UK FINANCE

HERBERT SMITH FREEHILLS

FAIR USE OF AI

A UK FINANCE WHITEPAPER

June 2022

CONTENTS

FOREWORD

We see today growing scrutiny of the use of artificial intelligence (“AI”) and algorithms, and a greater awareness of the potential for these technologies to exacerbate some consumer risks.

Questions touching on AI ethics and fairness are becoming more mainstream. In recent years we have seen greater public consciousness of what was once a relatively obscure academic issue. This includes the close attention paid to the use of an algorithm to determine 2020 A-levels by the Office of Qualifications and Examinations Regulation (or “Ofqual”).

There is no doubt that law makers and regulators around the world are taking note and working to update their own approaches. In the UK, policy makers at the Office for AI are working on an AI policy whitepaper, while the Department for Digital, Culture, Media and Sport has consulted on changes to data protection law to account for AI challenges. On the regulatory front, the Digital Regulatory Cooperation Forum is considering the approach of UK regulators to algorithm-related issues and the Equality and Human Rights Commission has included AI guidance in its strategic plan for 2022-25. And in financial services specifically, the AI Public Private Forum has produced its long-awaited report, with a discussion document from the Bank of England and FCA signposted as a next step.

This is all taking place within the wider context of a greater public awareness of social justice issues. These can be complex at times, with an interaction between current practices and the legacy of historical injustices, which can still be felt today.

Nonetheless, AI holds the promise to enable a leap forward in the provisions of financial services, not only in efficiency improvements for firms but also in real benefits for customers. The roll out of AI into more financial sector applications has the potential to bring more personalisation of products and services for consumers, to enable greater financial inclusion and to permit more effective protection against fraud and other economic crime.

It is therefore no surprise that questions of AI fairness and bias are front of mind. It will take time for UK Plc to work through all of the complexities so that consumers and society can enjoy the benefits of AI technology, with confidence that it is being used fairly and ethically. We hope that this whitepaper will be a helpful contribution to the debate.

Jana Mackintosh
Managing Director: Payments
and Innovation
UK Finance

Karen Anderson
Partner
Herbert Smith Freehills

INTRODUCTION

ARTIFICIAL INTELLIGENCE AND ETHICAL PRINCIPLES

Artificial intelligence ("AI") offers the potential to greatly enhance services and products across the economy, enabling greater personalisation, more accurate predictions and enhanced efficiency.

Nonetheless, as the use of AI becomes increasingly common, there is a growing interest among firms, as well as from government authorities and the public, in ensuring that it is used in an appropriate and ethical manner. Although ethical and other risks associated with AI have been a topic of discussion for some time, there remains a lack of clear guidelines on certain elements of the use of AI and how it should be deployed, including in financial services.

In order to help firms navigate the potential pitfalls associated with using AI – with the objective of promoting its ethical use and maintaining public trust – UK Finance produced as a whitepaper the following set of ethical principles for AI and advanced analytics ("**AAAI**") in financial services (the "**AAAI Principles**"):

- **Principle 1:** Explainability and Transparency – Be transparent about how we use AAAI and provide appropriate explanations on decisions.
- **Principles 2:** Integrity of AAAI – Adopt appropriate controls for the integrity, sourcing and sharing of AAAI and its associated data throughout the AAAI lifecycle.
- **Principle 3:** Fairness & Alignment to Human Rights – Design and use AAAI that produces fair outcomes (the "Fairness Principle").
- **Principle 4:** Contestability & Human Empowerment – Support the empowerment of AAAI subjects, respecting their decision making.
- **Principle 5:** Responsibility & Accountability – Be responsible and accountable for our AAAI.[1]

A high-level ethical framework is an important tool for firms building out their use of AI and algorithms. However, there are challenges to navigate when seeking to apply such principles in practice.

APPLYING ETHICAL PRINCIPLES IN PRACTICE

In order to deepen knowledge and understanding of ethical AI in practice, members agreed to explore in more detail how the Fairness Principle could be applied and what the challenges could be. AI offers the potential for decision-making to be fairer by enabling greater consistency and objectivity but there are certain risks that need to be managed.

As a way of illustrating and exploring these practical challenges, UK Finance and its members discussed how to apply the Fairness Principle to four scenarios where AI could be used in a financial services context. These were – intentionally – theoretical use cases that raise important fairness considerations, rather than necessarily being prevalent use cases currently seen in the market. In summary:

1. **Marketing:** AI offers firms an enhanced ability to identify customer product needs and to offer a wider pool of customers a more bespoke service (usually provided to a more limited pool of premium customers). This scenario focused on the potential to better target advertising to existing customers, using a range of data and data sources, such as transactional data, gender, age and marketing preferences received from the customer, as well as external Telephone Preference Service data.
2. **Credit:** AI can assist firms by detecting customers who may be experiencing financial difficulties, enabling them to assist these customers by taking appropriate action at an early stage. This scenario considered the use of the lender's internal data on the behaviour of the account holder in relation to loans and current accounts to identify borrowers showing signs of financial difficulty for relationship managers to review. This could help identify and manage financial difficulty early on.[2]
3. **Employee monitoring:** Employees can be monitored by AI with the objective of checking regulatory compliance and optimising their productivity; this technology can be used to track employees' productivity and even feed into their performance conversations with managers by producing measurements against objective metrics,

[1] These principles are expanded on in the whitepaper, available here.

[2] Although touched on tangentially, the use of AI for creditworthiness assessment or lending decisions was not discussed directly.

with a view to avoiding potential bias caused by human judgements. Monitoring technology can also help firms check employee regulatory compliance and productivity more effectively in a 'working from home' environment.

4. **Transaction monitoring:** AI can assist with detecting potential financial crime in transactions, enabling firms to protect customers from fraud and to notify authorities of suspicious activity, such as possible money laundering. The use of AI in transaction monitoring has clear benefits: AI applies consistently and therefore helps avoid human fallibility and can pick up on trends or irregularities that human monitoring alone may not be able to detect. Fully automated AI decision-making can also process far greater quantities of data in a shorter period of time than is possible with human review.

THIS PAPER

Building on the insights from these workshops, this whitepaper sets out four overarching themes, illustrating some of the key issues that were identified during the discussion of these scenarios:

- Defining 'fairness' and the requirement for fairness of both outcomes and the process surrounding the use of AI.
- The fair use of AI cannot be considered in siloes, as it requires the application of other AI principles and expertise from across firms.
- The use of AI involves trade-offs of competing outcomes and objectives, though existing laws help to justify certain outcomes as "fair" outcomes, and provide some parts of a fair process.
- Although existing laws assist, how to apply them to AI use cases is not always clear, so further guidance would help.

