Protection

The Swiss Federal Data Protection and Information Officer demands that AI, in addition to its primary purposes, must always protect the freedom of choice and the right to privacy of individuals.

A conflict between the protection of personal data and the data needs of AI systems could arise.

***Example:***
*In future, insurance premiums could be made even more dependent on various client data factors (such as food and book purchases, use of the Internet, travel behaviour).*

## Conduct

The Swiss Financial Services Act (FinSA) seeks to protect the client of financial services at the point of sale.

With the use of inaccurate data by AI, Swiss financial services providers could potentially breach behavioural, responsibility, information or accountability duties under the FinSA.

***Example:***
*Due to the use of manipulated and thus inaccurate data by a Robo Advisor, an independent asset manager could fail to carry out the required legal suitability and appropriateness tests for clients accordingly.*

## Fraud

AI consists of algorithms which are typically complex and not very transparent.

It opens up the possibility for misuses by third parties and developers to cause undesired actions of the code. As a consequence, fraudulent outcomes can occur.

***Example:***
*Banks could be defrauded by third parties or developers concealing functionalities in the code, resulting in the granting of unjustified credits or in the execution of illegitimate bank transfers.*

*According to a report published by the Swiss Federal Council ("Challenges of Artificial Intelligence"), December 2019, p. 79-80.

# Executive summary

## Overview

Customers are increasingly interacting with financial services (FS) firms through digital channels. Reduced human interaction requires firms to use AI and data analytics to understand and serve customer needs better. However, the combination of a digital channel and use of AI presents risks and opportunities, particularly from a governance and compliance perspective.

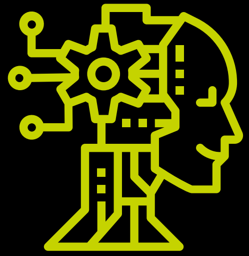
Take the case of detecting and supporting customer vulnerability in a digital journey. Without AI and data analytics, it is incredibly hard to detect patterns of vulnerable behaviour and therefore provide timely support. However, the use of AI requires careful consideration - data protection, conduct requirements, and robust review and challenge of customer outcomes are all essential to the safe and successful application of AI.

In addition, there is greater social pressure on firms to serve a purpose beyond pure commercial gain. This brings a third dimension to the use of AI and consumer data - the ethical use of data.

*In this report, we explore the alignment and potential regulatory uncertainty between conduct and privacy regulatory requirements and set out how ethics interacts with regulation and informs difficult judgment decisions and trade-offs when using AI-enabled solutions. We bring this to life through an illustrative case study - identifying and supporting vulnerability in a digital banking journey - and highlight what firms need to do to build trustworthy AI solutions. We conclude by making the case for further regulatory guidance to remove uncertainty to allow firms to innovate with confidence. Our analysis and exploration of these issues are designed to inform and support boards, senior management and digital leads who are responsible for AI-enabled solutions as they navigate their way through these complex issues.*

This report builds on our previous paper on AI and Risk management, where we explored the dynamic nature of AI models and the resulting risk management implications. We have not repeated the key elements discussed in the previous paper here, but they continue to be relevant.

While this report draws on UK regulations, the challenges and solutions proposed for firms will be relevant to other jurisdictions, especially in the EU.



What do we mean by AI and AI systems?

There is no consensus on a definition of AI. For the purposes of this paper, by AI and AI systems, we mean the theory and development of computer systems able to perform human tasks that normally require human intelligence. This is done by using various techniques such as machine learning, deep learning and natural language processing. Please see our paper on AI and risk management for a detailed description.

# Building trustworthy AI: key takeaways

### COMPREHENSIVE AND INTEGRATED APPROACH TO CONDUCT, DATA PROTECTION AND ETHICS

- AI conduct regulation and data protection requirements will intersect significantly across several areas.
- Some requirements will be aligned or complementary (e.g. in relation to transparency and explainability). Others might require assessment and case-specific interpretation from the design phase of the AI solution - e.g. GDPR lawful basis.
- A comprehensive and integrated approach to regulation and ethics is necessary to ensure good customer outcomes, compliance, and operational efficiency.

### AN ETHICAL FRAMEWORK BUILT ON COMPLIANCE BUT REFLECTING BROADER SOCIAL PURPOSE

- Strong ethical frameworks are necessary to identify, assess and choose the right course of action in relation to risks, opportunities and moral issues raised by the use of AI.
- They must be built on a solid foundation of regulatory compliance, but their purpose is to guide organisations where the current rules are silent or subject to interpretation - e.g. definition of fairness, or trade-offs between individuals' privacy and AI accuracy.

### DIVERSITY IN REVIEW, CHALLENGE AND DESIGN INPUT

- Skills and knowledge across boards, compliance, and AI design teams of the workings of case-specific AI models are hugely important to review, challenge, and interpret regulation in the spirit of the law, apply ethical judgements and understand customer outcomes.
- Testing AI systems with a diverse set of focus groups and stakeholders helps ensure they are fit for purpose and fulfil society's ethical expectations. This is important as society's consensus about what is acceptable in relation to AI continues to evolve and differs across countries/use cases.
- Internally, choosing the right course of action will require firms to nurture an open culture, with diversity of thought and perspectives.

### FIRM LEVEL CAPABILITIES TAILORED TO SUPPORT TRUSTWORTHY AI

- Risk appetite, governance and risk management need to be updated to enable the firm to innovate using AI. The level of risk management and governance will be context-specific and proportionate to the risk posed by the deployment of the AI solution. These boundaries need to be set and understood clearly throughout the firm.
- A key capability is the development of AI skills and training across control functions, boards and senior management to enable various stakeholders to ask the right questions, interpret the ethical and compliance requirements and understand the inherent risks and put the right mitigants in place, including a human-in-the-loop at the right points of the AI solution and related process flows.

### PROACTIVE REGULATORY ENGAGEMENT

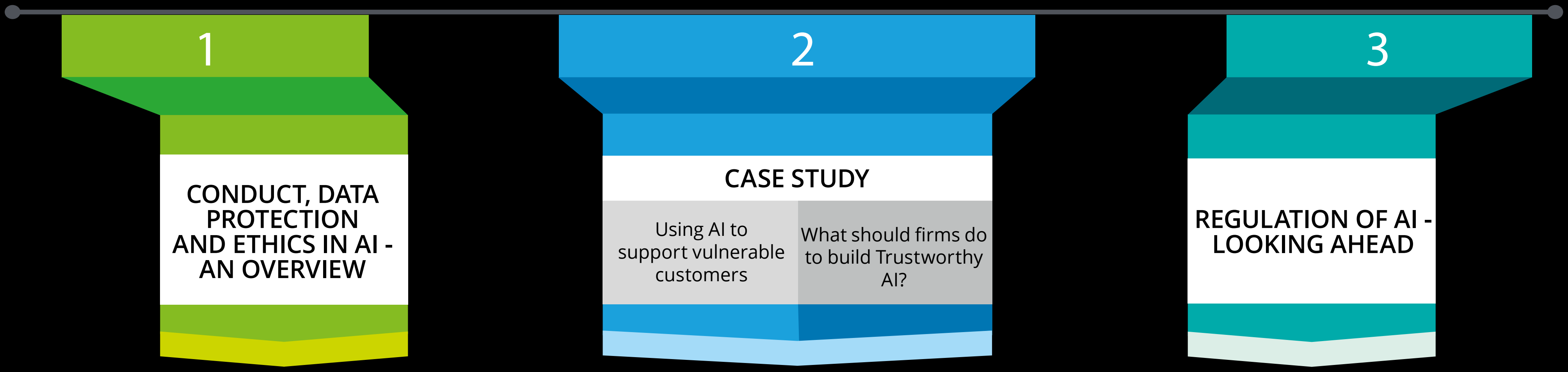
- Engaging early and proactively with data protection and conduct authorities can help resolve context-specific compliance challenges.
- One way to do this is through the use of regulators' innovation hubs, including direct advice teams and sandboxes.

### NEED FOR MORE REGULATORY GUIDANCE

- Our case study highlights that intepreting and complying with conduct requirements on the one hand, and data protection regulation on the other, can present important regulatory implementation challenges or uncertainty for firms.
- Conduct and data protection authorities should further support firms wishing to deploy AI applications aligned to significant public interest (e.g. vulnerable customer support), through co-ordinated guidance on areas of known regulatory uncertainty.

# The structure of our report

**This report has three key sections:**



1

CONDUCT, DATA PROTECTION AND ETHICS IN AI - AN OVERVIEW

2

CASE STUDY

Using AI to support vulnerable customers

What should firms do to build Trustworthy AI?

