

## Jumping into AI & big data FEAT first

Why companies must consider Fairness, Ethics, Accountability & Transparency when implementing AI programmes

“Your scientists were so preoccupied with whether or not they **could** that they didn’t stop to think if they **should!**” This iconic line delivered by Jeff Goldblum’s character in 1990s blockbuster movie *Jurassic Park* could be applied to every major technology innovation since the industrial revolution; change is disruptive, for better or worse.

It feels especially relevant to the fast-developing impact of AI and big data. The astonishing potential of AI and big data spurs us to greater enthusiasm for pushing the boundaries to develop commercial, economic, social and environmental applications that could solve challenging problems—both

for business enterprises and the world. As with any new technology, however, we must consider the wider impact of mass deployment of AI on individuals and society.

The challenge of safely adopting disruptive technology was a key theme at this year’s World

Economic Forum. Participants were asked to consider that “... the breakneck speed and sheer scale of this round of technical change is something else—it threatens the very definition of what it is to be human [...] Should we just slow down a bit?”<sup>1</sup>

<sup>1</sup> The World Economic Forum. “Tech for Good.” January 21-25, 2020. Accessed at: <http://bit.ly/2PlckeF>



**AI needs governance around how we mitigate bias, allocate responsibility and promote public accountability. But first and foremost, it requires a decision about whether an algorithm should or should not be used. The fact that something is technologically possible does not mean that it should be done.”<sup>2</sup>**

**Ivana Bartoletti, Technical Director and Data & Privacy professional, Deloitte**

There is strong movement to recognise the importance of pursuing AI innovation and application within a moral, legal and regulatory framework that promotes the principles of fairness, ethics, accountability and transparency (FEAT).

**This white paper examines why the FEAT principles are important, what initiatives are under way to support FEAT in AI and big data, and what businesses need to know to respond to this challenge.**

## The challenges inherent in AI and where FEAT fits in

Real-world artificial intelligence has been a long time coming. Its approach has been characterised by a cycle of “false dawns” followed by “AI winters” when unmet expectations led to reductions in funding and research. When AI finally did go mainstream, the reality looked quite unlike the imaginings of sci-fi writers. Most AI applications in use today are comparatively mundane. They are algorithms built to do heavy data lifting to enable better decision-making and to drive operational efficiencies. It is a far cry from humanoids endowed with sentient intelligence that have dominated popular culture for decades.



Sci-fi culture, however, has driven awareness of the potential risks of AI. Although those risks are considerably less dramatic than rogue AI robots taking over the world, they are arguably more insidious. As AI automation

proliferates through systems used to make decisions affecting people’s lives, the data on which it is based needs to be accurate and the algorithms that use it fair, unbiased and ethical.

<sup>2</sup> Bartoletti, Ivana. “What are the challenges of using Artificial Intelligence?” *The Telegraph*. March 25, 2019. Accessed at: <http://bit.ly/32kW2I9>

## Data bias

An advantage of using machines to make decisions would appear to be the removal of human bias from decision-making. However, while automated decision-making may eliminate the bias of a specific human decision-maker, it exposes the subject of the decision to the risk of big data bias.

AI is trained on vast amounts of historical data to help reach decisions. However, that data may contain bias. There is a significant risk that this bias will become institutionalised. A recent example is Amazon's AI-powered recruitment engine, which was found to be penalising women applicants for technical roles. The algorithm had been trained on resumes from the previous ten years, which reflected male dominance in the industry. The system concluded that male applicants were preferable, penalising resumes featuring the word "women" and downgrading those that referenced all-women colleges.<sup>3</sup> Another oft-cited example is the COMPAS criminal justice software used to predict a prisoner's likelihood of re-offending and suitability for parole.<sup>4</sup> This system was found to be unfairly biased against African American prisoners.

Identifying and eliminating bias to make data fair is a complex challenge that must be addressed if AI is to be used to make critical decisions.

## Individual privacy rights

The automated collection of personal data poses serious issues around individual privacy and information control. This is an area where ethical considerations have sometimes trailed technological capability. Consider the use of facial recognition technology employed by law enforcement bodies. Last year, UK police forces halted facial recognition trials being used to search for missing persons or known criminals. The issue? Privacy concerns raised by the collection of facial image data in public without the knowledge and consent of passers-by.<sup>5</sup> There has been significant concern from privacy protection

groups at the potential invasion of rights caused by "Big Brother" surveillance.

This issue is not limited to images. A 2018 briefing from analysts McKinsey suggests the spread of a contagious virus could be tracked by using Natural Language Processing (NLP) to analyse the content of text messages sent in the affected regions.<sup>6</sup> While this seems a valuable use of big data analytics, does the analysing organisation have the right to access and analyse private messages that could identify the individuals involved? At what point does public risk outweigh individual rights to privacy?