The discussion of these issues is in the context of a number of legislators, regulators and trade bodies considering how AI should be used in the financial services sector. Notably the Kalifa Review of UK FinTech identified the need for guidance from the Prudential Regulation Authority and the Financial Conduct Authority in relation to the application of AI across several areas (accountability, governance, explainability and human oversight).[3] More generally, in January 2022 the UK government announced an initiative with the intention for the UK to lead in shaping global technical standards for AI.[4] In addition, the European Commission has published a cross-sectoral legislative proposal on the harmonisation of rules regarding AI.[5]

Financial services firms already have detailed controls and governance to ensure compliance with regulatory and legal obligations and ensure effective service delivery. Nonetheless, firms deciding to employ AI to deliver services and engage with staff will need to reflect on how this new technology interacts with the rules and on whether existing governance arrangements may need updating.

The purpose of this whitepaper is to add to the debate surrounding how AI should be used in the financial services sector and highlight areas where further guidance may be beneficial to firms and customers.

The term 'artificial intelligence' is not concretely defined in this whitepaper. This work focused on systems using sophisticated techniques such as machine learning, but the issues discussed will at times be shared with simpler technologies and the use of analytics and algorithms more broadly.

[3] *Kalifa Review of UK Fintech*.

[4] More information on the AI Standards Hub is available here.

[5] Text of the European Commission's proposal for an 'AI Act' available here.

1. DEFINING AND APPLYING FAIRNESS

1.1. A key starting point is being clear on what is meant by "fairness". There is a risk that definitions relating to ethics and fairness regarding AI could be ambiguous and there are a number of ways in which these concepts are measured or quantified.[6]

1.2. The FCA has tended to approach the definition of fair treatment of customers (or 'treating customers fairly', sometimes abbreviated to 'TCF') in terms of consumer outcomes, publishing in 2006 six TCF outcomes that firms should be striving to achieve through the product/service lifecycle. These largely focused on internal processes, management and performance.[7] Both then and now, illustrative examples are primarily cast in terms of unfairness – see further discussion under 1.12 below.

1.3. Defining what fairness entails can be difficult to set out fully. There have been several publications which have considered how AI should be used within financial services[8], with different approaches taken to what is meant by "fairness". These approaches broadly fall into two categories. The first focuses on the outcomes of the AI application, while the second relates to the fairness of the process used.

OUTCOME FAIRNESS

1.4. Outcome fairness was the focus of the third principle in the UK Finance AAAI Principles whitepaper. An example of an outcomes focused definition of fairness includes the European Banking Authority's Report on Big Data and Advanced Analytics which states that "fairness requires that the model ensures the protection of groups against (direct or indirect) discrimination"[9].

1.5. Such an outcomes-focussed definition of fairness is a necessary component for ensuring the fair application of AI. Indeed, given AI is used in order to achieve certain outcomes, it is natural to consider whether such outcomes are fair. Examples discussed by members include:

1.5.1. Has previous unfair discrimination impacted the data which has been used by the AI tool, such that the outcomes of the AI tool perpetuate discrimination?

1.5.2. More generally has the application of AI led to unfair bias or discrimination against certain groups?

1.5.3. Even if fair for data subjects in general, might the use of AI have unintended consequences or be unfair on certain types of individuals?

1.6. The risk of 'unfair bias' in an AI system is a prominent concern among firms, policy makers, academics and consumer advocates. However, agreeing what in fact might constitute an unfairly biased AI model or system is not necessarily straightforward. Certain statistical approaches have been proposed, focused on either ensuring that individuals who are similar in relevant ways are treated similarly, or else on trying to remove or reduce differences in outcomes between different social groups such as between men and women (e.g. demographic parity,

[6] This is consistent with similar observations in the minutes from the first Artificial Intelligence Public-Private Forum, held 12 October 2020.

[7] See for example Conduct Rule 4 of the FCA Handbook.

[8] For example, the Monetary Authority of Singapore's *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector.*

[9] EBA Report on Big Data and Advanced Analytics January 2020 EBA/REP/2020/01.

conditional demographic parity, equalised odds).[10]

1.7. The ability to agree on what constitutes an unfairly biased AI model is also complicated by questions about from whose perspective fairness should be judged, what outcomes are being measured and how you measure them. For example, an objectively optimal outcome from the perspective of policy makers as stakeholders, driven by consumer protection objectives, may not meet the individual end user's expectations and immediate needs (discussed further below). There is then the related question of how you measure outcomes. What are the metrics you want to consider and how do they relate to what you consider a 'fair' outcome? For financial services, this question naturally leads to answers regarding financial outcomes. However, different stakeholders may have different concepts of "fairness" involving other values such as environmental or social values, including an end user's desire to be treated with respect and dignity. Some thought also needs to be given to the outcome for individuals or groups to whom the opportunity of participating in the process is not extended. If the AI model does not measure such values, it will be difficult to consider whether it is realising a 'fair' outcome from such perspectives.

CORE LEGAL AND REGULATORY CONSIDERATIONS FOR OUTCOME FAIRNESS

1.8. In parallel to such academic concepts, there are numerous laws and regulations that firms must consider. Although there is no 'AI fairness law', there are numerous layers of more general horizontal and sectoral rules for firms to consider.

1.9. Given the centrality of 'bias' to discussions of AI fairness, the Equality Act is a key starting point. The Equality Act sets out the 'protected characteristics' for the UK. These are broadly: age, disability, gender reassignment, marriage / civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation. The Act then sets out a framework for identifying illegal discrimination, broadly:

1.9.1. Direct discrimination, where an individual is treated worse because of one or more protected characteristics. Such conduct is illegal, though some narrow exceptions, for example some single-sex services are permitted, or services intended for specific age groups, where this targeting can be objectively justified.

1.9.2. Indirect discrimination, where an action has a worse impact on individuals who share a particular protected characteristic, compared to those who do not share that protected characteristic, despite this characteristic not being directly included in the decision-making. This is illegal also, unless the difference can be objectively justified.

1.9.3. An objective justification must be a proportionate means of achieving a legitimate aim. A key consideration for firms would be to identify a legitimate aim when first developing an AI system and before any problems emerge, rather than trying to back-fill a justification in the event of a complaint.

1.9.4. Positive action is where a firm considers that a certain protected characteristic group suffers a disadvantage connected to that characteristic, has different needs, or has disproportionately low participation in an activity, and then decides to take an action to address this, such as an emphasis on marketing a product to a particular community (as long as the product is available to all who met certain criteria). If the action is proportionate, this can be legal. However, subject to some limited exceptions, the firm cannot engage in positive discrimination, which involves trying to address the impacts of past discrimination by treating someone

[10] See for example the 2020 *Review into bias in algorithmic decision-making by the Centre for Data Ethics and Innovation.*
Also see Counterfactual fairness, from the Alan Turing Institute (2018, Kusner, Loftus, Russell, Silva) on developing a framework to ensure fairness through the use of causal methods to produce 'counterfactually fair' algorithms, based on the idea that a decision is fair towards an individual if the outcome is the same in reality as it would be in a 'counterfactual' world, in which the individual belongs to a different demographic.
It should also be noted that there is some dispute as to whether there is in fact a meaningful difference between 'individual fairness' and 'group fairness'; see for example: On the apparent conflict between individual and group fairness (2019, Binns).

with a protected characteristic more favourably (e.g. by having different application criteria for a product so that those with a particular protected characteristic will be more likely to be successful).