3

REGULATION OF AI - LOOKING AHEAD

- This section gives an overview of the relationship between conduct and data protection regulatory objectives and key areas of focus in relation to the use of AI.
- We highlight how the two regulatory frameworks are aligned or complementary, and where practical implementation challenges, including uncertainty around interpretation, may arise.
- We explore the ethical use of AI, how this sits alongside regulatory compliance, and why having strong ethical frameworks is critical to developing Trustworthy AI.

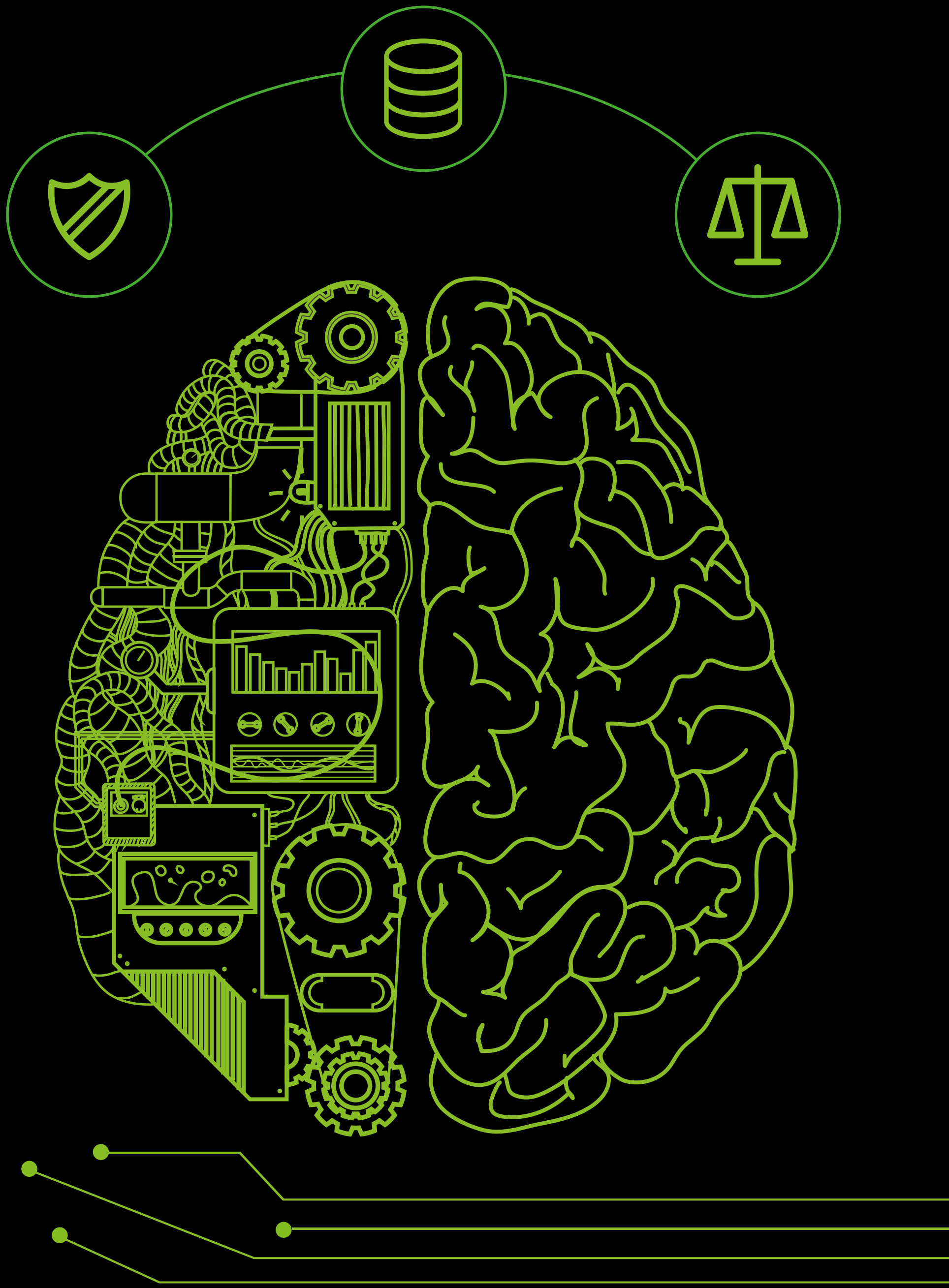
- This section brings our general considerations to life through a case study, to illustrate the important point that conduct and data protection requirements, as well as ethical issues, are context and use-case specific.
- The case study takes a closer look at the interaction between conduct, data protection and ethics in relation to an AI system designed to help vulnerability detection in UK retail banking customers.

**Click here for a selection of capabilities that firms need to deploy AI-enabled solutions safely to support vulnerable customers**

- This section highlights the developing nature of the AI regulatory landscape.
- We identify the areas where further regulatory collaboration and guidance would help firms to interpret the existing regulations to support their innovation journeys.

# THE RELATIONSHIP BETWEEN CONDUCT, DATA PROTECTION AND ETHICS IN AN AI CONTEXT

> "Trustworthiness demands two things: knowledge and skill; and good intentions and honesty."
>
> Andrew Bailey, then CEO of the FCA, 2018



# Conduct and data protection regulation: objectives, focus, and interaction

## UK Financial Conduct Authority (FCA) vs. the Information Commissioner's Office (ICO)

- The FCA and the ICO have a common objective: to protect individuals and consumers from harm.
- However, because of differing statutory mandates, they maintain two independent perspectives of what constitutes harm and what firms should do to prevent it.

### FCA

The FCA is focussed on ensuring that consumers can continue to access fair value financial products and services, suitable to their needs, without unreasonable barriers.

### ICO

The ICO is focussed on protecting individuals' information rights under the General Data Protection Regulation (GDPR) and ensuring firms use personal data lawfully, fairly and responsibly.

## Interaction between conduct and data protection in AI context

- The FCA and ICO's objectives and perspectives apply regardless of whether or not FS firms use AI to deliver their products and services.
- However, in the case of AI applications which use personal data and have a direct impact on customer outcomes, **data protection and conduct requirements will interact significantly and should be considered together** from the outset.
- **In a number of areas conduct and data protection requirements will be closely aligned and/or complementary** (e.g. explainability and transparency). Considering these requirements together will improve customer outcomes, and reduce duplication of effort across conduct, data protection, and AI/data science teams.
- **In some other areas there may be some regulatory uncertainty about how to implement data protection requirements, while fulfilling conduct expectations** (e.g. GDPR lawful basis). Firms must understand and resolve this uncertainty from the outset of any AI adoption programme, to avoid harming customers and/or being in breach of GDPR.
- We give an overview of **six key areas of alignment and regulatory uncertainty** on the next page.

**A firm's ethical framework, underpinned by its broader vision and social purpose, will play a significant role in informing how regulation is translated into policies. It will also influence the boundaries of how AI-driven results will be used to inform customer choice, access to products and services, pricing and customer support.**

# Regulation: areas of alignment and practical implementation challenges

While conduct and data protection requirements are context and use-case specific, the six key areas of regulatory focus set out below will be relevant whenever AI is used to profile customers, and make or support decisions about them.

We explore each of these areas in more detail as part of our case study.

## Areas of regulatory alignment or complementarity

Both the FCA and ICO require firms to:

**Fairness, bias and discrimination**

Ensure that the output of an AI system is not unfairly biased or discriminatory based on the datasets or methods used to build it.

Ensure customers/individuals are treated fairly, with a level of care appropriate to their capabilities and reasonable expectations.

**Explainability and transparency**

Ensure that they are transparent about their use of AI, and that their AI models and processes are sufficiently explainable to allow internal and external stakeholders, including individuals affected, to understand, challenge, and trust their outputs.

**Automated decision-making**

Minimise consumer harm, and fulfil more stringent regulatory requirements (e.g. in relation to consent or transparency), in relation to fully automated decisions, i.e. made by a system without meaningful human involvement.

## Areas of regulatory implementation challenges

Firms can face uncertainty in complying with specific data protection requirements which do not have a parallel or complementary conduct regulation. The extent of the challenge will depend on the specific use case. In some instances, such as in our case study on vulnerability detection, further co-ordinated guidance from the ICO and FCA may be necessary for firms to innovate with confidence. The ICO requires firms to:

**Lawful basis for processing personal data**

Have a valid lawful basis under GDPR in order to process personal data. Determining the most appropriate basis (e.g. contract), including how broadly it can be interpreted, is not always straightforward.

**Purpose limitation**

Only collect and use personal data for specified and legitimate purposes. Further processing for reasons incompatible with the original purposes is not allowed. This restricts a firm's ability to re-use data that it has already collected for secondary purposes.

