BAE SYSTEMS

Artificial Intelligence: Friend or foe for financial institutions?

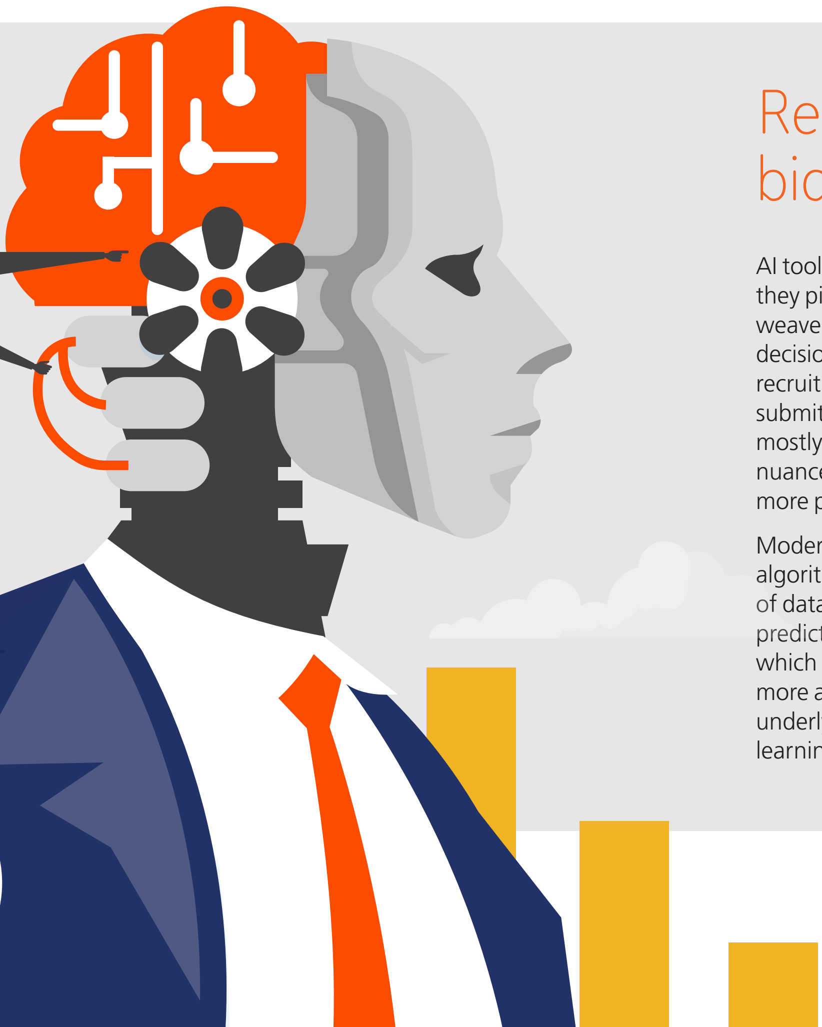
BANKING INSIGHTS

Introduction

The term artificial intelligence (AI) was first coined in 1955 by Herbert Simon and Allen Newell, who developed the first AI program that was used to prove 38 of the first 52 mathematical theorems, known as the 'Logic Theorist'. Since then, AI has become a 21st century phenomenon, with many industries eager to understand how it could be of benefit to them, none more so than the financial services (FS) industry when it comes to financial crime and fraud detection.

In the last few years, conversation has quickly turned to the consideration of fairness and the broader ethical issues associated with the use of AI (and algorithms in general). This follows a number of recently reported controversies in a variety of application sectors. In 2020, the UK saw tens of thousands of GCSE and A-Level students' exam results downgraded due to algorithmic bias in connection to their school's previous exam performance and other indicators. The police have also flagged concerns regarding bias in AI tools which can amplify prejudices and lead to some minorities being more likely to be stopped for street searches than others. Even the big tech firms haven't been completely spared from innate bias creeping into their algorithmic tools. In 2018, it was reported that one such firm had to scrap a 'sexist AI tool' used for recruitment that picked up on gender bias, and penalised women in its rating system.

When it comes to the FS sector, the need for fair and ethically sound AI systems is paramount. FS organisations are making increasing use of AI to inform critical decisions that can adversely affect their clients, such as the denial of loans and credit card applications. It is therefore important to ensure that any decisions generated by AI algorithms are not unfairly biased toward a particular demographic sector of the client population.



Reasons for biased AI

AI tools don't just become biased by themselves, they pick-up on innate human and societal bias weaved into data sets, algorithmic setups and decision outcomes. Returning to the aforementioned recruitment example, the tool was trained on data submitted by applicants over a 10-year period, which mostly came from men. The system picked-up on this nuance, and taught itself that male candidates were more preferable.

Modern AI solutions are driven by machine learning algorithms that predict outcomes for new instances of data. The investigated outcomes of those predictions are fed back into the learning process, which therefore adapts to new data and becomes more accurate over time. However, if there is underlying bias in the data or outcome decisions, the learning process can amplify this bias.