1.10. This legal model does not necessarily map well onto the types of group and individual fairness that have been proposed more academically, requiring consideration not only of patterns seen in outcomes but also of the reasons for these and the intentions of the firm. And there is clearly a difficult balance between positive action, and positive discrimination. In addition, whether an objective justification can be appropriate will depend on the surrounding circumstances. For example, requiring a university degree as part of recruitment criteria may be able to be objectively justified for certain roles (for example, where a professional body requires it). However, if a mortgage product had a requirement to hold a university degree as part of its application criteria, this may be more difficult to justify.

1.11. Data protection law also puts in place a horizontal set of rules that must be considered when implementing AI that involves personal data. The UK General Data Protection Regulation ("**GDPR**")[11] requires that the processing of personal data be 'processed fairly' as a part of Article 5(1)(a). This is linked to the requirements to have a legal basis for processing, to ensure transparency of processing and to enable information rights to be exercised, though these are not explored here in detail. GDPR fairness is, according to the brief ICO guidance, also a matter of considering individuals' interests and of determining whether the data processing could lead to 'unjustified adverse impacts' on them. This has some similarities to the concept of indirect discrimination and objective justification under the Equality Act (above) but applies beyond just differences in outcomes between protected characteristic groups.[12]

1.12. Within financial services, firms must consider the TCF rules, as noted above. But a future regulatory change will also create a fresh set of considerations for firms attempting to deploy AI. In particular, in the UK the FCA will be introducing a new Consumer Duty, the draft proposals for which include a new Principle for Businesses that 'a firm must act to deliver good outcomes for the retail consumers of its products' (the "Consumer Principle"). The regulator's examples of 'not good' outcomes may have relevance to AI use cases, for example, where they may involve the exploitation of behavioural biases, loyalty, inertia, informational asymmetries or characteristics of vulnerability, the use of negative friction (sludge practices), unreasonable post sale barriers, and the provision of 'poor' support.

1.13. The phrase 'good outcome' does not have an established legal meaning, and firms are left to grapple with the fact that the regulator's view of a 'good' outcome may well differ from that of the consumer, or indeed the Financial Ombudsman Service ("**FOS**"). For example, the rejection of an application for credit (or for a product) on the grounds of (un)affordability is an outcome the regulator may prefer but may not be the outcome desired by the customer.

1.14. Furthermore, calibrations made to models to produce 'fair' (or good) outcomes for the population as a whole (e.g. that may compensate for historical bias) may lead certain individuals to experience 'less good' outcomes than they had experienced in the past. When navigating and applying the FCA requirements, care will be needed to also consider the Equality Act rules relating to objective justification, positive action and positive discrimination, which may not map cleanly onto the FCA expectations. Firms will therefore need to ensure that the metrics they use to measure the outcomes of AI models are able to also measure 'good' outcomes for customers as well as 'fair' outcomes, in addition to being able to explain why calibrations to models achieve both (and if there is a trade-off between the two, why it is necessary and appropriate in the circumstances).

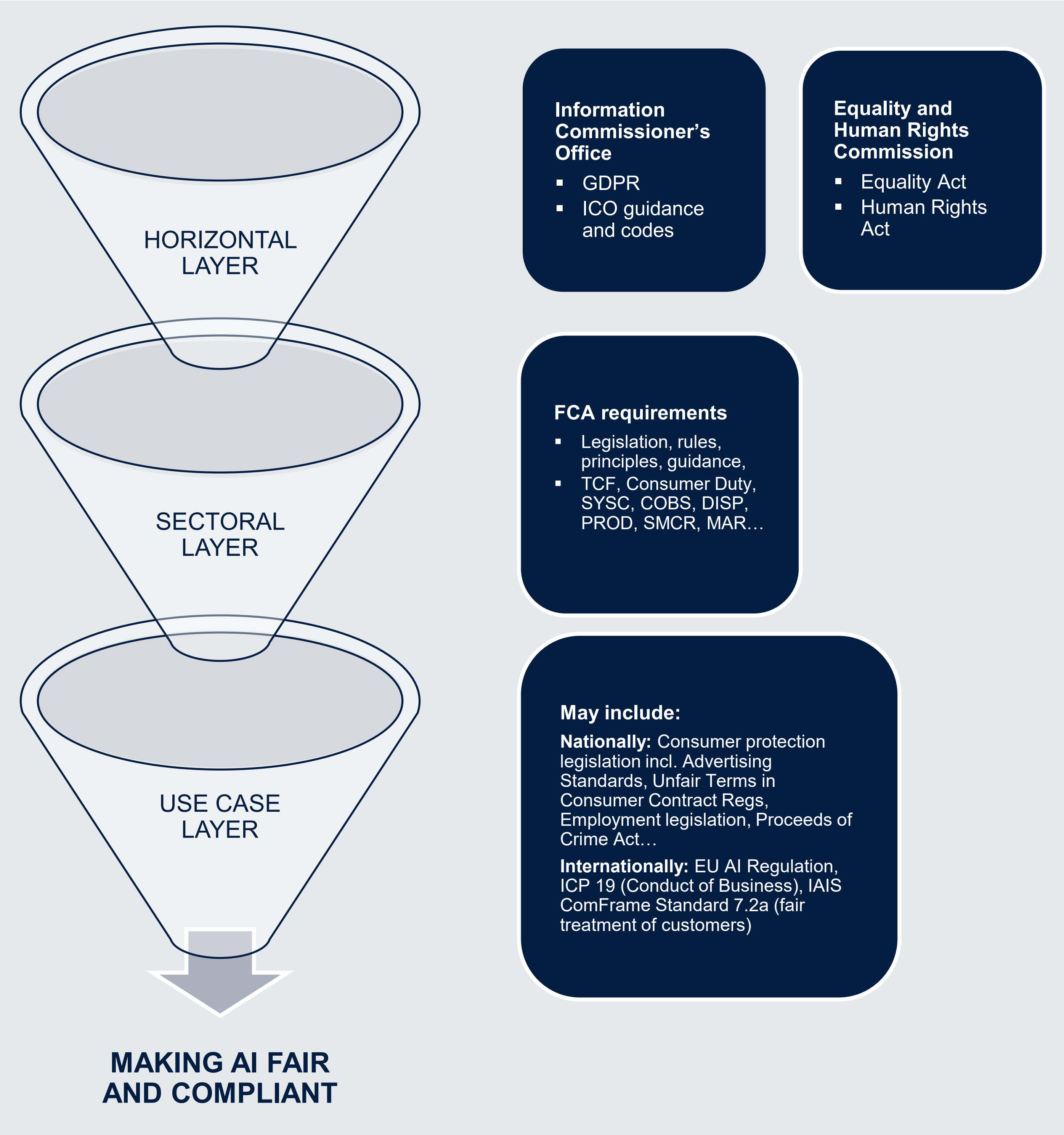
1.15. There can also be tension between the differing regulatory requirements with respect to what the best treatment of the customer is. Under the draft Consumer Duty guidance, the FCA expects firms

[11] The UK version of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

[12] ICO *Guide to the General Data Protection Regulation* - Principle (a): Lawfulness, fairness and transparency

to make particular efforts to support vulnerable customers and customers in financial difficulty. At the extreme, a firm could profile its customers with AI to identify such individuals then use the insights to consider and act on their individual needs. However, while these efforts would be intended to benefit those customers, they would arguably require the individual's explicit consent under data protection rules, meaning they might not be feasible in practice.