**Data minimisation**

Only process data that is adequate, relevant, and necessary to achieve the stated purpose of the processing. These terms are not defined in law. Firms must determine what they mean, and how broadly they can be interpreted, in the context of each use case.

# AI ethics: going beyond compliance to develop Trustworthy AI

*AI ethics is "a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies."* [1]

> *At a basic level, firms using this technology must keep one key question in mind, not just 'is this legal?' but 'is this morally right?'*
>
> Christopher Woolard, Interim CEO of the FCA, July 2019

Trustworthy AI means that a firm's use of the technology is ethical as well as lawful. Yet, the relationship between regulation and ethics is complex. In the words of the late Giovanni Buttarelli, former European Data Protection Supervisor; *"ethics comes before, during and after the law. It informs how laws are drafted, interpreted and revised. It fills the gaps where the law appears to be silent. Ethics is the basis for challenging laws"*.

## What have regulators already said?

- EU and UK regulators, and the FCA and ICO specifically, acknowledge that the increasing use of AI raises significant ethical questions. While these questions are often broader than data protection and conduct regulation, they frequently overlap and challenge the existing regulatory framework or its interpretation.
- Both the FCA and ICO are already leading contributors to the AI ethics debate and are reviewing their regulatory and supervisory approaches to determine if and how they need to change to support ethical AI innovation.
- However, they have also reiterated that many of the core regulatory principles and requirements already in place are directly applicable and fully aligned to the development of ethical AI. As part of their supervisory work, we expect them to focus on three issues:

GOVERNANCE AND ACCOUNTABILITY

CUSTOMER-CENTRIC CULTURE

TRANSPARENCY AND EXPLAINABILITY

## When do we need AI ethics to complement regulation?

It is likely that AI innovation will continue to outpace the policy making process and challenge the existing regulatory framework and perimeter. Therefore, ethical frameworks will be increasingly crucial to guide firms' behaviours and choices when:

| | |
|---|---|
| LAW IS SILENT OR UNCLEAR | If laws are silent or subject to interpretation, firms will have to decide what role to play to foster ethical AI. For example, taking proactive action (e.g. through positive discrimination) when developing AI applications to correct historical racial bias in society. |
| NAVIGATING REGULATORY TRADE-OFFS | AI systems will often require nuanced trade-offs between regulatory principles or requirements (e.g. between model accuracy and data privacy). The ethical way to approach these decisions, will depend on specific use cases. |
| COMPLIANT ≠ ETHICAL | Not all actions that comply with conduct or data protection regulation are necessarily ethical. For example, the use of automated decision-making may be technically legal in some circumstances, but it may be deemed unethical by customers and society. |

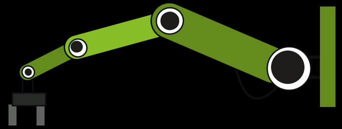
[1] "Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute."

# Case Study

> “Firms are increasingly using digital communication channels and these can be both a benefit and a barrier to vulnerable customers. With less direct contact with customers, digital channels can sometimes make it harder to pick up on indicators of vulnerability. However, firms can mitigate this risk, for example, by making it easy for consumers to disclose their needs on online platforms, or by using data analytics or software to identify indicators of vulnerability.”

Nisha Arora, Director, Consumer and Retail Policy, FCA, March 2020

# AI and vulnerable customers - regulatory context



A vulnerable customer is defined as **"someone who, due to their personal circumstances, is especially susceptible to detriment, particularly when a firm is not acting with appropriate levels of care".**

*The FCA's definition of vulnerability is very broad and includes temporary or longer-term vulnerability and a number of drivers that may contribute to that situation. See Appendix for further details.*

## Using data and AI to help vulnerable customers

- FCA commissioned research highlights that around 50% of UK customers are vulnerable.[2]
- The FCA expects firms to identify and support vulnerable consumers proactively, including understanding drivers and risks of vulnerability across groups of target customers.
- The FCA also recognises that data and advanced analytics will be important for digital customer journeys to identify characteristics of vulnerability in individual customers and offer them proactive support.
- Using AI in vulnerability detection could deliver significant benefits (Figure 1), but FS firms have so far been reluctant to deploy it in this highly sensitive area. A key deterrent is the uncertainty around how firms can comply with data protection requirements.
- AI is not a silver bullet. It should only form part of a firm's approach to vulnerability. AI design requires a mature understanding of vulnerability drivers and customer needs in target markets, as well as the IT infrastructure and data to support it. As such, it may not be suitable for all firms.

### Figure 1: Key benefits of an AI-enabled approach to vulnerability

1 **Improved and ongoing view of vulnerability drivers across customer cohorts**

2 **Integration with both traditional and digital journeys to enable real-time intervention**

3 **Increased oppurtunity for preventative intervention and support**

4 **Can reduce human bias and enable a more consistent approach to vunerability detection and support**

5 **Performance feedback loops can be used for ongoing oversight and improvement of vulnerability detection**

6 **Augmented, not replaced, human decision-making**

[2] 2, FCA, Findings from the FCA's financial Lives Survey 2017, October 2017, available at https://www.fca.org.uk/publication/research/financial-lives-survey-2017.pdf

# Introduction to our case study: using AI to detect and respond to vulnerability

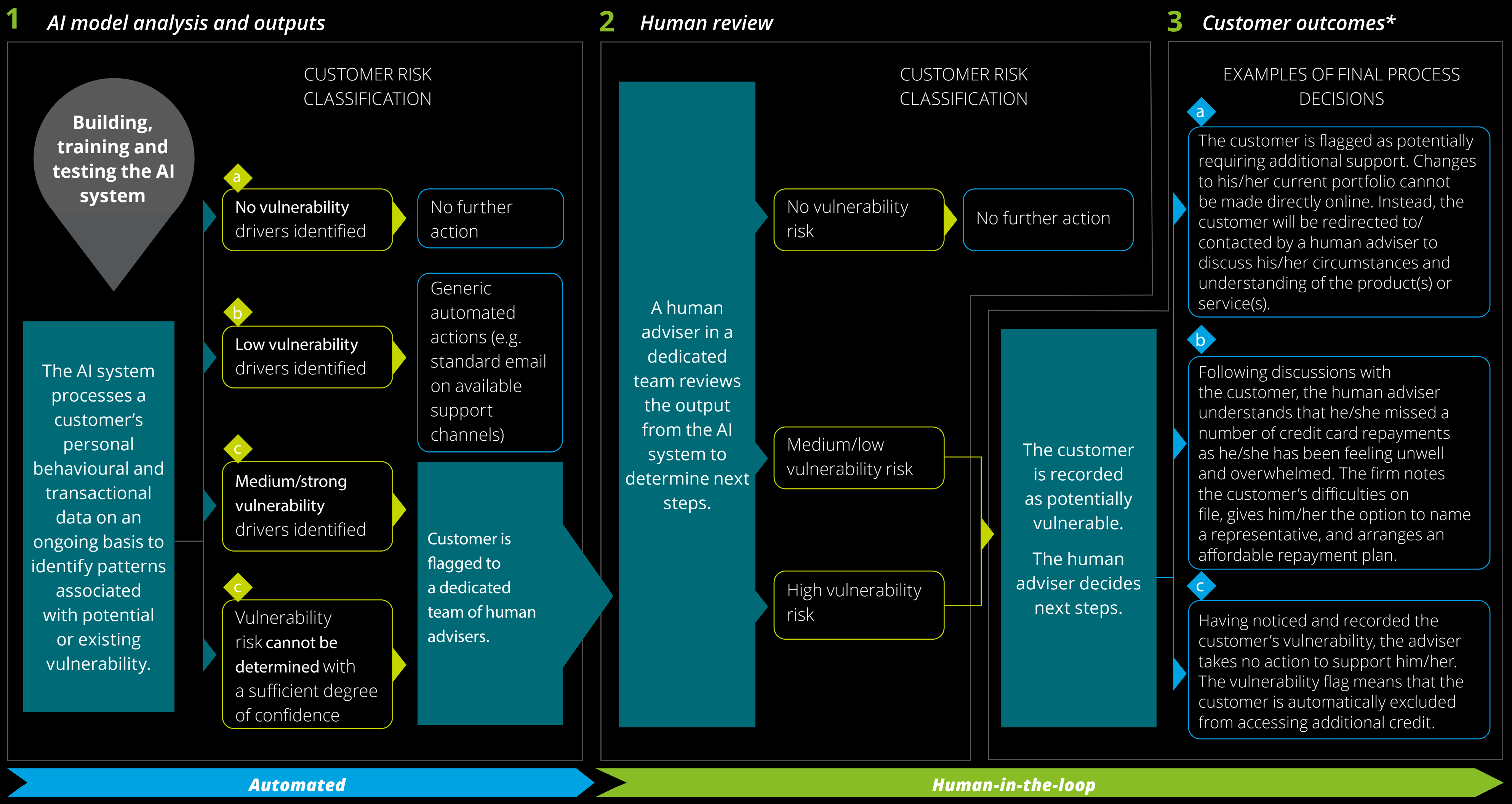
## Our case study brings the interaction between different pieces of regulation, ethics and AI to life.