Controlling AI bias

As the world becomes more dependent on AI-based algorithms for making a wide range of decisions, business leaders will be tasked with explaining the reason for inference or predictions generated by those algorithms, also known as 'AI explainability'. So, this leads us to ask: what controls can be put in place to ensure the decision outcomes underpinned by AI and machine learning algorithms are unbiased and fair?

An easy solution at first sight would be to remove protected variables in the input data, such as gender and race, but it's unfortunately not as simple as this due to their correlation with other variables. For example, in the US there is a perceived correlation between zip code and race, which has led to issues with predictive policing applications of AI. The impact of protected variables can proliferate through data sets through their correlation with proxy variables, which can confound efforts to remove bias if not brought under control.

Controls for mitigating bias can be inserted at three stages within algorithmic workflows:

1. Pre-processing controls resample and reweight data to mitigate the effects of bias before it enters the modelling stage and is used to inform a decision
2. In-processing controls apply modifications to the formulation of the learning algorithms, such as adaptations to their cost functions, to reduce biased outcomes
3. Post-processing controls are applied after the algorithmic/modelling stage, by applying de-biasing controls to their decision outcomes

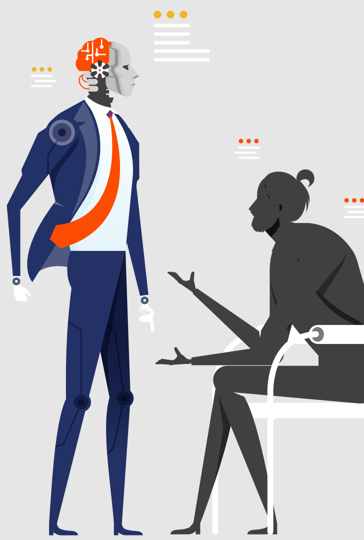
Financial institutions will need to have strong model risk governance practices in place to monitor these interventions. Research has shown that controls to reduce bias can have an adverse effect on accuracy, and vice versa. This has led to the formulation of solutions that attempt to optimise some balance or compromise between accuracy and fairness metrics. What compromise or tolerance is acceptable is a strategic decision for ultimate consideration by the ethics board within a financial institution.

The role of regulators, governments and big tech firms in ethical AI

Beyond data processing, financial regulators such as the Monetary Authority of Singapore have issued guidelines to strengthen the culture of responsibility and ethical behaviour within financial organisations, placing emphasis on the individual role each organisation has to play in developing an ethical culture.

The UK Government's Centre for Data Ethics and Innovation also released a report into bias in algorithmic decision making in November 2020, which looked at bias specifically in the recruitment, financial services, policing and local government sectors. The aim of the report was to make "cross-cutting recommendations that aim to help build the right systems so that algorithms improve, rather than worsen, decision-making".

But, this is an issue that is further reaching than regulatory and government bodies. Big tech organisations have a role to play in informing safeguards and policy development. For example, Microsoft created the FATE community group to define the fairness, accountability, transparency and ethical principles and processes in AI. The group studies "the complex social implications of AI, machine learning, data science, large-scale experimentation, and increasing automation", with the aim "to facilitate computational techniques that are both innovative and ethical, while drawing on the deeper context surrounding these issues from sociology, history, and science and technology studies".



However, it is worth noting that the current safe guards and policies created by such groups and government bodies acknowledge their thinking is preliminary. The reality is that ethics in AI is a relatively new concept that has grown in prominence over the last three to four years and is a concept we are all still trying to fully understand, unpick and control. Current safe guards and controls need to be fully worked through FS use cases and tightened up over time, becoming a cornerstone of FS ethics principles.

In summary, AI and its underpinning machine learning algorithms can be a game-changer for financial institutions, driving significant detection improvement. However, it can also amplify bias that may enter at multiple points during the journey from data, through modelling, to decisions. Organisations must lean on strong model governance principles and practices to recognise and remediate the effect of such bias, which could lead to unfair or unethical outcomes if left uncontrolled.

We are

BAE SYSTEMS

At BAE Systems, we provide some of the world's most advanced technology defence, aerospace and security solutions.

We employ a skilled workforce of 82,500 people in over 40 countries. Working with customers and local partners, our products and services deliver military capability, protect people and national security, and keep critical information and infrastructure secure.

Global Headquarters
BAE Systems
Surrey Research Park
Guildford
Surrey GU2 7RQ
United Kingdom
T: +44 (0) 1483 816000

BAE Systems
8000 Towers Crescent Drive
13th Floor
Vienna, VA 22182
USA
T: +1 720 696 9830

BAE Systems
19, Boulevard Malesherbes
75008 Paris
France
T: +33 (0) 1 55 27 37 37

BAE Systems
Mainzer Landstrasse 50
60325 Frankfurt am Main
Germany
T: +49 (0) 69 244 330 040

BAE Systems, Surrey Research Park, Guildford, Surrey, GU2 7RQ, UK

E: learn@baesystems.com | W: baesystems.com/bankinginsights

linkedin.com/company/baesystemsai

twitter.com/baesystems_ai