CASE STUDY ANALYSIS

This case study will outline the ethical considerations and legal obligations that must be considered during the project’s development.

**Ethical**

For ethical dilemmas, the vast majority will lie within the AI chatbot project. Due to the lack of a human controller, there must be guarantees that the automated process is acting fairly and accurately.

AI

There are growing worries of inbuilt bias 1. (Bias, no date) “The action of supporting or opposing a particular person or thing in an unfair way” grows exponentially due to the increased use of automated systems. Along with evidence that such systems can be subject to bias **5. (New study shows AI chatbots reflect human biases, 2023)** “A new study co-authored by Dr Joe Stubbersfield, Senior Lecturer in Psychology at the University of Winchester, shows that large language models (LLMs) such as GPT-3 often reflect human biases” customers may feel morally obligated to avoid the use of such systems that discriminate based immutable characteristics. For in our society, immutable characteristics are seen as unfair and immoral to judge a person.

Examples of unfair discrimination

* Microsoft’s chatbot (named Tay) was released on twitter and **2. (Tay, 2016**) "As a result, Tay tweeted wildly inappropriate and reprehensible words and images.”
* **3. (Amazon's sexist hiring algorithm could still be better than a human, 2018)** “Amazon decided to [shut down](https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G) its experimental artificial intelligence (AI) recruiting tool after discovering it discriminated against women”

Though there are ethical ways to use personal date to provide more effective services. The business must ensure high ethical standards beforehand.

Transparency

Transparency will be defined in this case as **4. (Ethics guidelines for trustworthy AI, 2019) “**AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned.” It is the measurement of honesty and openness the business enacts concerning its products and services. This contains the algorithms source code itself, the training data used, and the reasoning behind the choices, and the output of the services (explainability of the answers the chatbot provides). Lack of transparency will result in stakeholders losing trust in the business and its utilities. This may result in losing business overall. Additionally, without the business being clear on its services, it allows new outlets and independent journalists to speculate on the businesses’ internal functions. For many surveys, they class **3. (Leading Business Problem #3: Lack of Transparency, No Date) “**lack of transparenc**y** as the third leading business problem businesses are facing today.”. For example, in some businesses **3.**” 82% of employees don’t trust their boss to tell the truth”.

For being honest and open is also vital for a healthy workforce in the business itself. Lack of may result in decreased productivity and poor staff retention.

SOLUTIONS

There are many solutions to these issues. For the issue of unfair bias and lack of transparency, these solutions can cover most concerns.

* Fairness metrics. Evaluating response quality according to criteria such as: Relevance, accuracy, completeness, Clarity, politeness … analysing patterns between customers characteristics and chatbot responses.
* 3rd party verification. Impart partials to test claims of fairness and transparency.
* Open-source AI code, maximizing transparency.
* Follow the public guidelines from the Responsible Tech Adoption Unit (RTA, previously known as CDEI) and many institutions such as the Alan Turing institute, The British Standards institute, National physics lab on how to ethically maintain AI projects.
* Public accountability, like Microsoft **2.** “We take full responsibility for not seeing this possibility ahead of time."

With these steps taken, the businesses capacity of explainability **6. (Explainable AI, no date)** “is the concept that a machine learning model and its output can be explained in a way that “makes sense” will increase dramatically. The ability to justify the Ais output will increase the business ethical and transparent reputation.

**LEGAL**

For the legal side, much of the focus will be adhering to the DPA 2018 (Data protection Act). Focusing on how the key principles **7. (Data Protection Act 2018 Summary, 2014)** “Fair, lawful and transparent processing. Purpose limitation, Data minimisation, Accuracy, Data retention periods Data security and accountability.” are applicable to these projects. Though the UK GDPR and Equality act 2010 would also be vital to consider during development, we will be focusing solely on the DPA 2018.

AI

**8. (At a glance, no date) “**Where AI uses personal data it falls within the scope of this legislation. This can be using personal data to train, test or deploy an AI system.” Though the DPA 2018 does not specifically mention AI, it does show focus on 8.” the user of profiling and automated decision-making”.

Here are just a few examples of from the DPA that should be considered during development. All underlined statements are quoted from **9. (Data Protection Act 2018, no Date)**

* Part 7, Section 184 - makes it illegal for businesses to require an individual to provide information contained in a health record when deciding if they will provide services to that individual
* Part 3, Chapter 3, Section 45 – Right of access by the data subject. This gives rights to customers to request confirmation as to whether or not personal data concerning him or her is being processed and access to the personal data
* Part 6, Section 170 – This makes it an offence for the business to obtain, disclose or retain personal data without the consent of the controller
* Part 6, Section 173 - alter, deface, block, erase, destroy or conceal information with the intention of preventing disclosure. This would be after a request to view the data from (for examples) the data subject.

