# Artificial Intelligence (AI)

As Artificial Intelligence advances at a rapidly increasing pace, so does the need for ethical understanding and awareness of the developers of such technologies. Responsible AI is the practice of ensuring that AI systems are designed with the intention of optimising decision-making and ultimately augmenting the human experience with a focus on meeting societal and legal standards, providing transparency in the decisions made by the systems, and ensuring privacy and security of the data. By developing explainable AI, where the route taken by AI systems to reach decisions is explicitly explained, the user can be made aware of any potential biases within this decision-making and decide whether this meets their personal ethical expectations. With this transparency in the development of such systems, users can be confident of what is being done with their data and who it is available to, and can be certain that they understand the processes taken to reach decisions, eliminating any concerns of malicious intent or predetermined prejudices in the systems.

There are many examples of AI being used maliciously or failing due to bias. Bias in AI systems can stem from unconscious biases in the engineers or from their intention to create accurate models with the available data and little consideration for the wider context and impact of the results, however often we are faced with the problem of data or algorithmic biases. Data bias refers to biases formed by the AI system based upon the training data it is given. Often these data sets are unbalanced and reflect societal biases, feeding into the system a restrictive perspective resulting in the AI learning biased patterns and making assumptions based upon these patterns. Algorithmic bias refers to the amplification of biases in data due to algorithms within the system.

An example of algorithmic and data bias is Amazon's attempt in 2018 to speed up their recruitment process by automating their candidate selection. It was found that the system favoured white males, since the training data was based on their current engineering employees, who were predominantly white and male. Google faced similar problems with its translation services, which when translating from an ungendered language, would automatically refer to all doctors as male and nurses as female, an assumption generated from the analysis of large sets of written data reflecting societal biases [1]. Further to this, there are many examples of facial recognition software being racially biased, including Amazon's Rekognition which incorrectly matched 28 members of Congress to a database of criminal mugshots and of which 39% of the falsely returned matches were people of colour who made up only 20% of the sample tested. Facial recognition software by Microsoft, IBM and Face++ also proved to be less successful in analysing the faces of people of colour than lighter-skinned subjects. Though Amazon refused to submit their AI systems for Rekognition to the National Institute of Standards and Technology, making it difficult to pinpoint the exact reasons for its failures, the latter three tech giants tested their AI on two large facial databases, IJB-A and Adience, which both meet

industry standards despite 79.6% of IJB-A's and 86.2% of Adience's data set containing light-skinned subjects [2].

Aside from societal biases influencing AI systems, there is also the ethical concern of what certain AI systems are being used for and what is being done with the information they produce. In 2016 Cambridge Analytica obtained data from 87 million Facebook profiles and used it as a political weapon, creating psychological profiles and targeting ads at voters whose opinions could likely be swayed in favour of the Republican Party, ultimately leading to the success of Cruz and Trump in the 2016 elections. Prior to this becoming public knowledge, in 2017 Facebook launched a "suicide detection algorithm" which predicts the mental state of users based on data gathered without their consent from their posts and determines from this their inclination to commit suicide [3]. Genetics testing company 23andMe offer to analyse DNA in order to provide healthcare information tailored to the individual, however their Biobanking Consent Document does not clearly or explicitly detail what DNA information is stored or how it is used, just that it can be stored for up to 10 years - unless they notify you otherwise. These companies legally sell data to pharmaceutical and biotechnology firms, and while this can advance drug development and lead to medical breakthroughs, the ambiguity in their legal documents provides the company with a considerable amount of freedom in what they choose to do with the data.

As discussed, the implications of AI and data misuse can have societal and political consequences on a global scale. Organisations which hold our data have the power to influence us without our awareness through media we are subsequently exposed to, and AI has the power to amplify societal bias and further embed it into our culture. In order to avoid such failures of AI, companies and individuals developing these technologies must take responsibility and be held accountable for the outcomes of the AI systems they create.

Engineers should ensure that training data is relevant, cleaned and verified and that the data is understood before a model is trained so that sources of bias, as well as factors influenced by these biases, can be identified and eliminated. To avoid data and algorithmic biases, it is possible to train models to identify when it learns a bias and to penalise it for doing so, resulting in fewer biased outcomes. The EU's General Data Protection Regulations (GDPR) mandate that organisations operating in this region must explicitly inform users of the intended use of their data and have their consent before collecting sensitive data. To uphold standards of security and privacy, organisations operating outside this region should maintain a similar level of transparency, as should the engineers developing the AI systems to ensure the system is not subject to personal bias. The GDPR also gives people the right not to be subject to decisions solely based on automated processes which result in legal or similarly significant effects [4].