Layers of fair AI law and regulation



Information Commissioner's Office

- GDPR
- ICO guidance and codes

Equality and Human Rights Commission

- Equality Act
- Human Rights Act

FCA requirements

- Legislation, rules, principles, guidance,
- TCF, Consumer Duty, SYSC, COBS, DISP, PROD, SMCR, MAR…

May include:

Nationally: Consumer protection legislation incl. Advertising Standards, Unfair Terms in Consumer Contract Regs, Employment legislation, Proceeds of Crime Act…

Internationally: EU AI Regulation, ICP 19 (Conduct of Business), IAIS ComFrame Standard 7.2a (fair treatment of customers)

PRACTICAL CHALLENGES IN ACHIEVING OUTCOME FAIRNESS

1.16. Whatever metrics are used, a fundamental challenge is access to data. This can arise in three main ways:

1.16.1. The firm might not be legally allowed to collect the data needed to check for unfair bias or indirect discrimination, particularly if an individual objects. GDPR tightly restricts the collection of data relating to 'special categories of personal data', including in particular: race, ethnicity, health, religion and sexual orientation. While there is provision for processing special categories of personal data for the purpose of equal opportunities monitoring in the Data Protection Act 2018, this not permitted where either the data will be used to make a decision about an individual or the individual objects.

1.16.2. If the law is clarified and firms were to start collecting such data, they would still need to develop a suitable way of doing so and of communicating with customers, who would likely be surprised to be asked about their religion, etc, when signing up for a loan or bank account. Indeed, customers could well be suspicious, notwithstanding the good intentions of firms and policy makers. This would be a challenge at the point of customer onboarding but could be particularly challenging for existing (back book) customers, who may not respond to requests for information. Even if customers do respond, the accuracy of the demographic data obtained might be questionable and may not be easily verifiable, especially if customers are mistrustful of the data collection.

1.16.3. Beyond the data on customer characteristics, data about outcomes might not be available, impeding firms from analysing whether patterns of outcomes across different groups might cause concern and need investigation. For example, if AI is used to inform lending decisions, the lender will be able to follow the performance of customers approved for a loan but cannot know whether customers turned down for a loan would have paid it back, had they been accepted. In the context of lending, tools like reject inference provided by credit reference agencies help firms to manage this risk at present.[13]

1.17. An additional practical challenge is determining the appropriate level of analysis. A review of small business loans might reveal a significant difference in approval rates for different demographic groups. However, it could be the case that this is due to these groups tending to apply for different types of financing, with demographic differences in approval rates disappearing at a more granular product level.

1.18. Similarly, significant systems work would be needed. Firms would need to develop appropriate labels for different protected groups, with the appropriate level of granularity. Systems would also need to be able to accommodate individuals that identify with multiple demographic groups, such as individuals with mixed ancestry.

BOX 1:

In discussing the application of AI to monitoring transactions for signs of fraud and money laundering, which need to be notified to law enforcement in Suspicious Activity Reports, members discussed the hypothetical scenario where a higher fraction of transactions from Wales were being flagged as high risk by the system, relative to transactions from other areas. The firm might attempt to adjust the model to bring the results for the different UK nations closer together in order to come closer to demographic parity and avoid discrimination on the basis of national origin. However, doing so could be considered positive discrimination, risking a breach of the Equality Act. Discussions identified that the relative accuracy of analysis for different nations would likely be more important than the number of high-risk transactions identified, but firms would struggle to measure this, as authorities do not provide feedback on how many suspicious activity reports result in criminal charges.

[13] For a short explanation, see for example here.

BOX 2:

A paradigm example of how potentially unfair outcomes can occur is the use of postcodes or geolocation by AI in processes. The risk of bias and discrimination is often cited as a risk in relation to the outcomes of AI processes that use postcodes. While postcode and address are not currently used in creditworthiness assessments, they could potentially be used as inputs in other processes such as for marketing or financial crime risk assessment. There may be a strong correlation at the population level between certain postcodes and marketing interests or financial crime risk, but it could nonetheless be unfair to use postcodes as a predictor on an individual level. There is, for example, the risk of bias or discrimination against protected characteristics such as race, where certain postcodes could be an unintended proxy. The firm might however be able to show that there is an objective justification if accuracy is sufficiently improved by the use of post code data, particularly if the use case helps protect customers, for example by better identifying signs of fraud against them. As an added complexity, the new FCA Consumer Duty may lead to expectations that firms in fact start to use such data in new ways if doing so could help ensure 'good' outcomes for customers through improved accuracy.

1.19. Beyond considerations of 'unfair bias' and discrimination, achieving fairness can also require identifying potential customer or employee segments that might not be accounted for effectively by an AI system. Such groups could experience potentially impactful unintended consequences. This connects to the broad GDPR fairness concept of not processing personal data in ways that could have unjustified adverse impacts on individuals. For example, older customers might have less available social media data, which would impact accuracy for any algorithms using this as an input. Similarly, employee monitoring software might not anticipate employees with certain disabilities; if the staff member has not disclosed an 'invisible' disability to the employer, this would be particularly difficult to account for.

1.20. Firms would need to ensure they have an approach in place to identify and manage exceptions within their overall processes. Indeed, beyond just ensuring fair outcomes, fair process is in and of itself an important component in ensuring overall fairness.

1.21. In particular, how would a firm respond to an individual challenging the outcome of an AI process? For example, could a firm explain how the AI process used a postcode in its decision making? The complexity of the AI process would make it difficult to assess what direct impact the postcode had on any individual result. It would have been one of many variables that were inputs that the AI process will have combined in thousands of ways. The ability to explain the process in a manner that is understandable is not straightforward. Yet this fundamental challenge is one that needs to be addressed to have a fair process in using AI. We explore this further below.[14]

PROCESS FAIRNESS

1.22. An example of a broader definition of 'fair AI', which includes a fair process, is the definition used by the European Commission's Independent High Level Expert Group on Artificial Intelligence's Ethics Guidelines for Trustworthy AI. It states "... fairness has both a substantive and a procedural dimension. The substantive dimension implies a commitment to: ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation... The procedural dimension of fairness entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them. In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable."[15]

[14] We also explore the issue of AI transparency and explainability in more detail in a separate paper, available here.

[15] Page 12 of European Commission's Independent High Level Expert Group on Artificial Intelligence's *Ethics Guidelines for Trustworthy AI*.

1.23. In other words, in order to ensure fairness, it is not just the model and its outputs that must be considered. Firms also need to consider the overall process and how customers and other stakeholders interact with it.

