AI/Data Science Professional COSC2778/COSC2792/COSC2818

Thu 4:30- Group 1

Project Proposal

Improving Transparency and Interoperability of Heathcare AI systems

Mrwan Fahad A Alhandi	s3969393
Eric Cheung	s3868588
Luke Dale	s3964888
Richard Doherty	s3863706

RMIT UNIVERSITY

Introduction

Eric Cheung

- Better clinical decision and user experiences
- Stakeholders
- Pain points
- Digital Trust and Responsible AI are important

RMIT UNIVERSITY

Problem Definition

Luke Dale

Interoperability

RMIT UNIVERSITY

Image: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health

Problem Definition

Luke Dale

Interoperability

RMIT UNIVERSITY

Image: https://www.pexels.com/photo/photo-of-doctor-examining-patient-7088829/

Problem Definition

Luke Dale

Interoperability

Solutions are required which:

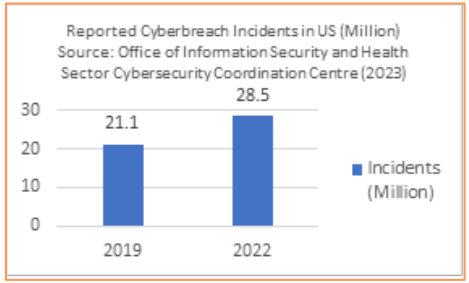
- Support greater interconnectivity between data infrastructures
- Standardise semantics and syntax
- ”Democratise” access to healthcare data

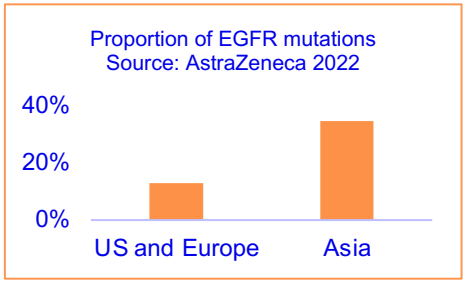
RMIT UNIVERSITY

Problem Significance

Eric Cheung

- Privacy and cybersecurity
- Transparency
- Train data availability





RMIT UNIVERSITY

Proposed Data-Driven Solution

Eric Cheung

- Consolidation and Coordination
- WHO – new regulatory and guideline standard
- Ethical concerns – resources to implement
- Job security concerns
- Education

RMIT UNIVERSITY

Methodology

Mrwan Fahad A Alhandi

8 Core Principles for AI

1. Generates net-benefits
2. Do not harm
3. Regulatory & legal compliance
4. Privacy protection
5. Fairness
6. Transparency & Explainability
7. Contestability
8. Accountability