Resolving these ethical and accountability issues to protect individual rights to privacy and data protection is central to legitimately using AI and big data.



<sup>3</sup> Dastin, J. "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters. October 20, 2018. Accessed at: <https://reut.rs/2vXLG4M>

<sup>4</sup> Richardson, Rashida and Schultz, Jason and Crawford, Kate. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York Law Review*. February 13, 2019. Accessed at: <http://bit.ly/2wJENEH>

<sup>5</sup> Townsend, M. "Police forces halt trials of facial recognition systems." *The Guardian*. August 17, 2019. Accessed at: <http://bit.ly/3a3WXPB>

<sup>6</sup> Manyika, James and Bughin, Jacques. "The promise and challenge of the age of artificial intelligence." McKinsey Global Institute. October 2018. Accessed at: <https://mck.co/2HPv2a7>



Lack of trust in AI and big data is a real risk for businesses investing in the technology, but few organisations are actively addressing it. McKinsey's most recent survey into AI adoption found that "while 39% of respondents say their companies recognize risk associated with 'explainability' (the ability to explain how AI models come to their decisions), only 21% say they are actively addressing this risk."<sup>8</sup>

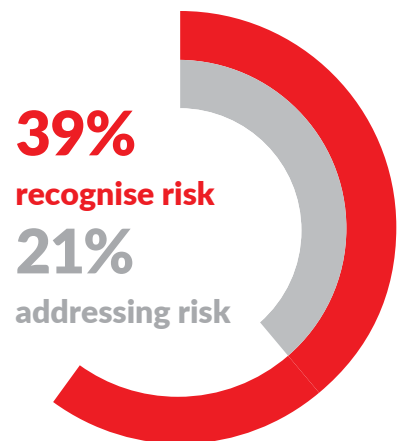
Trust in AI can only be achieved if there is confidence both in the data on which it bases its analysis and the decision-making process behind the analytics.

## Explainability

A practical barrier to the widespread adoption of AI and big data analytics is the challenge of explainability—sometimes called the "black box" problem. Teradata CEO Stephen Brobst summarised this in an interview with Bernard Marr, saying, "With multi-layer neural networks, the average human doesn't understand, so now we're making predictions based on things that people don't understand and that's going to make people uncomfortable."<sup>7</sup>

A decision made by AI may not coincide with human intuition but, as Brobst explains, the complexity and volume of data inputs that contribute to an automated decision mean its rationale, while sound, may not be explainable in a format accessible to humans.

This leads to a lack of trust in AI decisions. Research by McKinsey indicates that only 16% of employees trust AI-generated insight. This causes problems in scenarios where decisions have high human impact, such as health diagnoses and finance.



<sup>7</sup> Marr, Bernard. "The Biggest Business and Social Challenges for AI." Bernard Marr & Co. Accessed on 2/20/2020 at: <http://bit.ly/32rdURx>

<sup>8</sup> "Global AI Survey: AI proves its worth, but few scale impact." McKinsey & Company. November 2019. Accessed at: <https://mck.co/2w3lqVn>

# Take a closer look at FEAT principles

## Fairness

It is often said that Artificial Intelligence is only as good as the data it uses. However, even when the data is fair, if the algorithm analysing it has been built with bias, the decisions it makes will be skewed. Organisations need to consider fairness in both big data selection and algorithm design.

David Weinburger, writing in the *Harvard Business Review*, describes bias as “machine learning’s original sin,” noting that it picks up all the historical bias contained in the human-generated data it uses.<sup>9</sup> These include well-known biases such as gender or race, but also contingent biases that may occur as a result of multiple factors in combination that are less easy for a human—or a machine—to detect.

“Cleaning” data to remove bias and promote fair decision-making is a complex challenge that is a very active research area among computer scientists. Work is ongoing to develop fair data classifiers and design algorithms to check for fairness. In support

of this, last year Amazon made grants of up to \$10M available to researchers exploring how to ensure fairness in AI. From a company pioneering consumer-facing AI, that is an indicator of the importance it places on fairness.

Businesses that deploy machine learning algorithms for automated decision-making must keep abreast of developments in data fairness and ensure data used to train algorithms is as balanced as possible.

- Pre-processing to eliminate bias in the original dataset
- In-use processing to mitigate for bias that emerges during a specific application
- Post-processing monitoring of outcomes

Humans design AI algorithms, so these algorithms are also vulnerable to human bias. An example might be bias contained in the choice of decisions that an AI is asked to make. If an option is excluded as a result of programmer bias an individual or group could be disadvantaged.