- A retail bank has a digital channel to interact with customers.
- It has an AI-enabled process to identify characteristics of vulnerability in individual customers and deploy preventative intervention to avoid customer detriment.
- The AI system will analyse customers' transactional and behavioural data to spot patterns that are typically associated with vulnerability across profiles and product portfolios.
- The AI system is intended to support and augment human decisions. Once identified as vulnerable, customers are automatically directed to a dedicated team of human reviewers who will determine the appropriate response for each customer.

*Note - The AI vulnerability monitoring process we describe is a simplified illustrative example to highlight key areas of risks.*

# An illustrative case study: AI to support vulnerable customers

**The diagram below sets out a simplified AI-enabled vulnerability detection process for a bank. We will use this process flow to highlight the areas of interaction between conduct and data protection compliance, and where ethical issues may arise.**



## 1 *AI model analysis and outputs*

CUSTOMER RISK CLASSIFICATION

**Building, training and testing the AI system**

The AI system processes a customer's personal behavioural and transactional data on an ongoing basis to identify patterns associated with potential or existing vulnerability.

a No vulnerability drivers identified → No further action

b Low vulnerability drivers identified → Generic automated actions (e.g. standard email on available support channels)

c Medium/strong vulnerability drivers identified

c Vulnerability risk cannot be determined with a sufficient degree of confidence

→ Customer is flagged to a dedicated team of human advisers.

## 2 *Human review*

CUSTOMER RISK CLASSIFICATION

A human adviser in a dedicated team reviews the output from the AI system to determine next steps.

No vulnerability risk → No further action

Medium/low vulnerability risk

High vulnerability risk

→ The customer is recorded as potentially vulnerable.

The human adviser decides next steps.

## 3 *Customer outcomes**

EXAMPLES OF FINAL PROCESS DECISIONS

a The customer is flagged as potentially requiring additional support. Changes to his/her current portfolio cannot be made directly online. Instead, the customer will be redirected to/contacted by a human adviser to discuss his/her circumstances and understanding of the product(s) or service(s).

b Following discussions with the customer, the human adviser understands that he/she missed a number of credit card repayments as he/she has been feeling unwell and overwhelmed. The firm notes the customer's difficulties on file, gives him/her the option to name a representative, and arranges an affordable repayment plan.

c Having noticed and recorded the customer's vulnerability, the adviser takes no action to support him/her. The vulnerability flag means that the customer is automatically excluded from accessing additional credit.

***Automated***

***Human-in-the-loop***

* Note: outcomes 3a and 3b are likely to be acceptable from a conduct regulation perspective, while outcome 3c is an example of bad practice.

# Conduct and data protection: key areas of regulatory alignment or uncertainty
# A closer look

**Click on each box** to explore in more detail how data protection, conduct requirements, and AI ethics interact in the context of this particular use case.

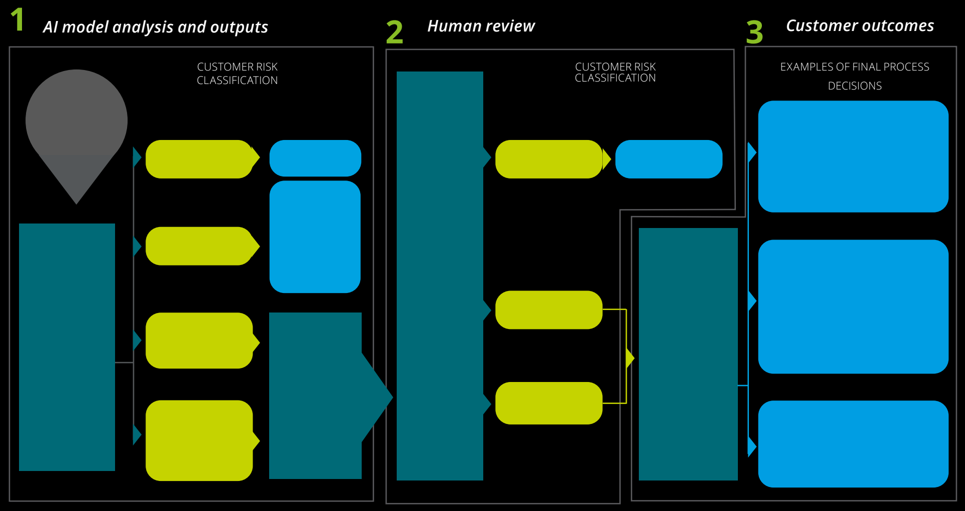
**Areas of regulatory alignment or complementarity**

**Areas of regulatory implementation challenges**

Fairness and transparency

Explainability of AI decisions

Automated decision-making


1 AI model analysis and outputs
CUSTOMER RISK CLASSIFICATION
2 Human review
CUSTOMER RISK CLASSIFICATION
3 Customer outcomes
EXAMPLES OF FINAL PROCESS DECISIONS

Lawful basis for processing personal data

Purpose limitation

Data minimisation

How can firms respond?

# Fairness and transparency

- The ICO and FCA's fairness requirements complement each other, especially across the three areas set out below: bias and discrimination, customer outcomes and rights, and transparency.
- While regulation sets the wider boundaries, firms will need to apply an AI ethical framework to interpret and fulfil supervisory, customer, and social expectations of what is fair for each specific AI use case.

| | Overview | Case study example |
|---|---|---|
| **BIAS AND DISCRIMINATION** | • Use of AI should not lead to unfair adverse discrimination, based on factors that are irrelevant to the risk being assessed (FCA) or protected characteristics such as gender or race (ICO). | • The bank should ensure that customers are not under or over identified as vulnerable based on, for example, postcode or gender. The bank should also reject vulnerability classifications based on proxy variables for protected characteristics - e.g. occupation as a proxy for gender. See "Appendix 1" for more information about identifying unintended bias. |
| **CUSTOMER OUTCOMES AND RIGHTS** | • The ICO and FCA expect firms to ensure their behaviour is aligned to customers' reasonable expectations and protects their customers' interests adequately.<br>• Both regulators also require firms to ensure that their AI systems and processes uphold customers' statutory rights. | • From a conduct perspective, the bank must ensure customers understand its actions, and that these are appropriate to vulnerable customers' needs - e.g. help consumers understand a product, rather than automatically excluding them from it.<br>• From a data protection perspective, the bank must inform customers about its use of personal data to detect vulnerability, and the effect it may have on them. If applicable, the bank must also support their right to object to challenge the profiling. |
| **TRANSPARENCY** | • The ICO and FCA require firms to be transparent about their use of AI, and ensure their AI systems are sufficiently explainable to support accountability, compliance, and regulatory scrutiny. | • The bank must be transparent about its use of AI for vulnerability detection, especially with its customers.<br>• The bank must be able to provide customers, the FCA and ICO, and its senior managers with the information they need to determine whether the vulnerability detection process is lawful and fair. See Explainability of AI decisions for more details. |

## AI ETHICS SPOTLIGHT
## DEFINING FAIRNESS

- Defining what treating customers fairly means in practice in an AI context will require a significant degree of ethical judgement.
- A firm's definition of fairness will need to fulfil customers' and regulators' expectations, while balancing its legitimate commercial interests.
- A firm's fairness policy will need to be aligned with its ethical framework, and articulated in a way that can be coded into the AI system, as well as tested and monitored.

For example, the bank in our case study will need to make a number of ethical choices that will determine the "fairness" of its AI system. Such choices include:

- Defining the types of vulnerability and the vulnerability drivers the AI systems should monitor. These factors are subjective, specific to the bank's customer base and products, and may change over time.
- Determining the acceptable level of false positives (people wrongly identified as vulnerable), and false negatives (people wrongly identified as non-vulnerable) in the AI system's outputs.
- Deciding how and how often to consult with customers to understand their reasonable, and changing, expectations about the use of their personal data for automated profiling.