Above all, it is important to remember that AI systems are tools and are

not a sufficient replacement for human reasoning. While AI is an excellent implementation for completing laborious tasks and repetitive routines, complex problems which have non-trivial solutions are likely to lead to lots of system errors. The more serious the decisions being made by AI, the more serious the problems it causes, and so engineers of AI should be ethically aware of biases in data and of the issues surrounding the outcomes.

## Business Challenges of Implementing AI Solutions

AI can transform a business by saving a company time and money, and offering valuable insights. The main challenges an organisation is faced with when implementing AI strategies can be categorised into three branches: Talent, Time and Trust. Business leaders must address these challenges in order to successfully utilise AI strategies.

The challenge of talent refers to the growing demand of data and AI specialists and the lack of individuals with the adequate technical skills required to fill these roles. 56% of technology professionals agree that this is the biggest challenge companies currently face when adopting AI strategies [8]. Companies should work to develop a culturally collaborative and data-driven environment to promote the effective use of AI, keeping in mind that the process of developing such strategies requires a constructive team of individuals with different skill sets and technical abilities. A key factor in successful teamwork is ensuring effective communication, so that all individuals understand the constraints and desired outcomes and can contribute to delivering relevant insights and solutions. It is also crucial that the data is presented in a way which is accessible to all appropriate members within a company to encourage collaborative success.

Although implementing AI strategies will ultimately save a company time, the time taken to develop and deliver those strategies is one of the challenges companies face. Metrics should be identified in order to measure the success of implementing AI within a business to ensure goals are being reached. Again, communication within a team is of the utmost importance when faced with this challenge as it is crucial that the goals of the company can be clearly translated into a business-centred problem where the outcomes from the AI systems can be determined and interpreted. To directly tackle the constraint of time, insights and solutions can always be obtained with the current available data sets, and models can later be retrained on updated data. As these technologies develop, it is encouraged to share insights and knowledge with those in the wider machine-learning community in order for AI to grow and for businesses to gain insights on how they can further use this technology to their advantage.

Finally, Trust in AI solutions can challenge companies adopting these strategies. Transparency in developing and using these systems is important in ad-

dressing this problem, as is the ability of those within the business to coherently explain how the AI arrives at its decisions and predictions. The explainability of a model is crucial and companies working with this technology should thoroughly consider the most ethical and efficient ways in which they can gather and save data, as well as how it can be made accessible.

# References

[1] Sam Genway, *Three Causes of AI Bias and How to Deal with Them*, 2020. [Online: https://www.tessella.com/insights/three-causes-of-ai-bias-and-how-to-deal-with-them]

[2] Noah Blier, *Bias in AI and Machine Learning: Sources and Solutions*, 2019. [Online: https://www.lexalytics.com/lexablog/bias-in-ai-machine-learning]

[3] Andrea Kulkarni, *AI in Healthcare: Data Privacy and Ethics Concerns*, 2021. [Online: https://www.lexalytics.com/lexablog/ai-healthcare-data-privacy-ethics-issues]

[4] *UK General Data Protection Regulation (GDPR)*, Article 22(1), 2018. [Online: https://gdpr-info.eu/art-22-gdpr/]

[5] Andrew Patel, *Malicious Use of AI*, 2019. [Online: https://blog.f-secure.com/malicious-use-of-ai/]

[6] Dominic Delmolino and Mimi Whitehouse, *Responsible AI: A Framework for Building Trust in Your AI Solutions*, 2018. [Online: https://www.accenture.com/acnmedia/PDF-92/Accenture-AFS-Responsible-AI.pdf]

[7] John Spooner, *How to effectively deliver an AI transformation strategy*, 2020. [Online: https://www.itproportal.com/features/how-to-effectively-deliver-an-ai-transformation-strategy/]

[8] Ingrid Burton, *What Business Leaders Need to Know About AI*, 2019. [Online: https://builders.intel.com/ai/blog/h2o-ai-business-leaders-ai]