1.24. In particular, this requires:

1.24.1. defining the objective of using AI, explaining the potential benefits and risks to relevant stakeholders;

1.24.2. testing the AI system across a diverse mix of data subjects before implementation to understand the likely outcomes in common or impactful scenarios and explore the possibility for bias and/or discrimination;

1.24.3. being able to explain how and why certain outcomes were reached by the AI system;

1.24.4. allowing stakeholders to be able to challenge outcomes and correct data errors.

1.25. The Fairness Principle does not exist in a vacuum. During the workshops, it was clear that to be able to ensure fair outcomes and show a fair process in the use of AI, it was necessary for firms to consider and apply other AAAI Principles:

1.25.1. Explainability & Transparency – in order to understand the decisions made by AI and to be transparent about the process and its outcomes.

1.25.2. Contestability & Human Empowerment – this is a core requirement of any fair process, though how best to apply it will vary depending on the nature of the customer impacts and risks of 'gaming the system'.

1.25.3. Responsibility & Accountability – in order to ensure fair outcomes a firm needs to monitor the application of AI and the decisions it makes, as well as be accountable for the results (and the objectives the firm wants to achieve).

1.26. These three requirements are crucial to developing trust in the system, and the success of the application of AI in customer-facing roles (such as marketing) will depend on developing and maintaining trust. This is even more important for certain customers, given existing apprehensions about firms using their data.

1.27. This can be partly addressed by ensuring customers recognise the potential benefits of applying AI to the use case in question. While the benefits to firms of using AI tools seem clear in terms of efficiency and scale, it can be less obvious for customers in the context of financial services. This is compared to other customer experiences – such as using online marketplaces or streaming sites – where the use of AI is embedded in the customer experience.

1.28. Accordingly, putting the customer experience front and centre of any customer-facing AI process will be key. This would assist firms in being able to clearly articulate how AI processes benefit customers, such as more efficient banking engagement or the benefits of 'industrialised personalisation', where the efficiency of AI enables firms to offer a premium level of service to a larger pool of customers.

1.29. Being able to explain the benefits and objectives of AI would, in turn, improve the accountability of firms for their use of AI and help empower customers and other stakeholders to contest its use.

BOX 3:

In the US, under the Under the Equal Credit Opportunity Act and the Fair Credit Reporting Act, applicants for credit are generally entitled to receive the reasons why creditors take adverse action on their applications and, when creditors use a credit score, the key factors adversely affecting that score (an Adverse Action Notice). In October 2020, the Consumer Financial Protection Bureau hosted a virtual tech sprint focussing on innovative electronic ways of notifying consumers of – and informing them about – adverse credit actions. Various approaches were explored, including:

- Ways to offer more complete information such as explaining how changes to an application could lead to credit approvals in the future or including other features within the notice such as links to a credit report and enhanced information about how to request credit report corrections when appropriate.
- Ways in which the adverse action notice could provide financial literacy and other information and coaching to educate and empower consumers (e.g. about loan programs that have credit score or other specific lending guidelines).
- Creative approaches to improve the format and presentation of the notice itself in order to better engage consumers, particularly when delivered online and on mobile devices (including chat bot driven engagements, customized videos, and links to useful consumer-facing resources).
- Methods for identifying the principal reasons for credit denial for underwriting models that use artificial intelligence/machine learning, seeking to offer effective means to improve their credit profile and improve their chances of credit approval (e.g. an interactive “approval simulator” powered by machine learning that a consumer could use to see what actions, or combination of actions, would most easily yield a credit approval).

1.30. As noted above, a central consideration for fair process from the customer perspective will be the firm’s ability to manage a concern, complaint or challenge against an AI-based decision effectively. Effective customer explanations of decisions and decision-making processes are central to this.[16] In guidance on Explaining decisions made with AI,[17] prepared jointly with the Alan Turing Institute, the ICO stresses that where an AI-assisted decision is made about someone without some form of explanation, this is unlikely to be fair, as it may limit their autonomy and scope for self-determination.

1.31. In this regard, communicating how a decision specific to a consumer was made in language which can be understood by a layperson is usually more important to fairness of process than transparency about, for example, the specifics of the algorithm or algorithms in use.[18] In its guidance, the ICO identifies six main types of explanation for AI decisions, which relevantly include a ‘Fairness Explanation’. This is an explanation of the steps taken across the design and implementation of an AI system to ensure that the decisions it supports are generally unbiased and fair, and whether the individual has been treated equitably. A useful list of what firms may need to show, and how to go about preparing such a fairness explanation, can be found in the Workbook on Use Case 1 which was prepared to build on the ICO’s guidance.[19]

[16] In explaining their use of AI, firms may often find themselves face two ‘audiences’: a consumer (or end-user) audience, and a model audience made up of developers, compliance teams and regulators. See for example the minutes from the fourth Artificial Intelligence Public-Private Forum, held 1 October 2021).

[17] ICO guidance available here.

[18] For an expert model audience, firms need to be able to accurately explain decision flows and assure that any decisions are reliable. So called ‘black-box’ models, that may use deep learning or neural networks (which work through complex interactions between many variables, may infer attributes and put different weights on different attributes), can make it particularly challenging for firms to explain to any audience - let alone a consumer - in simple and readily understandable terms what data was used and how that affected the decision.

[19] Leslie, David, & Briggs, Morgan. (2021). *Explaining decisions made with AI: A workbook (Use case 1: AI-assisted recruitment tool).*

BOX 4:

Hypothetical scenarios in the case studies discussed by members regarding algorithmic assessments for identifying potential financial difficulties among borrowers brought into focus potential challenges in responding to complaints where there may appear to the customer to be discrimination against certain protected characteristic groups. On one level, the fact that an algorithm has been designed to exclude protected characteristics provides an answer to a dissatisfied customer. However, in some scenarios, discrimination or bias may present themselves as the intuitive explanation for differences in outcome, such as different credit outcomes between a husband and wife. The fact that the algorithm was blind to the customer's gender may not be a sufficient response in such situations. An inability to provide more information about how the AI system reached its outcome may make it more difficult to respond to the customer's presumption of discrimination.

Being able to explain the reasons for the decisions made, and potential reasons for differences in outcomes, is an important part of reassuring customers that they have been treated fairly. Being able to explain human oversight arrangements can also help provide reassurance to customers. This has been highlighted in cases investigated by US regulators where, with the benefit of data and detailed explanations, it was concluded there was no bias or discrimination in credit decisions by the firm in respect of husbands and wives, contrary to initial suspicions. However, the firm's inability to provide a sufficient explanation at the time as to why credit decisions could legitimately be different resulted in customers assuming algorithmic bias.

BOX 5:

The workshop discussions regarding transaction monitoring also illustrated where there can be limits to the scope of the application of AI in financial services, which can flow through to fairness considerations and the importance of explainability.

Under the relevant legislation[20], it is necessary for the financial institution's Nominated Officer to form a suspicion of money laundering in order to file a Suspicious Activity Report (SAR)[21]. Conceptually it is only possible for a human to form subjective suspicions.

