AI/Data Science Professional

COSC2778/COSC2792/COSC2818

Assessment Task 1:
Case Study of Data-Driven Decision-Making System

Florence 2.0

Report

Thu 4:30- Group 1

Mrwan Fahad A Alhandi	s3969393@student.rmit.edu.au
Eric Cheung	s3868588@student.rmit.edu.au
Luke Dale	s3964888@student.rmit.edu.au
Richard Doherty	s3863706@student.rmit.edu.au

Introduction

The AI project discussed in this report is "Florence 2.0" (WHO 2022), it is a web-based application hosted by World Health Organization (WHO). The previous version of "Florence" was designed to help delivering correct information about Covid - 19 from the beginning of the pandemic in English only. In Florence 2.0, it can also deliver information on mental health, tobacco products, and healthy living, in 6 additional languages.

The application was developed by a San Francisco & New Zealand based company, Soul Machines. It combines Natural Language Processing (NLP) and the designed digital human faces to deliver a virtual person experience to customers, available for 24/7 (WHO 2022) with the support from Amazon Web Services and Google iCloud. When the application is launched, the website prompts users to grant access to camera and microphone, the application can interact with both voice and words.

Like most AI applications, Florence 2.0 is also a black box operation, it is difficult for patients and the public to fully understand the process and limitation. To foster the development of AI, Data Science practitioners need to unpack and understand the issues of responsible data science practice.

In this project, we studied Fairness, Accountability, Transparency, Ethics (FATE) and other responsible data science issues around Florence 2.0 as an example of AI usage in Healthcare sector.

For Florence 2.0, relevant stakeholders can include a diverse range of individuals, such as:

Health professionals**:** World Health Organization (WHO), the professionals at WHO with the help with Qatar Ministry of Health, launched Florence 2.0 to fill the gap in health care sector to deliver messages to protect and promote people's physical and mental health (WHO 2022).

Technology companies and researchers: Companies which are developing the software, algorithms, hardware for AI usage, for Florence 2.0, it is a collaborative project between WHO and Soul Machines.

Patients and users: They are the group including the users receiving the health care information and who participate in the development of the service in providing feedback to developers.

Regulators: Governments and regulatory bodies such as Qatar Ministry of Health are the stakeholders. They are the bodies which set policies for areas such as privacy, security, transparency, bias, accountability and regulate the use of AI.

Background

The first version of Florence was designed for delivering Covid-19 information since the beginning of the pandemic when people were working from home and lack of physical contact with families and friends. The aim was to deliver correct information about Covid-19 to fight misinformation when people spent a lot of time seeking information online.

It has been reported that people suffered mental illness because of Covid-19. (WHO 2023), and the Florence 2.0 is now able to deliver other information such as mental health, tobacco products, and healthy living to address the follow-on issues from Covid-19.

Chatbot has been used in healthcare sector for years, and became more popular during Covid-19 pandemic, because it can engage patients 24/7 and on demand. Its aim is not to replace the role of human medical practitioners, but to help patients to figure out what was their ailment and get the patients to the right services, so that more resources can focus on treating patients. Also, it can deliver cost savings in administration and money can be spent on improving the health care industry in general. (Lerman 2021)

Data Privacy and Digital Trust

Artificial intelligence (AI) presents remarkable new capabilities; however, these advancements also introduce the possibility of violating privacy regulations in unprecedented ways. Each interaction with Florence 2.0, generates data derived from various sources, such as cameras, audio, or text through user engagement. Recently, Facebook has developed a ground-breaking technology called "DeepFace" that can determine whether two photographs depict the same person with accuracy comparable to that of the human brain (Dormehl, 2014). This data that comes from various sources can be transformed into several types of information, including personal, behavioural, location-based, sensory, financial, or health-related details.

Given its function as a digital health worker, Florence 2.0 primarily collects health-related data. Such data can be extremely sensitive and personal, encompassing aspects such as patients' medical histories, mental health, and other private elements of their lives. Safeguarding this information is crucial to protect consumers from potential harm, discrimination, or embarrassment. Consequently, it is essential to prioritize privacy and adhere to stringent data protection standards when utilizing AI technologies like Florence 2.0.