The AI chatbot must must adhere to these requirements, or the business would be subject to enormous fines, and restrictions to data handling.

MOBILE APP

For mobile apps, the concerns would lean more towards data security and transparency. Due to the portability and accessibility of mobile phones, security for the app must be paramount.

This heavily involves the act of data sharing, since this is a very common practice for mobile apps. Including documenting all action taken on the data. (All lined statements are from 9.)

Chapter 4, Section 62. A controller must keep logs for at least the following processing operations … collection, alteration, consultation, disclosure (including transfers), combination, erasure

Additionally, Section 66 requires the organisation to evaluate risks. And implement measures to

* Prevent unauthorised processing
* Ensure that stored personal data cannot be corrupted
* Ensure the possibility to establish precise details of any processing that takes place

SOLUTIONS

There are many solutions that can be done during development and can be implemented as a routinely actions through the life of the projects. (Projects will refer to the AI chatbot and the Mobile bank app)

* Program the projects to only withhold relevant data. This excludes medical or criminal data, ensuring data relevance.
* Ensure projects contain functionality to provide data subject with all stored information concerning them, adhering to their right of access under section 45 of the DPA 2018.
* Security measures such as data encryption, and secure storage systems. To prevent data breaches and unauthorised access to sensitive information.
* Present a mandatory consent form before accessing the projects. This will outline all potential data handling procedures and the data subject’s rights to their data.
* Exceedingly Detailed audit logs of all processes that occurs in the automated systems. This includes but not limited to the deletion, addition, manipulation, transfer and use of any personal data that is stored by the business.
* |Secure and reliable backups of any data, in case of corruption or breach of security.

This will allow the projects to fall in the lines of the DPA 2018, keeping the business accountable but reliable to the safety of their customers data.

REFERENCES

1. Cambridge University Press & Assessment 2025 (No date) Bias. Available at: [BIAS | English meaning - Cambridge Dictionary](https://dictionary.cambridge.org/dictionary/english/bias) (accessed 12th January 2025)
2. BBC News (25th March 2016) Tay: Microsoft issues apology over racist chatbot fiasco. Available at: [Tay: Microsoft issues apology over racist chatbot fiasco - BBC News](https://www.bbc.co.uk/news/technology-35902104) (accessed 13th January 2025)
3. IMD (November 2018), Amazon’s sexist hiring algorithm could still be better than a human. Available at: [Amazon’s sexist hiring algorithm could still be better than a human - IMD business school for management and leadership courses](https://www.imd.org/research-knowledge/digital/articles/amazons-sexist-hiring-algorithm-could-still-be-better-than-a-human/) (accesses 13th January 2025)
4. European Commission (8th April 2019), Ethics guidelines for trustworthy AI. Available at: [Ethics guidelines for trustworthy AI | Shaping Europe’s digital future](https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai) (accessed 13th January 2025)
5. University of Winchester (1st November 2023) New study shows AI chatbots reflect human biases and focus on threat, negativity and gossip. Available at: [Media Articles - University of Winchester](https://www.winchester.ac.uk/news-and-events/press-centre/media-articles/new-study-shows-ai-chatbots-reflect-human-biases-and-focus-on-threat-negativity-and-gossip.php) (accessed 15th January 2025)
6. C3.ai (Not date) Explainable AI. Available at: [What is Explainability? | C3 AI Glossary Definition](https://c3.ai/glossary/machine-learning/explainability/#:~:text=Explainability%20(also%20referred%20to%20as,being%20at%20an%20acceptable%20level.) (accessed 17th January 2025)
7. Liz Burton-Hughes (2014) Data Protection Act 2018 Summary. Available at: <https://www.highspeedtraining.co.uk/hub/data-protection-act-summary/> (accessed 19th January 2025)
8. ICO (Not date) Part 1 The basics of explaining AI. Available at: [Legal framework | ICO](https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/legal-framework/#:~:text=The%20General%20Data%20Protection%20Regulation,or%20deploy%20an%20AI%20system.) (accessed 19th January 2025)
9. Legislation.gov.uk (No date) Data Protection Act 2018. Available at: [Data Protection Act 2018](https://www.legislation.gov.uk/ukpga/2018/12/section/45) (accessed 20th January 2025)