Thus, while there are many benefits in AI assisting the Nominated Officer to identify potentially suspicious transactions, it cannot replace human decision-making. This being the case, the firm must consider the risk of human overreliance on AI, which could lead to human review of the results failing to identify errors in its output.[22]

Accordingly, the Nominated Officer can only form their own suspicion appropriately if they understand the basis for the identification of the activity as significantly anomalous or otherwise in need of review by the AI, and takes this into consideration, with any other information available, when assessing the activity. This illustrates the importance of appropriate training and a suitable approach to explanations for internal stakeholders in order to achieve fair outcomes for the customer whose activity is being assessed.

[20] Proceeds of Crime Act 2002 (POCA).

[21] In addition to the reporting obligations which will arise in the case of knowledge of, or reasonable grounds to suspect, money laundering.

[22] A classic example of such 'automation bias' is someone following the route suggested by GPS, even where there is no road or even into water or off cliff edges.

2. FAIR USE OF AI CANNOT BE CONSIDERED IN SILOES

LOOKING AT THE FULL CUSTOMER JOURNEY

2.1. It is necessary to consider the fairness of the whole customer interaction with the firm, not just whether a given algorithm is fair in and of itself. Building on the 'process fairness' points above, the same algorithm outputs might result in fair or unfair customer outcomes, depending on how they are acted on by the firm and on the communications with the individual.

2.2. If a firm identifies a pattern of outputs that has the potential to be problematic, such as unexpected differences in outcomes between two protected groups, fully investigating the causes will require more than just examining the workings of the algorithm. For example, if a firm observes a potentially concerning pattern in its AI-based lending decisions, it may need to consider the interaction with its marketing. If the firms marketing practices – potentially themselves underpinned by an AI model's analysis – lead to some specific communities being targeted more than others, this could feed through to lending decision patterns. In either case – lending or marketing – the root cause may stem from the AI model or the live customer data being processed and might or might not be unfair or unjustified.

APPROPRIATE AND PROPORTIONATE USE

2.3. When stepping back and looking at the overall picture of the use case and customer journey, the purpose and effectiveness of the use of AI need to be considered. In other words, is the use of AI appropriate for what the firm wants to achieve and, if so, is the benefit of its use proportionate considering the potential risks of its use?

2.4. Consider for example the hypothetical scenario of employee monitoring with an AI tool that monitors employees' performance using various 'productivity' metrics. These metrics could include the length and number of calls made and meetings attended, periods of activity and inactivity on work devices, keystrokes, social media use on work devices and geolocation data.

2.5. For many roles, such metrics might not in fact reflect the productivity and effectiveness of an employee, particularly for roles where the quality of work cannot be assessed through easily quantified measures, such as the number of meetings attended or emails sent.[23] There may therefore be limited benefit to using an AI tool in such circumstances. It

BOX 6:

Having a fair process and considering the information available to data subjects, along with their ability to 'challenge' decisions, supports fair outcomes. Where AI is used to monitor transactions and identify signs of fraud against the customer, how these insights are acted on and how customer communications are managed could make for a fair or an unfair AI use case. If signs of fraud on a debit card are detected and the card is blocked to protect the customer, there is a risk of unfair treatment if the customer is not notified or is not given an easy route to unblocking the card where in fact there has been no fraud. Similarly, if ongoing inaccuracy is left unresolved and this leads to the card being blocked repeatedly, customers are likely to feel they are being treated unfairly.

[23] The paradigm use case for this sort of productivity monitoring has been in call centres – where productivity is sometimes measured in terms of numbers of calls received, number of breaks, app or web usage, customer waiting times, numbers of calls resolved/closed out; number of complaints or referrals to supervisors etc. often accompanied by a manual random sampling of call quality.

is important to start by identifying the objectives of any AI tool and of ensuring that the tool would in fact meet those needs. Where an AI tool would be of only marginal benefit but would involve a high level of monitoring or intrusion, this would be unlikely to be fair, and indeed might breach privacy laws. In practice, firms may encounter a desire to find new purposes for a tool for which the firm holds a licence, but each new use case has to be considered rigorously.

2.6. In any event, even where an AI tool could be of benefit to an employer in assessing productivity, the use of AI could cause a loss of employee 'agency' in their roles and a corresponding loss of trust in their employer. This would likely be counterproductive to the employer's relationship with its employees.

2.7. In contrast, using AI for employee monitoring may well be more widely accepted by employees, and of clearer benefit to employers, when used to detect fraud and other financial crime. In such cases, the monitoring can act as a deterrent as well as a detector when employees know that surveillance is being undertaken. Clearly explaining the purpose and use of AI in such circumstances before the AI tool is deployed within the workforce would be an important component in having a fair AI process. This allows employees to understand why AI was being used to monitor them, the extent of the monitoring and the extent of any ability to contest its use in the event of concerns. Similarly, it would be important to communicate clearly which management team is accountable for its use and responsible for its results. As this example demonstrates, all of these elements of a fair process bring in and apply other AI Principles.

THE IMPORTANCE OF A MULTIDISCIPLINARY APPROACH

2.8. We can therefore see that achieving fairness and appropriately applying the AAAI Principles, requires a multi-disciplinary approach. For example:

2.8.1. Frontline business must be clear on the objective of the use of AI, the risks to individuals and to the business, and the extent to which risks of unfair treatment will be managed and explained to stakeholders.

2.8.2. Data scientists are central to the technical aspects of the use, testing and monitoring of AI.

2.8.3. Legal and Compliance need to be involved (including in any preliminary stages) to provide appropriate challenge, to oversee testing and to assist with fair process and related transparency principle.

2.9. Where firms are challenged about decisions, a likely approach is for a Complaints team to investigate, potentially supported by members of Legal and Compliance, and Risk. Traditionally that would involve speaking to the decision makers and understanding what matters were considered when decisions were made and analysing the reasons for the decisions taken in light of what policies, regulation and law require. However, such an approach would have clear limitations when considering decisions made by AI. The operation of an algorithm is often opaque, involving data sets too large for humans to assimilate and a lack of clarity over what factors were (and were not) considered. Colleagues from Complaints,

BOX 7:

In assessing the proportionality of the amount of data collected relative to the utility of the data, members also noted that AI systems built to collect wide or diverse data sets might inadvertently collect more than the firm had intended. This could create heightened risks for customers and employees. For example, monitoring employees via a webcam in order to meet compliance obligations in a 'working from home' environment could accidentally result in capturing the data of children or other family members. Similarly, AI analysing customers for marketing purposes might derive their age from the time they have held a product, or an employee productivity monitoring tool might gather health or other sensitive data if social media activity is monitored. Firms would need to consider such risks and take measures to minimise accidental or disproportionate data collection.

and Legal and Compliance teams can help ensure that systems are designed such that the firm will have the documentation and evidence on hand to demonstrate fairness, in the event of complaint or query from a regulator.