The two main issues to chatbots' like Florence 2.0 security are threats and vulnerabilities. A security threat refers to a hazard that may potentially compromise an organization and its systems. System vulnerabilities are flaws that can be used by an attacker to breach privilege limits (e.g., execute unauthorized tasks) within a computer system. Such vulnerabilities may arise due to inadequate coding, outdated hardware drivers, insufficient security system protection, and other similar factors. Most system vulnerabilities are the result of human mistakes (Hasal et al., 2021). Some of the data security and privacy technologies that can help deal with threats and vulnerabilities include Cloud Data Protection (CDP), tokenization, Big Data encryption, data discovery, flow mapping, and many more (Press, 2017). WHO in their privacy policy states that they employ a range of technologies and security measures to protect information maintained in their systems from loss, misuse, unauthorized access or disclosure, alteration, or destruction.

Establishing trust with consumers goes beyond mere security measures. Digital trust encompasses an organization's ability to safeguard consumer data, implement robust cybersecurity practices, deliver trustworthy AI-powered products and services, and provide transparency regarding the use of AI and data (Hamilton, 2022). To achieve transparency in AI and data usage, it is imperative for organizations to clearly articulate their ethical principles and guidelines.

Ethics

New healthcare technologies, often untested and unproven across sufficiently large population groups, may create risks which go unrealised for years (Ghalambor 2022). Whilst medicine has established moral and legal frameworks which synergistically promote the health, equity, and safety of patients, AI is an emerging field driven largely by private enterprise where commercial interests are prioritised (Pasricha 2022). As chatbot tools like Florence continue to be introduced across new settings and applications, ethical challenges have emerged.

Exploring the concept of fairness, Giovanola and Tiribelli (2021) note that biases are not limited to the quality and representativeness training data — they can also be structural, resulting from the design of a system itself. In this instance bias emerges when the scope of a machine learning model has been limited to easily measured cohorts or categories (Giovanola and Tiribelli 2021). This lack of granularity can be observed in the vaccination advice feature of Florence. One of the primary reasons the system was deployed was to help users comprehend the safety and clinical development of approved COVID-19 vaccines, but the information supporting this resource skews to Western audiences (WHO n.d.). Florence makes reference to six vaccines, of which the majority are produced by American and European manufacturers (WHO n.d.). However, as many as 12 vaccines have been fully approved worldwide with a further 21 authorised for early or emergency use, many of these being manufactured in China, India and Russia (Zimmer et al. 2022). Florence therefore runs the risk of generating informed mistrust, with prospective audiences from non-Western geographies developing feelings of apprehension or scepticism towards Florence because of perceived discrimination or partiality (Giovanola and Tiribelli 2021).

Shaban-Nejad et al. (2020) meanwhile contend that a lack of explainability in machine learning models means that prescribed health interventions are difficult to justify given the lack of understanding around how recommendations are selected and prioritised. Due to technical complexity or commercial confidentiality, AI-based systems are often black boxes that inherently limit the ability for stakeholders to scrutinise the soundness of decisions (Sheban-Nejad et al. 2020). Despite Florence being launched by the World Health Organisation, a globally recognised public health agency, the technology powering this platform is developed and maintained by a private company (WHO n.d.). It’s therefore difficult to understand whether the system was trained on clinically supported material, or if the model semi-independently generates responses based on a wider and less reliable information base. When asking Florence to describe itself, the system responds by stating: "Since my creation, I've searched the internet to find the best information … and I want to use this to now help you". (WHO n.d.). Whilst this may simply be a narrative technique employed by the provider to personalise Florence, it creates ambiguity and leaves the user to decipher the underlying logic themselves (Harrer 2023). For example, only through discovery of a press release do stakeholders understand that Florence has been designed to employ specific "brief advice" techniques when building a tobacco cessation plan. This strategy largely impresses upon the user that they should quit "cold turkey" (Soul Machines 2020: WHO n.d.). However, QUIT Victoria advocate that an approach combining immediate cessation with Nicotine Replacement Therapy or other prescription medications leads to more effective outcomes (QUIT Victoria 2022). With conflicting information like this, the trustworthiness of the platform is diminished.