## Regulating fairness in AI

The topic of fairness in AI is still a new discipline. However, it is receiving considerable attention from governments and international bodies. In 2019, the Algorithmic Accountability Act was proposed in the U.S., which would require large corporations to audit machine learning-powered systems and assess them for bias, discrimination, security and privacy risk.<sup>10</sup>

The European Union is also exploring policy options to address algorithmic transparency and accountability. In early 2020, MEPs passed a resolution addressing challenges associated with the rapid evolution of AI, among which was the edict that automated decision making systems “... should only use high-quality and unbiased data sets and “explainable and unbiased algorithms” to boost consumer trust and acceptance.”<sup>11</sup>

Despite this increasing scrutiny and the likelihood of regulation, McKinsey’s 2019 AI survey found that just 26% of organisations recognised the risk that fairness poses in AI deployment, and only 13% were actively working to address it.<sup>12</sup>

<sup>9</sup> “Big data expertise seals innovation partnership”, [www.cardiff.ac.uk](http://www.cardiff.ac.uk), published November 6, 2019

<sup>10</sup> Robertson, Adi. “A new bill would force companies to check their algorithms for bias.” *The Verge*. April 10, 2019. Accessed at <http://bit.ly/2veZQ1g>

<sup>11</sup> “Artificial intelligence: EU must ensure a fair and safe use for consumers.” European Parliamentary Research Service. January 23, 2020. Accessed at <http://bit.ly/2PnPIdl>

<sup>12</sup> “Global AI Survey: AI proves its worth, but few scale impact.” McKinsey & Company. November 2019. Accessed at: <https://mck.co/2w3lqVn>

# Ethics

The ethics of AI and big data are fascinating and complex. To what degree should someone's personal information be used by third parties to make decisions affecting their lives? Even if a decision is ostensibly in a person's best interests, can or should it be made unilaterally? What kinds of decisions must remain the preserve of humans and when might artificial intelligence encroach on the concept and practice of free will?

There is the issue of conflicting rights, such as with the contagious disease example: at what point does community safety outweigh an individual's right to privacy? And should companies be developing AI in sectors such as the military? These are just a few illustrations of the many questions that the ethics of AI and big data raise.

One of the challenges lies in the classic "slippery slope" argument: What starts as valid use of AI, through unintended consequences, unrecognised bias or new applications, becomes a threat when it proliferates. But the point at which this transition from benign to malignant occurs varies depending on individual and group attitudes.



This risk and the complexity of ethically managing AI is leading businesses to establish and fund think tanks and research centres to explore these issues and consider frameworks in which AI and big data use can be safely developed.

Salesforce appointed a Chief Ethical and Humane Use Officer at the start of 2019 with a remit to develop a strategic framework for the responsible use of technology across the business. Similarly, Microsoft has identified six ethical principles to guide the development and use of AI, established research group "to study the complex social implications of AI, machine learning, data science, large-scale experimentation, and increasing automation."<sup>13</sup>

While this demonstrates that businesses recognize their obligations around AI, data privacy and protection expert Ivana Bartoletti points out that independent governance is needed, noting, "The companies dominating the digital sphere

are, right now, the very same organisations leading the debate around its ethics. There is a lack of accountability and, more importantly, a lack of democracy in who decides what the ethics are that shape the world around us."<sup>14</sup>

Addressing this potential bias in framework development are new national institutions such as the UK Centre for Data Ethics and Innovation. The Centre is charged with exploring risks and opportunities, establishing best practice for responsible use of data and AI and advising on the need for laws and regulations.

At an industry level there are also moves to support ethical implementation of AI—particularly in those sectors where AI has immediate potential and customers must be protected, such as finance and healthcare. The Monetary Authority of Singapore has formally adopted FEAT principles, for example, and in the UK the establishment of an NHS AI Lab in late 2019 was prefaced with an in-depth exploration of AI governance.<sup>15</sup>

<sup>13</sup> "FATE: Fairness, Accountability, Transparency, and Ethics in AI." Microsoft | Research. Accessed on 2/20/2020 at: <http://bit.ly/2umyHJo>

<sup>14</sup> Bartoletti, Ivana. "What are the challenges of using Artificial Intelligence?" *The Telegraph*. March 29, 2019. Accessed at: <http://bit.ly/37Tdc7f>

<sup>15</sup> "Artificial Intelligence: How to get it right." NHS. Accessed on 2/20/2020 at: <http://bit.ly/3a1vMFj>

# Accountability

The fast pace of AI deployment and lack of controls have led some commentators to refer to it as a “Wild West” landscape. However, accountability is essential to building public trust in automated decision-making and times are changing. Recent developments in privacy legislation have raised public awareness of rights and responsibilities and this, together with growing government and regulatory interest, looks set to act as a restraint on the more concerning possibilities of AI deployment.