2.10. The involvement of data scientists and computer programmers to try and help and explain decisions would be clearly important as their understanding of coding and relevant models would be fundamental in any response. In particular, their assistance would be crucial if one was attempting to replicate the model's reasoning to explain a decision or to audit the outcomes of decisions (e.g. for signs of bias or error).

2.11. However, there would still be limits to what could be done after the event. It is important for potential issues to be considered in the AI design and implementation stages. Firms should work with front-office business lines to agree on appropriate objectives, identifying risks and how risks will be mitigated. The Risk function would likely have oversight over this process.

2.12. That is not to say that the role of Legal and Compliance would be diminished. Legal and Compliance should be involved at preliminary stages, including the analysis of balancing objectives against potential risks, and at the design stage of any AI tool to oversee testing before implementation. Gathering information through testing before implementation allows organisations to understand an AI tool's likely outcomes in common scenarios and identify any issues before problems arise. This requires Legal and Compliance to understand the AI tool and its design in sufficient detail to be able to meaningfully challenge.

2.13. In addition, upon receipt of challenges or disputes, the traditional methods of investigation, advice and clear communication will continue to be important. For example, it may be necessary to explain the intentions and objectives of the model, the type of model used, its parameters and inputs to the complainant. It may also be necessary to explain the testing and monitoring for bias or other undesirable outcomes carried out, both before implementation and as a result of a particular complaint.

3. USE OF AI INVOLVES TRADE-OFFS AND GREY AREAS

3.1. The use of AI to make decisions in relation to the activities of financial firms inevitably involves trade-offs between different interests and objectives. This is not unique to AI. The same trade-offs occur in any decision-making process.

3.2. However, the potential risks of the use of AI are greater in light of the enormous volume of decisions that can be made by AI processes very quickly and the "black box" nature of decision making by AI. In terms of outcomes, the "fair" use of AI therefore requires an understanding of the intended outcomes, why they are considered "fair" and appropriate monitoring to ensure that unintended outcomes do not occur. In terms of process, in addition to transparency and the ability for subjects of the AI process to contest its decisions, there should be appropriate consideration of the data inputs used in the AI process and what negative impact using them could have. This will involve cross-functional teams and adversarial inputs to test the performance of the AI against multiple criteria.[24]

3.3. For example, in the context of AI being used to make credit decisions, there is a balance to be struck between the type and amount of data that could be used by an algorithm – which may provide more reliable outcomes – versus the appropriateness of using such data for making a credit decision.

3.4. Consider information that is published by individuals on social media or information about how individuals use social media. In China, this information has already been used by companies to consider an individual's creditworthiness. In particular, it has been reported that certain firms consider social media use as a proxy for an individual's concern with their reputation and integrity. This informs the firms' assessment of the individual's creditworthiness, as do more traditional approaches to credit assessment involving new forms of commerce such as analysing internet purchasing history (including in-game purchases).[25]

3.5. There are potential benefits to this approach, assuming reasonable accuracy can be attained, such as providing credit to those who do not have a credit history. This was the reported motivation behind the approach in China, where it was used to seek to provide credit to those who had no credit history or access to formal channels of credit.

3.6. However, there are also potential negative aspects to using such data. There are clear risks of infringing on privacy for those who use social media as well as potentially poor outcomes for others in society who choose not to engage with social media. For example, it seems hard to see why an individual's credit score should be affected if they choose to communicate with others via email or the phone rather than social media.

3.7. Firms would need to consider such competing values and document why their AI processes are calibrated in certain ways and how potential risks or poor outcomes to others would be mitigated, for example by also making alternative services readily available.

3.8. It is important to look at the full customer journey and process when seeking to achieve this. For example, for corporate clients the AI tool could provide relationship managers with information about the AI process for them to consider as part of making a credit decision. This would enable the relationship managers to provide meaningful, informed human input into the decision-making process to mitigate the risks associated with purely automated credit decisions and potentially take the process outside of the scope of automated decision making to which various regulatory obligations are attached under the GDPR.

[24] See similar observations in the minutes from the first Artificial Intelligence Public-Private Forum, held 12 October 2020.

[25] *China P2P lender banks on social media*, FT 30 August 2015.

3.9. Avoiding looking at 'fairness' in siloes and also leveraging the framework under existing laws can help firms navigate these trade-offs. For example, the 'contestability and human empowerment' principle, combined with the requirement under GDPR to establish a 'legal basis' for processing under Article 6[26] could be helpful to firms wishing to mitigate certain risks and establish a "fair" process. In the example above, the firm could decide that the best 'legal basis' for accessing social media data to inform a lending decision is the consent of the individual. Obtaining consent under GDPR in this way would contribute towards establishing a 'fair' process because the firm would need to be very transparent about the data collection and use, and could only collect the data for individuals who choose to actively 'opt-in', rather than opting for an alternative lending product that does not involve social media. Achieving a fair process would of course also have other prerequisites, such as providing information to customers about how they can contest results of the AI process, amongst other things.

BOX 8:

Another set of use cases which pose difficult trade-offs are those where an individual's data is collected or processed for the good of the individual, for example in response to FCA guidance on identifying customers with characteristics of vulnerability (see 1.12 to 1.15). This could include monitoring a customer's general spending with AI in order to identify early signs that they might start struggling to make debt repayments. Such a use case requires balancing the customer's interest in avoiding financial difficulty against their right to privacy. One could imagine more extreme scenarios, where a firm might do transaction analysis to identify wider risks of harm, such as 'unwise spending' on alcohol or gambling. Without freely given customer consent or a clear regulatory requirement, such further steps may be more difficult to justify as fair treatment, although they could prevent the customer from suffering what the regulator or FOS might consider to be a poor outcome.

[26] See UK GDPR Article 6. See also footnote 10, above.

4. GUIDANCE ON THE USE OF AI

4.1. Consistent with previous surveys[27], workshops did not identify current laws and regulation as a barrier to the use of AI in financial services. However, the process of applying the fair use of AI principle to practical scenarios illustrated several areas where there is merit in exploring the need for further guidance.

4.2. How existing laws and regulation are applied to AI appears ripe for further guidance, especially where laws or regulation may vary between sectors or jurisdictions, and where there is overlap between different regimes.

BIAS AND DISCRIMINATION

4.3. It would be helpful to have guidance on how organisations should approach competing concepts such as the legal definition of discrimination within the Equality Act 2010 (based on the concept of equality of treatment and equality of outcome) with more academic notions of AI fairness such as demographic parity.

4.4. Such guidance could take the form of de minimis thresholds for the fair or compliant use of AI in different contexts. For example, guidance could provide an indication of the minimum level of testing required before implementation or more detailed guidance on the form that such testing is expected to take. Competition would encourage firms to seek more effective ways to ensure fairness but having such guidance as a starting point would help firms in applying existing legal and regulatory standards to the application of AI.

4.5. In addition, there is currently a lack of practical guidance to support firms on the lawfulness of bias mitigation techniques, so that they can understand what they can legally do. Such guidance would support firms' own work in using internal legal and data science expertise in interpreting legal requirements and how bias mitigation techniques work.