Executive Summary Contents Case study

# Explainability of AI decisions

- A firm's use of AI cannot be fair or transparent unless it can explain how AI decisions are made to its customers, oversight functions, senior managers, and supervisors. In particular, senior management accountability and customer explanations are two areas of regulatory focus for both the FCA and ICO.
- The type and granularity of information firms should provide will depend on the use case, stakeholder type, and stakeholders' interests in obtaining an explanation. Firms should go beyond basic compliance and focus on building trust in their use of AI. This will strengthen their brand and customers' loyalty.

| | Overview | Case study example |
|---|---|---|
| SENIOR MANAGEMENT ACCOUNTABILITY | • ICO and FCA require senior managers to take responsibility for their firm's treatment of customers and compliance obligations.<br>• This means senior managers are accountable for understanding and governing the way AI models and processes work, their limitations, and how they could harm customers. | • The bank must identify an owner for its AI vulnerability detection system who will be responsible for reviewing and signing-off the AI system's' explainability requirements for different recipients (e.g. customers vs. "humans-in-the-loop").<br>• The owner will be responsible for ensuring that the system can explain, for example, the weight that different data features have in the risk classification for different types of vulnerabilities or customer groups. |
| CUSTOMERS' EXPLANATIONS | • ICO and FCA require firms to give customers a meaningful explanation about how and why a decision about them was made.<br>• Explanations should provide customers with enough information to exercise their rights to challenge a decision or seek recourse, where applicable. | • The bank must explain clearly to each customer how and why they were identified as vulnerable, what elements of the process were automated, and what the implications for them are.<br>• The explanation's content and delivery should suit customers' capabilities and cover both conduct and data protection perspectives.<br>• The bank should not overwhelm customers with information, but give them the tools to understand, and, where it is their right to do so, challenge the bank's actions and decisions. |

## AI ETHICS SPOTLIGHT
## EXPLAINABILITY TRADE-OFFs

- Explainability is fundamental in building society's trust in AI, but it can involve trade-offs.
- For example, in some cases prioritising explainability may require firms to adopt simpler AI models, which could have an adverse impact on the accuracy of outputs.
- In other cases, explanations may reveal a firm's commercial approach, or increase the risk of customers "gaming" the system.
- Ethical frameworks can help firms to determine their position with respect to such trade-offs.

- For example, in our case study, the AI system works by finding correlations between customers' personal data (e.g. patterns in financial transactions or voice sentiment analysis) and known drivers of vulnerability.
- Depending on the type and range of vulnerabilities and data in scope, the system may become too complex to explain to human reviewers or customers, especially if the latter are vulnerable.
- Reducing the complexity of the system may increase explainability, but may also reduce the number of vulnerable customers identified correctly.
- A fit for purpose ethical framework should help enable the bank to determine the right balance between explainability and statistical accuracy.

# Automated decision-making

- The role of automated decision-making, and the safeguards that firms must put in place to protect customers from potential harms, are increasingly hot topics in regulatory policy, as well as society more broadly.
- GDPR has very specific and stringent requirements in this area. While the FCA does not have specific rules, in our experience its supervisors are increasingly questioning firms' ability to ensure fair treatment of customers in a digitised setting, and especially when using automated processes with no or limited human intervention.

| | Overview | Case study example |
|---|---|---|
| **FULLY AUTOMATED DECISIONS WITH SIGNIFICANT EFFECT ON CUSTOMERS** | • Both ICO and FCA require firms to remain accountable for any fully-automated decision that have a significant effect on customers.<br>• This includes ensuring that customers are treated lawfully and fairly, and that potential harms are mitigated appropriately. | • Although the AI system is intended to augment human-decisions only, in practice it still involves fully automated decision-making for those customers classified as displaying no/low vulnerability drivers.<br>• The bank must mitigate the risk that wrongly classified customers (false negatives) may not receive the support they need.<br>• The bank should ensure the system minimises false negatives, that customers can easily ask for support through a range of channels (email/app/phone), and that front line staff are trained to identify possible vulnerabilities in their interactions with customers. |
| **THE ROLE OF THE "HUMAN-IN-THE-LOOP"** | • If AI is only intended to augment human decisions, both ICO and FCA require firms to ensure that human intervention (human-in-the-loop) is meaningful.<br>• "Humans-in-the-loop" must have the tools, authority and incentives to understand, validate, and reject/accept the AI output. | • The bank must ensure human reviewers have access to a clear explanation about why customers are flagged as potentially vulnerable by the AI system.<br>• The bank should consider which additional information reviewers can take into account to validate the AI's output (e.g. talk to customers).<br>• Performance targets should not pressure human reviewers into accepting/rejecting the AI's vulnerability classification by default. |

## AI ETHICS SPOTLIGHT
## FULLY AUTOMATED AI DECISIONS

- The use of fully automated decisions and profiling can be lawful under specific conditions.
- However, society is challenging increasingly whether the use of such practices is morally acceptable, especially when decisions made solely by software have a significant impact on people's lives.
- To build trust in AI, firms will need to go beyond compliance. They will need to adopt their own positions about whether the use of fully automated decision-making in their business is ethical and in what circumstances.

- In our example, the bank's AI system uses fully automated decision-making in the case of customers flagged as displaying low vulnerability drivers - the classification triggers a generic email reminding these customers of all available support channels.
- The bank will need to decide whether its use of fully automated decision-making is ethical, as wrongly classified customers will not be given proactive support.
- This may depend on the accuracy and explainability of the system for each type of vulnerability (e.g. financial resilience vs. mental health).
- The bank should also consider the opportunity cost of not using the AI system - i.e. fewer customers receiving proactive support overall.

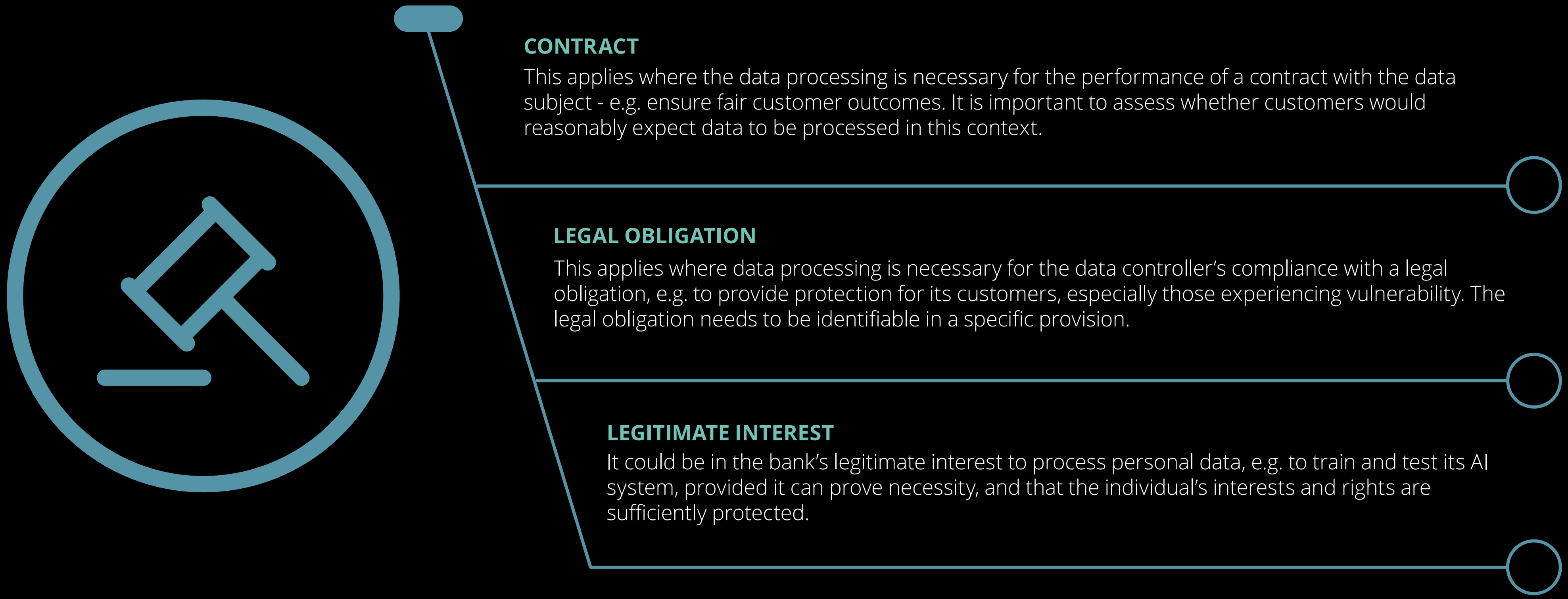
# Regulatory implementation challenges

Firms with predominantly digital customer journeys may need to use customer data and technology to detect and provide additional support to vulnerable customers, as noted by the FCA. GDPR does not prevent them from doing so in principle, but in practice some of its requirements that do not have a parallel in conduct regulation may limit firms' ability to leverage personal data and fulfil the FCA's expectations in relation to vulnerable customers.