Insofar as best practice, a number of mechanisms and processes have been developed to support the ethical and effective operation of AI-based chatbots. Horn et al. (2020) have determined that micro-service system architectures function best for healthcare applications. The finer control of query interpretation, made possible by deep and accessible knowledge bases, allows modular machine learning models to be built with an understanding of discipline-specific requests and requirements (Horn et al. 2020). This prevents systems from being trained upon generalised data sources, improving fairness and reducing risks for error. Harrer (2023) also suggests that machine learning models could be assigned an index to measure reliability and explainability. Whilst a relatively new concept, with only one comprehensive framework having been developed to date, such metrics could offer a more simplistic solution to evaluate trustworthiness compared with existing proprietary solutions that focus more on bias (Harrer 2023).

Future iterations of Florence may benefit by strengthening existing capabilities rather than rushing to expand and integrate new services. Offering new technology as an opportunity to test proprietary systems, or to simply be at the "cutting edge" of product development, is ethically questionable (Ghalambor 2022). The provider states in their privacy terms that the collection of personal information is helping to advance human-to-machine interaction (WHO n.d.). Refocusing resources on building and testing the base of medical information underpinning Florence, rather than refining the animations which give life to this "digital person", will ensure the machine learning model acts with beneficence.

Florence should also be reinforced to handle more severe presentations. For example, the mental health function insufficiently directs users experiencing acute stress or depression to professional help. Just like how country-specific tobacco cessation hotlines can be provided, Florence should similarly maintain a directory of emergency services and crisis contacts (e.g., Lifeline). Additionally, users should be more thoroughly educated on the limitations of the system through product disclaimers and adjusted dialogue. These changes would enhance the degree to which risks and responsibilities are understood, leaving the user more informed as to whether alternative healthcare services should be sought and clarifying the accountabilities on part of the provider (Harrer 2023).

Other desirable additions include clearer references to source material to enhance the trustworthiness and explainability of advice, making guides (e.g., for the vaccination and nutrition features) region-specific to ensure socio-cultural differences are adequately catered to, and engaging test groups to ensure unintended structural bias is appropriately reported to development teams and subsequently addressed.

Conclusion

In this case study report, we examine the healthcare AI application, Florence 2.0. The aim of Florence 2.0 is to deliver accurate information about Covid 19 and follow on issues in several languages. We have identified the stakeholders including health professionals, technology companies, patients, and regulators. Florence 2.0, a digital health worker, collects health – related information to provide advice to users, that raises the issue of privacy and the importance of data protection standard.

Establishing digital trust with consumers requires organizations to safeguard consumer data, implement robust cybersecurity practices, and deliver trustworthy AI-powered products and services while providing transparency regarding the use of AI and data. Organizations must clearly articulate their ethical principles and guidelines to achieve transparency in AI and data usage.

Exploring the concept of fairness, we learnt that Florence 2.0 only refers of 6 Covid-19 vaccines (manufactured in US and Europe) when there are actually many other vaccines have been fully approved worldwide or authorised for early / emergency use (manufactured in China, Russia and India). This arises the risk of generating informed mistrust and discrimination or partiality. While Florence, like other AI application, the technical complexity and commercial confidentially, makes it difficult to understand whether the system was trained on clinically supported material, or if the model semi-independently generates responses based on a wider and less reliable information base.

Insofar as best practice, Horn et al. (2020) have determined that micro-service system architectures function best for healthcare applications. Harrer (2023) also suggests that machine learning models could be assigned an index to measure reliability and explainability. Future iterations of Florence may benefit by strengthening existing capabilities rather than rushing to expand and integrate new services. Florence should also be reinforced to handle more severe presentations, Additionally, users should be more thoroughly educated on the limitations of the system through product disclaimers and adjusted dialogue. Other desirable additions include clearer references to source material to enhance the trustworthiness and explainability of advice, making guides region-specific to ensure socio-cultural differences are adequately catered to, and engaging test groups to ensure unintended structural bias is appropriately reported to development teams and subsequently addressed.

Bibliography / References

Dormehl, L. (2014) Facial recognition: Is the technology taking away your identity? The Guardian. Guardian News and Media. Available at: https://www.theguardian.com/technology/2014/may/04/facial-recognition-technology-identity-tesco-ethical-issues (Accessed: April 25, 2023).

Hasal, M. et al. (2021) "Chatbots: Security, privacy, data protection, and social aspects," Concurrency and Computation: Practice and Experience, 33(19). Available at: https://doi.org/10.1002/cpe.6426.

Press, G. (2017) Top 10 Hot Data Security and Privacy Technologies, Forbes. Forbes Magazine. Available at: https://www.forbes.com/sites/gilpress/2017/10/17/top-10-hot-data-security-and-privacy-technologies/?sh=4beef7d66b3f (Accessed: April 25, 2023).