4.6. In particular, it would be helpful to clarify specific areas of uncertainty on how firms can lawfully approach situations where outcomes vary between different customer groups (particularly where differences in outcomes might arise out of wider societal bias or patterns of unfairness). Greater clarity is needed on whether it is possible to take steps to reduce such group differences, while avoiding direct discrimination against individuals in groups that might currently experience (higher rates of) better outcomes at present, perhaps due to wider patterns in society or as a legacy of having been being directly favoured in the past. This should avoid the result where firms have an obligation – whether from their own policies applying fairness principles or due to law or regulation – to monitor algorithmic bias risks, but then being unable to deploy proportionate methods to address any unfair bias or group differences that they find.

DATA QUALITY

4.7. As with many issues in AI, the monitoring of the outputs of algorithms and reviewing for unfair bias is also related to the quality of data put into the algorithm and/or which is used in model training or for later monitoring for unfair bias. Currently there is limited guidance on the standards of data that should be used by firms. While this could not be too prescriptive, given the range of data and uses, it would be helpful to inform the approach to data governance, development of AI tools and auditing of outcomes.[28]

4.8. The quality of data available to firms (particularly in respect of special categories of personal data) may in part be heavily dependent on the extent to which individuals, including customers or employees, voluntarily provide the relevant information. As mentioned above, this may present both legal and cultural challenges, particularly where firms seek to

[27] Such as the joint FCA and Bank of England 2019 paper *Machine learning in UK financial services*.

[28] See also minutes from the second Artificial Intelligence Public-Private Forum, held 26 February 2021.

gather such data in a consistent way across a number of jurisdictions. One alternative may be for firms to use synthetic data, in which case firms would benefit from guidance on the required quality of such synthetic data.

4.9. In this regard, the FCA has issued a call for input on the use of 'synthetic data' within financial services.[29] The regulator recognises that the utility and analytical value of synthetic datasets depends on the quality of the model and data used to generate them, and that there is a risk that biases may be replicated in the synthetic dataset, unless the generation process takes this into account. The FCA believes that high quality data sets could potentially be used to evaluate and compare AI decision-making by firms and as a tool to ensure that customers are being treated fairly.

LEXICON AND COMMUNICATIONS

4.10. There is a lack of an agreed lexicon regarding the use of AI which would assist firms explain, and customers and other stakeholders understand, AI's objectives, the decisions made and the outcomes of those decisions. Currently there can be a gap in understanding between those involved in developing and using AI and those who are subject to its decisions.

4.11. A range of different initiatives seek to assist firms in organising themselves in a way which maximises their ability to demonstrate their fair use of AI. In addition to the joint work of the ICO and the Alan Turing Institute on Fairness explanations (both process- and outcomes-based) referred to at 1.31, iTechLaw has, as part of its work on Responsible AI: A Global Policy Framework, published a 'Responsible AI Impact Assessment' tool designed to assist firms in measuring the impact of a proposed AI solution in measurable and quantifiable terms.[30]

4.12. As noted above, there can be two different audiences for AI information: non-experts, such as consumers, general employees or other 'data subjects', and specialists with technical expertise. However, to some extent, the gap between these audiences needs to be bridged. While there is concerted effort from model developers and industry bodies to construct ethical principles and work on how they could be applied in practical scenarios, this can lead to two levels of discussion. On one level there are high-level principles, which are easy to understand and follow for consumers. There is then another level of detail and complexity when model developers and data scientists seek to put these principles into practice.[31]

4.13. These different levels of discussion do not assist when seeking to explain the outcomes of AI processes – for example, when seeking to respond to a complaint. In such circumstances, guidance from regulators on expectations and more widely understood common terminology would assist. Without it, there is a risk of firms not being able to sufficiently explain the reasons for the outcomes of AI and/or consumers not understanding such explanations. In turn this can lead to suspicion of bias and AI being unfair, despite firms' best efforts.

GUIDANCE VERSUS STATUTE

4.14. While guidance would be appreciated, new hard regulation or laws on AI fairness do not seem to be required at this time, though gaps needing to be patched might emerge in due course. Statutes would struggle to keep pace with the change in technology and so should remain technology neutral. For that reason, guidance would help firms to apply existing rules to AI, while being more easily updated to keep pace with technology.

4.15. We recognise however that, given the increased focus on AI globally, it is likely that regulation and law will develop in various sectors and countries over time. There is a risk such new laws or regulations may conflict or lack coherence and thereby create

[29] Consultation available here.

[30] Their framework is available here.

[31] The distinction between the two is illustrated in the differences between the Monetary Authority of Singapore's Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics and the application of the principles to credit risk scoring and customer marketing by the Veritas Consortium.

further uncertainty for firms, stifling innovation.[32] This risk is heightened by the overlapping nature of equality, data protection and financial sector rules. Regulators could avoid duplicative compliance efforts by aligning regulatory requirements, or jointly issuing guidance. Ideally, authorities would also seek to collaborate at an international level. Incremental change, with careful consideration of how amendments and additions to the rules cohere with other existing laws and regulation would help avoid uncertainty and conflicts of laws.

CONFLICTING POLICY OBJECTIVES

4.16. Where policy imperatives for different regulators may from time to time come into tension or even conflict (see the discussion at 1.14 to 1.15 and in Box 8), it would be helpful for regulators to work together with a view to assisting firms in seeking to manage and resolve such conflicts. The Digital Regulation Cooperation Forum is a step in the right direction.[33]

[32] This has also been recognised in the Artificial Intelligence Public-Private Forum. Minutes from the first meeting, held 12 October 2020.

[33] See here for more information on the Digital Regulation Cooperation Forum.

5. CONCLUDING REMARKS

5.1. As described in this whitepaper, the exercise of seeking to apply the Fairness Principle to practical scenarios has helped refine our understanding of the fair use of AI. There are a wide range of considerations that must be taken into account, with important nuances unique to different scenarios and use cases. As such, building a comprehensive understanding of how to apply AI technology ethically, to the benefit of firms, customers and wider society, will involve an ongoing conversation between industry, regulators, civil society and policymakers.

5.2. We look forward to continuing to contribute to this debate.

UK Finance and Herbert Smith Freehills LLP

CONTRIBUTORS

HERBERT SMITH FREEHILLS

Karen Anderson
Partner

Jon Ford
Senior Associate

Miriam Everett
Partner

Sian McKinley
Senior Associate
(employed barrister)

UK FINANCE

Walter McCahon
Principal, Privacy and Data Ethics

The content of this report does not constitute legal advice and should not be relied upon as such. Specific advice should be sought about your specific circumstances.

Herbert Smith Freehills LLP

This report is intended to provide general information only and is not intended to be comprehensive or to provide legal, regulatory, financial or other advice to any person. Information contained in this report based on public sources has been assumed to be reliable and no representation or undertaking is made or given as to the accuracy, completeness or reliability of this report or the information or views contained in this report. None of UK Finance or any of their respective members, officers, employees or agents shall have any liability to any person arising from or in connection with any use of this report or any information or views contained in this report.