The end of the last decade witnessed a significant shift in the understanding of the degree to which businesses are accountable to individuals over personal data collection, use and protection. Legislation such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) requires businesses to take responsibility for the personal data they collect, their justification for holding it and how they protect it. It also gives individuals the right to obtain information about their data and force companies to delete it if requested. This has implications for any business that uses

personal data in AI-powered processes as the subject must specifically consent to this use in advance.

However, AI accountability must go beyond basic authorisation to collect and use data. Research from the ICO found that more than 50% of people are concerned about machines making automated decisions about them.<sup>16</sup> If trust is to be built, the interface between humans and AI needs careful management. Therefore, when AI makes consequential decisions there must be mechanisms available enabling people to understand how they have been reached and to challenge decisions—the black box problem notwithstanding. This means providing recourse to human oversight and devoting resources to supporting AI accountability.

As the AI environment matures, expect more legislation and regulation governing business and public sector use of AI. As with development of global financial and privacy regulations, organisations will have to comply with rules specifically relating to AI and big data, requiring governance frameworks, monitoring protocols and risk management strategies to drive visibility and accountability.

With one AI bill already under way in the U.S. and the EU paying close attention to ethics and fairness, it is a matter of time before accountability and consumer protections come into force.

# Transparency

Accountability cannot exist without transparency. Organisations must be open about how they use AI algorithms and big data and the effect of this use. However, being sufficiently open about technology can be difficult for organisations that are staking their competitive edge on AI innovation and typically adopt a highly secretive and protective approach. Transparency that eliminates advantage is unlikely to be embraced by those at the cutting edge of AI.

There are also risks to full AI disclosure. Writing about AI’s “transparency paradox” Andrew Burt, chief legal officer at data governance specialist Immuta, notes, “Explanations can be hacked, releasing additional information may make AI more vulnerable to attacks, and disclosures can make companies more susceptible to lawsuits or regulatory action.”<sup>17</sup> Burt goes on to underline that the more information revealed about an algorithm, the more damage a malicious actor can do. However, he acknowledges that this is an issue already dealt with in the realms of cybersecurity.

Despite these difficulties inherent with transparency, its importance to trust and the moral use of AI means organisation will be forced to address it.

<sup>16</sup> “Information Rights Strategic Plan, Trust and confidence.” The ICO. July 2019. Accessed at <http://bit.ly/38QPvr0>

<sup>17</sup> Burt, Andrew. “The AI Transparency Paradox”. Harvard Business Review. December 13, 2019. Accessed at <http://bit.ly/2Ve32oJ>

# FEAT for the future

It has always been true that technology is neither good or bad but can be applied for both good and bad purposes. It is therefore positive that, as AI becomes more mainstream and enters the public consciousness issues of trust, fairness and accountability are being addressed.

Fairness, ethics, accountability and transparency in AI and big data—the FEAT principles—are intrinsically linked and central to the development and successful implementation of AI and big data. Adopting FEAT principles in the early stages of AI implementation also delivers benefits for businesses:

- It helps address the black box problem by building trust from employees and citizens that data and algorithms are accurate, fair and ethically applied.
- AI performs better when the data it uses is robust and fair, ultimately driving better decisions for the organisation.
- It puts the business in a better position to respond to the looming likelihood of legislation and regulation of big data and algorithm use. By building FEAT principles into deployments from the ground up, the compliance burden will be mitigated.
- It helps protect the business from CSR and reputational risks arising from misuse of data and AI.

AI is here to stay, and its use is accelerating. Nobody can ensure that its use is only ever for positive reasons. However, the leadership demonstrated by many larger companies in addressing the FEAT principles demonstrates a commitment to addressing ethical issues in parallel with the wider deployment of AI.

As businesses and governments explore its potential to solve complex global challenges and advance human understanding, it is important that they are good custodians of the power that AI confers. Taking a FEAT-first approach is a critical step to building trust, confidence and acceptance of AI as a force for good.

## About Nexis® Data as a Service

With an unmatched global content collection enhanced through normalisation and metadata enrichments, Nexis® Data as a Service (DaaS) delivers the highly relevant, archival and current data organisations require for a broad range of artificial intelligence applications.

### Learn more

Talk to a Nexis Data as a Service specialist to arrange access to our developers' portal, sample files, and sample schema

 [LexisNexis.com/internationalsales](https://www.lexisnexis.com/internationalsales)

 +31 20 485 3456

 [information@lexisnexis.com](mailto:information@lexisnexis.com)

LexisNexis, Nexis and the Knowledge Burst logo are registered trademarks of RELX Inc. Other products or services may be trademarks or registered trademarks of their respective companies. ©2020 LexisNexis. All rights reserved. SEG-DaaS-BigData-FEAT-First 0320