| | Overview of GDPR requirements | Case study example |
|---|---|---|
| **LAWFUL BASIS FOR PROCESSING PERSONAL DATA** | • Firms must identify the lawful basis most appropriate for their purpose before processing any personal data. This applies to new data, but also to data it collected previously.<br>• In an AI context, firms might need to identify two lawful bases to process personal data: one to develop and test the AI system, and one to classify customers once the system is deployed. | • There are several relevant lawful bases that the bank could consider to develop and run its vulnerability identification AI system - e.g. contract, legal obligation, legitimate interest.<br>• Given the sensitive nature of the personal data the AI system is likely to use and infer, and the impact on potentially vulnerable customers, it is particularly important for the bank to choose appropriately. However, given the lack of leading practice examples and guidance, the bank may determine that the regulatory risk is greater than its risk appetite. |
| **PURPOSE LIMITATION** | • Firms' purposes for collecting and using personal data must be clear and documented from the outset.<br>• Firms exploring secondary use of data, must consider whether this is compatible with the original purpose for which the data was collected, or whether further consent from data subjects is required. | • The bank's purposes for processing personal data are the development and running of an AI system to enable the proactive identification and support of vulnerable customers.<br>• As such, the bank may not be able to process data it already holds on its customers without obtaining further consent, if the data was collected for different purposes, such as marketing or other commercial objectives. Lack of usable personal data may prevent the bank from training and/or running its AI system. |
| **DATA MINIMISATION** | • Firms must only use the minimum amount of personal data necessary for their specified purposes.<br>• Data minimisation applies to the training, testing, and deployment stages of the AI lifecycle.<br>• Additional personal data should only be used if the benefits outweigh the potential additional harms for individuals. | • The bank must limit the use of personal data to what is sufficient to identify vulnerable customers properly.<br>• The bank needs to assess how each data feature contributes to the correct identification of vulnerable customers, and balance that contribution against potential customer harms. For example, if customers' location data makes a limited contribution to identifying vulnerability, it should be excluded to avoid further erosion of customers' privacy.<br>• The assessment of the contribution of each data feature could be challenging given the broad and evolving definition of vulnerability, and the number of potential drivers. |

# Using AI to support vulnerable customers: potentially relevant lawful bases under GDPR

Balancing privacy considerations and data collection is a complex exercise in this field but, if done properly, AI systems can help firms meet their legal obligations. Although privacy compliance needs to be addressed on a case-by-case basis, we believe that there are several legal bases under GDPR for firms to rely on when processing data for the purpose of identifying vulnerable clients. The bank will choose a lawful basis based on an assessment of its appropriateness and legal strength, and this will be driven in part by how confident the bank feels that it understands the ICO's supervisory approach and expectations.



**CONTRACT**

This applies where the data processing is necessary for the performance of a contract with the data subject - e.g. ensure fair customer outcomes. It is important to assess whether customers would reasonably expect data to be processed in this context.

**LEGAL OBLIGATION**

This applies where data processing is necessary for the data controller's compliance with a legal obligation, e.g. to provide protection for its customers, especially those experiencing vulnerability. The legal obligation needs to be identifiable in a specific provision.

**LEGITIMATE INTEREST**

It could be in the bank's legitimate interest to process personal data, e.g. to train and test its AI system, provided it can prove necessity, and that the individual's interests and rights are sufficiently protected.

Finally, the bank could be able to process **special categories of data** (e.g. gender, ethnicity, health - whether collected directly or inferred) for the purpose of **identifying and mitigating bias and ensuring equality of opportunity or treatment between groups of people** (Data Protection Act 2018 - Schedule 1, Part 2).

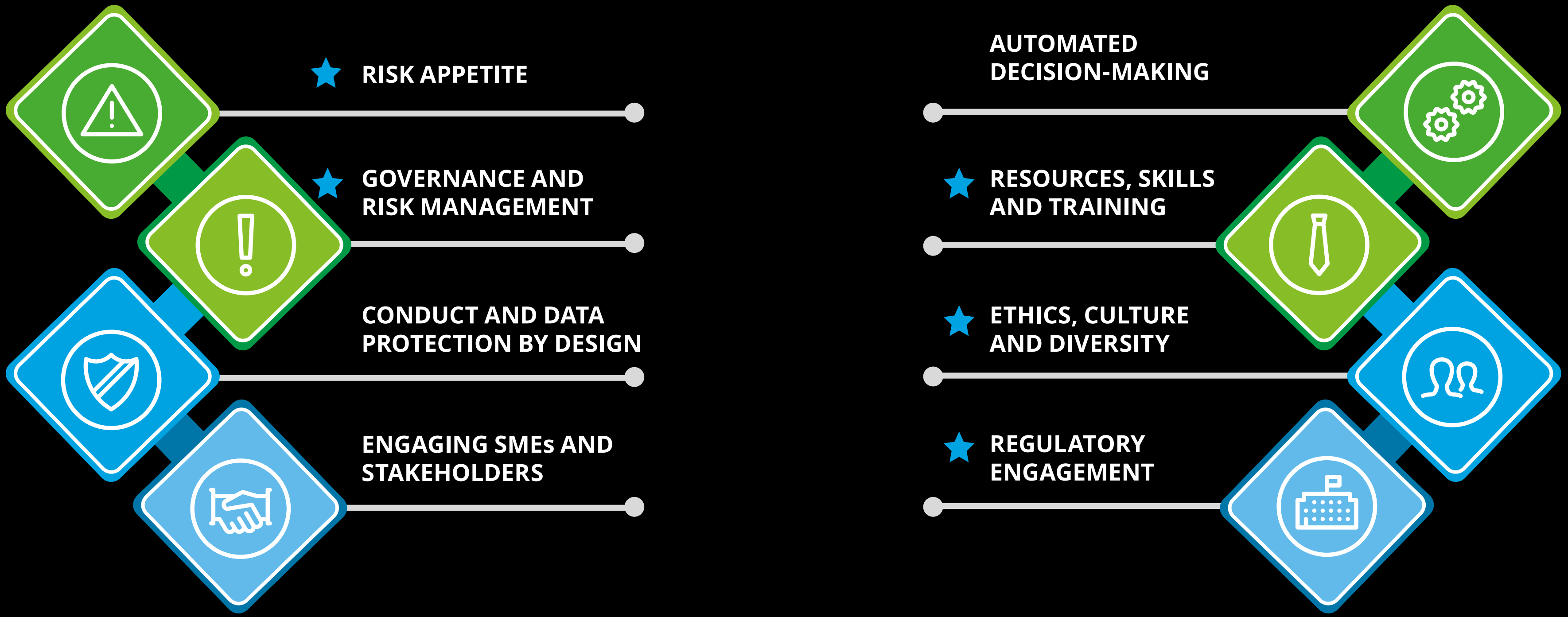
# Using AI to support vulnerable customers: firm level capabilities
# Protecting consumers and ensuring a lawful and ethical approach to AI

We have explored how conduct, data protection regulation and AI ethics interact in the case of an AI system used to detect customer vulnerability in retail banking. The bank in our case study will need to consider all three dimensions jointly to ensure good customer outcomes.

**Below we highlight some of the key areas that the bank will need to consider as it develops its approach to Trustworthy AI.**

**Click on each box for more details.**

## Areas of focus



★ RISK APPETITE

★ GOVERNANCE AND RISK MANAGEMENT

CONDUCT AND DATA PROTECTION BY DESIGN

ENGAGING SMEs AND STAKEHOLDERS

AUTOMATED DECISION-MAKING

★ RESOURCES, SKILLS AND TRAINING

★ ETHICS, CULTURE AND DIVERSITY

★ REGULATORY ENGAGEMENT

★ theme particularly relevant for boards.