Privacy policy (no date) World Health Organization. World Health Organization. Available at: https://www.who.int/about/policies/privacy (Accessed: April 25, 2023).
Hamilton, H. (2022) Digital Trust. why it's important for your business, Digital trust. Why it's important for your business. Jamf. Available at: https://www.jamf.com/blog/digital-trust-5-reasons-it-matters-for-your-business/#:~:text=%E2%80%9CDigital%20trust%20is%20the%20confidence,around%20AI%20and%20data%20usage.%E2%80%9D (Accessed: April 25, 2023).

WHO (World Health Organization) (2022) WHO and partners launch world's most extensive freely accessible AI health worker, WHO website, accessed 1 April 2023. (https://www.who.int/news/item/04-10-2022-who-and-partners-launch-world-s-most-extensive-freely-accessible-ai-health-worker)

World Innovation Summit for Health (WISH) (2022), WISH website, accessed 1 April 2023. (https://2022.wish.org.qa/)

WHO (World Health Organization) (2023) Mental health & COVID-19, WHO website, accessed 1 April 2023. (https://www.who.int/teams/mental-health-and-substance-use/mental-health-and-covid-19)

Lerman R (World Health Organization) (2021) The robot will see you now: Health-care chatbots boom but still can't replace doctors, The Washington Post, accessed 1 April 2023. (https://www.washingtonpost.com/technology/2021/07/26/healthcare-chatbots-pandemic-rise/)

Comes S, Chauhan R and Schatsky D (2021) Conversational AI Five vectors of progress, Deloitte Insight, accessed 1 April 2023. (https://www2.deloitte.com/us/en/insights/focus/signals-for-strategists/the-future-of-conversational-ai.html/)

Kalinin K (2022) Medical Chatbots: The Future of the Healthcare Industry, Topflightapp, accessed 1 April 2023. (https://topflightapps.com/ideas/chatbots-in-healthcare/)

Ghalambor M (2022) ‘Ethical challenges in applying new technologies in orthopaedic surgery’, in Ehansi S, Glauner P, Plugmann P and Thieringer F (Eds) *The future circle of healthcare: AI, 3D printing, longevity, ethics and uncertainty mitigation*, Springer Cham, https://doi.org/10.1007/978-3-030-99838-7.

Giovanola B and Tiribelli S (2021) ‘Beyond bias and discrimination: Redefining the AI ethics principle of fairness in healthcare machine-learning algorithms’, *AI & Society.* https://doi.org/10.1007/s00146-022-01455-6.

Harrer S (2023) ‘Attention is not all you need: The complicated case of ethical using large language models in healthcare and medicine’, *eBioMedicine*, 90. https://doi.org/10.1016/j.ebiom.2023.104512.

Horn M, Xiang L, Chen L and Kafle S (2020) ‘A multi-talent healthcare AI bot platform’, in Sheban-Nejad A, Michalowski M and Buckeridge D (Eds) *Explainable AI in Healthcare and Medicine*, Springer Cham. https://doi.org/10.1007/978-3-030-53352-6_1.

Pasricha S (2022) ‘AI ethics in smart healthcare’, *arXiv.* https://doi.org/10.48550/arXiv.2211.06346.

QUIT Victoria (2022) *Cold turkey*, QUIT Victoria website, accessed 7 April 2023. https://www.soulmachines.com/2020/09/florence-digital-health-worker/.

Sheban-Nejad A, Michalowski M and Buckeridge D (2020) ‘Explainability and interpretability: Keys to deep medicine’, in Sheban-Nejad A, Michalowski M and Buckeridge D (Eds) *Explainable AI in Healthcare and Medicine*, Springer Cham. https://doi.org/10.1007/978-3-030-53352-6_1.

Soul Machines (2020), *Florence: Digital health worker*, Soul Machines website, accessed 7 April 2023. https://www.soulmachines.com/2020/09/florence-digital-health-worker/.

WHO (World Health Organisation) (n.d.) *Florence*, WHO website, accessed 7 April 2023. https://www.who.int/campaigns/Florence.

Zimmer C, Corum J, Wee S and Kristofferson M (2022) *Coronavirus vaccine tracker*, New York Times website, accessed 7 April 2023. https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html.