# Drawing on our case study - what can firms do?

| THEME | ACTIONS | EXAMPLES OF EFFECTIVE CONTROLS |
|---|---|---|
| **RISK APPETITE** | The bank's risk appetite should include AI-specific conduct and data protection considerations to set clear limits, e.g. in relation to the use of automated decision-making or explainability, within which the AI system must operate. | *The bank has enhanced its risk appetite statement to include AI-specific considerations. For example, "we will never use fully automated decision-making to determine whether or not a customer should be considered vulnerable".*<br>*The risk appetite clearly ties in with the governance process, whereby high risk AI-enabled solutions, to be delivered at scale, are escalated to the right level of seniority for debate and challenge, including to boards, before being deployed.* |
| **GOVERNANCE AND RISK MANAGEMENT** | The bank's structure; roles and responsibilities; and model risk management should support a holistic approach to, and compliance with, both conduct and data protection requirements across the AI system's lifecycle. | *The bank has established an AI oversight committee, with representatives from across the business, and an AI centre of excellence to deliver AI systems. There is a clearly identified owner for the AI vulnerability system who is responsible for verifying its compliance with data protection and conduct requirements at each key stage (training, testing, and deployment). The bank has also created a vulnerability policy that includes information on the vulnerabilities present - and likely to be present - in its customer base and target market.* |
| **CONDUCT AND DATA PROTECTION BY DESIGN** | Conduct, data protection, and AI ethics should be considered from the design phase. AI technical and business requirements should be compatible with relevant regulatory obligations, and with the firm's ethical principles and risk appetite. | *The bank has a formal process, involving relevant experts from across the business, to ensure that personal data used to build/run the AI system are processed in a lawful, fair, transparent and ethical way.* |
| **ENGAGING SMEs AND STAKEHOLDERS** | Given the broad range of vulnerability types and drivers, the firm should involve a wide range of SMEs and stakeholders in the design of its AI system. This would also test its social acceptability and highlight any ethical concerns. | *The bank has engaged with external research - e.g. mental health charities' reports on vulnerable customers - and data to identify the characteristics of vulnerability that may be present in its customer base, and the bank's response is suitable for, and ethically acceptable to, its customers.* |

# Drawing on our case study - what can firms do? (continued)

| THEME | ACTIONS | EXAMPLES OF EFFECTIVE CONTROLS |
|---|---|---|
| **AUTOMATED DECISION-MAKING** | A firm should mitigate the risk that decisions supported by the AI system could *de facto* be considered fully automated (e.g. in the case of false negatives or ineffective human reviews). | *The AI system is not the only process the bank uses to identify vulnerability. Human reviewers/front line staff have the tools, skills, incentives and authority to use their judgment to support vulnerable customers.* |
| **RESOURCES, SKILLS AND TRAINING** | The bank must review and address the capacity, skills and training needs of staff involved in the governance, development, validation, and use of the AI system. Moreover, the bank's frontline staff should have the necessary skills to recognise and respond to a range of characteristics of vulnerability. | *Data protection and conduct compliance teams have been trained and are able to engage with each other and with developers on the regulatory implications and risks of different AI design, training, and testing choices. Technical playbooks are in place to help guide AI developers. Vulnerability champions - with expertise in different types of vulnerability - are available to discuss complex cases and support front line staff.* |
| **ETHICS, CULTURE AND DIVERSITY** | The bank's board and senior leaders should create and champion a culture that prioritises the fair treatment of vulnerable customers. Using AI to support vulnerable customers will involve ethical trade-offs - e.g. between privacy and statistical accuracy. Choosing the right course of action will require effective ethical frameworks, and diversity of thought and perspectives. | *The bank's board and senior management are diverse and engage openly with different views in relation to the balance between privacy and proactive support for vulnerable customers. The firm's underlying purpose is driven by a desire to achieve good customer outcomes - including for vulnerable customers.* |
| **REGULATORY ENGAGEMENT** | A firm should engage early, proactively, and openly with both the FCA and ICO about its plans to use AI to detect and support vulnerable customers, and seek their views on its compliance approach. | *The bank is making use of the FCA and ICO innovation support teams. It is also working with industry groups to create an industry code that addresses data protection issues around using AI to detect vulnerable customers, with support from the ICO. The bank is applying to the FCA's Digital Sandbox, whose areas of focus include use of data to support vulnerable customers.* |

# REGULATION OF AI - LOOKING AHEAD

> "While the widespread use of AI presents us with complex, ethically-charged questions to work through, it also holds enormous promise....It's crucial that we engage with these issues now, not least because we expect the application of machine learning in financial services to increase substantially over the next few years. This is going to require a combined effort. We - regulators, academics, industry and the public - need to work together to develop a shared understanding that will determine our approach over the years ahead."

Christopher Woolard, then Executive Director of Strategy and Competition at the FCA, July 2019

# Looking ahead - the case for further regulatory collaboration and guidance

- AI and data are set to remain key priorities for policy-makers for the foreseeable future. Over the next twelve months the EU and UK will launch a number of strategic policy initiatives aimed at fostering data-driven technological innovation and competition, while also setting strong expectations around data privacy, consumer protection and ethics. Several of these initiatives will be cross-sectoral in nature, but we expect parallel and complementary FS-specific proposals.
- In the meantime, we believe FS regulators and data protection authorities should collaborate more closely on additional regulatory guidance, as well as supervision, to clarify how they should interpret conduct and data protection rules when there is uncertainty around how they interact in an AI-context.

## The case for collaboration

- We believe there is a great opportunity for regulators to collaborate further and more systematically to support innovation in FS areas of significant public interest - e.g. support for vulnerable customers, financial inclusion, fraud detection.
- It is up to industry to voice specific use-case challenges that would benefit from further regulatory co-ordination and guidance. Forums such as the BoE/FCA financial services AI public-private forum provide a valuable platform for regulators to work together with industry to develop best practices to address data protection issues in key AI use cases.
- The FCA and ICO could also benefit from more structured collaboration in relation to the day-to-day practicalities of AI supervision, especially in areas where conduct and data protection requirements are aligned.
- The case for collaboration and harmonisation resonates within the EU as well. The EBA recently underlined the importance of improved dialogue between consumer protection, prudential supervision and data protection authorities.
- Cross-jurisdictional harmonisation is also increasingly important to support international firms. Longer-term projects such as the EU's AI regulatory framework will help to harmonise AI regulation across jurisdictions, but in the short-term networks such as the BIS Innovation Hub can help facilitate collaboration and harmonisation between regulators.

## Using AI to support vulnerable customers: more regulatory guidance is needed

- GDPR has only been in place for a couple of years, and the use of AI to support vulnerable customers is immature. Firms would welcome more clarity on the ICO's expectations around the use of AI and data in this area. The UK Regulatory Network recently noted this as well.
- Through its Innovation Hub, the ICO is already working with other regulators on how to support the use of data for the benefit of vulnerable customers.
- As part of this work, we believe that the ICO and FCA should consider working together, and with industry, to clarify which lawful bases are likely to be suitable and under what conditions. This could have significant benefits in improving FS firms' identification of and support for vulnerable consumers.
- FCA guidance to help banks practically apply AI ethics principles would also be welcome. For EU insurers, EIOPA's work on guidance to operationalise digital ethics principles should be useful.
- **We hope that this paper starts a fruitful debate on how we can collectively innovate safely to support vulnerable consumers**

# APPENDIX 1

## Model Guardian

*Deloitte’s tool to investigate unintended bias in AI*

# Model Guardian
## Deloitte's solution to investigate unintended bias in AI decision-making

Firms have a regulatory and ethical responsibility to **ensure that their AI systems do not unfairly discriminate** particular groups of individuals.

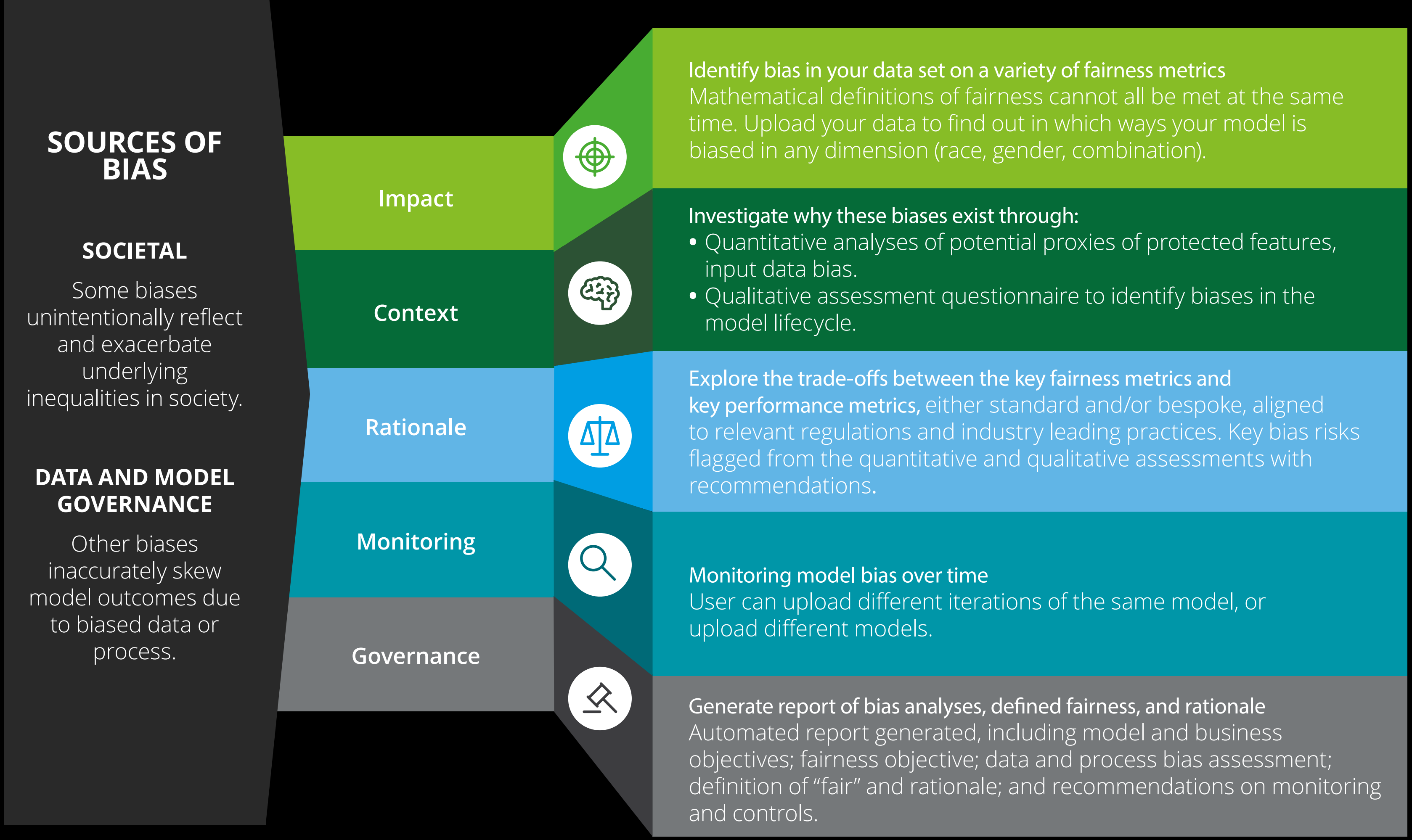
As such, there are **both legal** as well as **reputational risks** associated with deploying an AI system that shows signs of bias and unfair discrimination.

Deloitte's solution - a model **GUARDIAN**

**Guardian is an end-to-end, customisable tool which helps firms identify, investigate and track biases in AI models.**

Guardian should be deployed alongside an effective AI governance and controls framework.

### Mitigating the risks of bias with Deloitte's model Guardian



**SOURCES OF BIAS**

**SOCIETAL**

Some biases unintentionally reflect and exacerbate underlying inequalities in society.

**DATA AND MODEL GOVERNANCE**

Other biases inaccurately skew model outcomes due to biased data or process.

**Impact**

Identify bias in your data set on a variety of fairness metrics
Mathematical definitions of fairness cannot all be met at the same time. Upload your data to find out in which ways your model is biased in any dimension (race, gender, combination).

**Context**

Investigate why these biases exist through:
- Quantitative analyses of potential proxies of protected features, input data bias.
- Qualitative assessment questionnaire to identify biases in the model lifecycle.

**Rationale**

Explore the trade-offs between the key fairness metrics and key performance metrics, either standard and/or bespoke, aligned to relevant regulations and industry leading practices. Key bias risks flagged from the quantitative and qualitative assessments with recommendations.

**Monitoring**

Monitoring model bias over time
User can upload different iterations of the same model, or upload different models.

**Governance**

Generate report of bias analyses, defined fairness, and rationale
Automated report generated, including model and business objectives; fairness objective; data and process bias assessment; definition of "fair" and rationale; and recommendations on monitoring and controls.

# APPENDIX 2
Understanding vulnerability - a broad definition

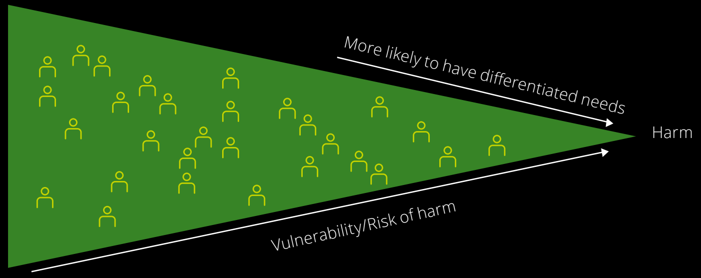
# The FCA's approach to understanding the multiple facets of vulnerability

## TYPES OF VULNERABILITY

TRANSIENT

PERMANENT

## LOOKING AT THE SPECTRUM OF RISKS


More likely to have differentiated needs
Harm
Vulnerability/Risk of harm

Source: FCA draft vulnerability guidance

- All customers sit on a spectrum of vulnerability, and different types/levels of vulnerability will require different forms of support. Firms should address the needs of all customers across this spectrum, but should be particularly careful to address the needs of those most at risk of harm. These are more likely to require more support and adaptations.
- Firms should understand the characteristics of vulnerability likely to be present in their target market and customer base.

A vulnerable customer is defined as **"someone who, due to their personal circumstances, is especially susceptible to detriment, particularly when a firm is not acting with appropriate levels of care".**[3]

## DRIVERS OF VULNERABILITY

| HEALTH | RESILIENCE | CAPABILITY | LIFE EVENTS |
|---|---|---|---|
| e.g. poor mental health | e.g. low or erratic sources of income | e.g. low English language skills | e.g. changes in caring responsibilities |

## IMPACT OF VULNERABILITY

- Increased stress levels
- Increased time pressures as a result of having to fulfil other additional responsibilities
- Reduced ability to cope and manage resulting from an increase in stress levels ("less headspace")
- Reduced processing power and ability because of the side effects (physical or emotional) of the vulnerability
- Lack of perspective and understanding of the implications (including financial)
- Changing attitude towards risk

[3] 3, FCA, Guidance for firms on the fair treatment of vulnerable customers, July 2020, available at https://www.fca.org.uk/publication/guidance-consultation/gc20-03.pdf

# Contacts

David Strachan
Partner
Head of EMEA Centre for Regulatory Strategy
dastrachan@deloitte.co.uk

Paul Garel-Jones
Partner
Risk Advisory
pgareljones@deloitte.co.uk

Peter Gooch
Partner
Risk Advisory
pgooch@deloitte.co.uk

Cindy Chan
Partner
Risk Advisory
cichan@deloitte.co.uk

Reny Vargis-Cheriyan
Partner
Risk Advisory
rvargischeriyan@deloitte.co.uk

Ivana Bartoletti
Technical Director
Risk Advisory
ibartoletti@deloitte.co.uk

Matt Papasavva
Associate Director
Risk Advisory
mpapasavva@deloitte.co.uk

# Authors

Suchitra Nair
Director
EMEA Centre for Regulatory Strategy
snair@deloitte.co.uk

Valeria Gallo
Senior Manager
EMEA Centre for Regulatory Strategy
vgallo@deloitte.co.uk

Morgane Fouché
Manager
EMEA Centre for Regulatory Strategy
mfouche@deloitte.co.uk

Ben Thornhill
Senior Associate
EMEA Centre for Regulatory Strategy
benjaminthornhill@deloitte.co.uk

# Swiss Contacts

Ralph Wyss
Partner
Head Risk and Regulatory Assurance
rwyss@deloitte.ch

Sandro Schoenenberger
Partner
Audit and Assurance
sschoenenberger@deloitte.ch

Carlo Schmid
Assistant Manager
Risk and Regulatory Assurance
cschmid@deloitte.ch

CENTRE *for*
**REGULATORY STRATEGY**
**EMEA**

The Deloitte Centre for Regulatory Strategy is a powerful resource of information and insight, designed to assist financial institutions manage the complexity and convergence of rapidly increasing new regulation.

With regional hubs in the Americas, Asia Pacific and EMEA, the Centre combines the strength of Deloitte's regional and international network of experienced risk, regulatory, and industry professionals – including a deep roster of former regulators, industry specialists, and business advisers – with a rich understanding of the impact of regulations on business models and strategy.

**Deloitte.**

This publication has been written in general terms and we recommend that you obtain professional advice before acting or refraining from action on any of the contents of this publication. Deloitte LLP accepts no liability for any loss occasioned to any person acting or refraining from action as a result of any material in this publication.

Deloitte LLP is a limited liability partnership registered in England and Wales with registered number OC303675 and its registered office at 1 New Street Square, London, EC4A 3HQ, United Kingdom.

Deloitte LLP is the United Kingdom affiliate of Deloitte NSE LLP, a member firm of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"). DTTL and each of its member firms are legally separate and independent entities. DTTL and Deloitte NSE LLP do not provide services to clients. Please see www.deloitte.com/about to learn more about our global network of member firms.