RMIT UNIVERSITY

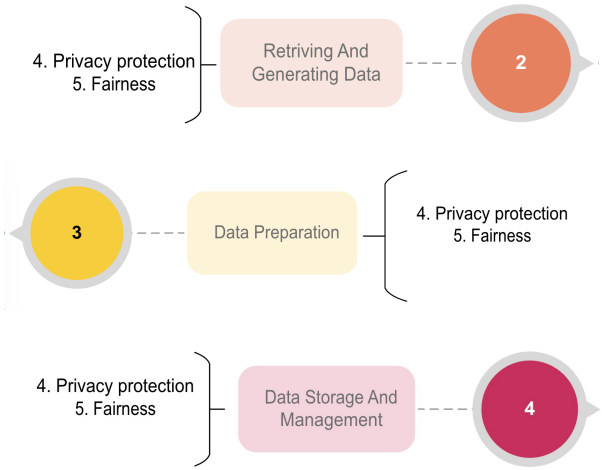
Methodology

Mrwan Fahad A Alhandi

RMIT UNIVERSITY

Methodology

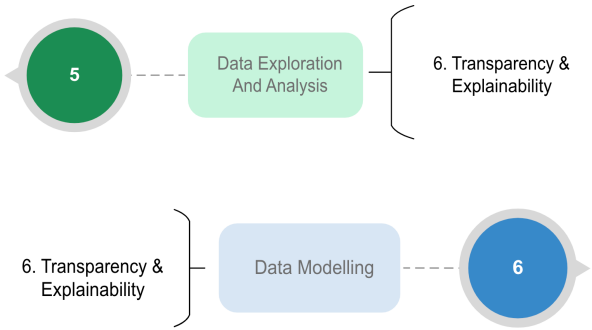
Mrwan Fahad A Alhandi



RMIT UNIVERSITY

Methodology

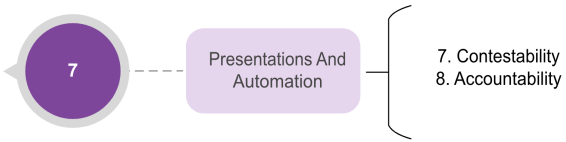
Mrwan Fahad A Alhandi



RMIT UNIVERSITY

Methodology

Mrwan Fahad A Alhandi



RMIT UNIVERSITY

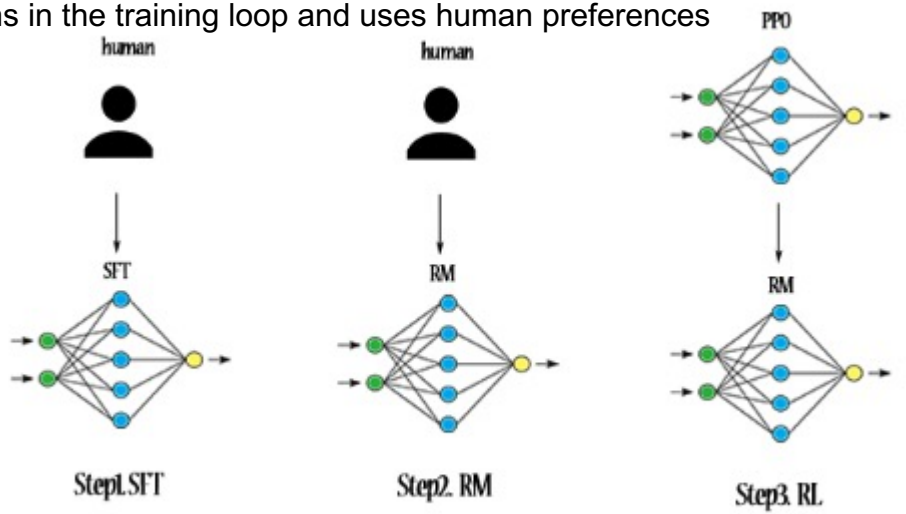
Design Prototype - LLMs

Richard Doherty

- LLMs – Large Language Models. The ability of OpenAI's LLM-enabled ChatGPT to engage in human-like conversations is now widely known and accepted and will influence future user expectations for all products.

- *"LLMs have shown remarkable capabilities in a wide range of NLP tasks. However, these models may sometimes exhibit unintended behaviours, e.g., fabricating false information, pursuing inaccurate objectives, and producing harmful, misleading, and biased expressions. (La Vivien, 2023)."*

- To address these key trust-related concerns (especially in the context of Healthcare AI systems), we propose that the technique of Reinforcement Learning with Human Feedback (**RLHF**) be embedded in the training methods of Healthcare AI systems. RLHF incorporates humans in the training loop and uses human preferences as a "reward signal" to fine-tune LLMs. (La Vivien, 2023).



RMIT UNIVERSITY

Design Prototype – RLHF steps

Richard Doherty

(La Vivien, 2023)

1. Supervised fine-tuning (**SFT**)

“Collect demo data & train a supervised policy. Labellers provide desired behaviour demos on input prompt distribution. Team fine-tunes a pretrained GPT-3 model on this data using supervised learning."

2. Rewording model training (**RM**)

“Collect comparison data and train a reward model. The team collects a dataset of comparisons between model outputs, where the labellers indicate which output, they prefer for a given input. Then trains a reward model to predict the human-preferred output.”

3. Reinforcement Learning fine-tuning (**RL**)

*“Optimize a policy against the reward model using **PPO**. The team uses the output of the RM as a scalar reward. Then fine-tunes the supervised policy to optimize this reward using the Proximal Policy Optimization PPO algorithm."*

"Steps 2 and 3 are iterated continuously; more comparison data is collected on the current best policy, which is used to train a new RM and then a new policy. In practice, most of comparison data comes from supervised policies, with some coming from PPO policies.” (La Vivien, 2023)

RMIT UNIVERSITY

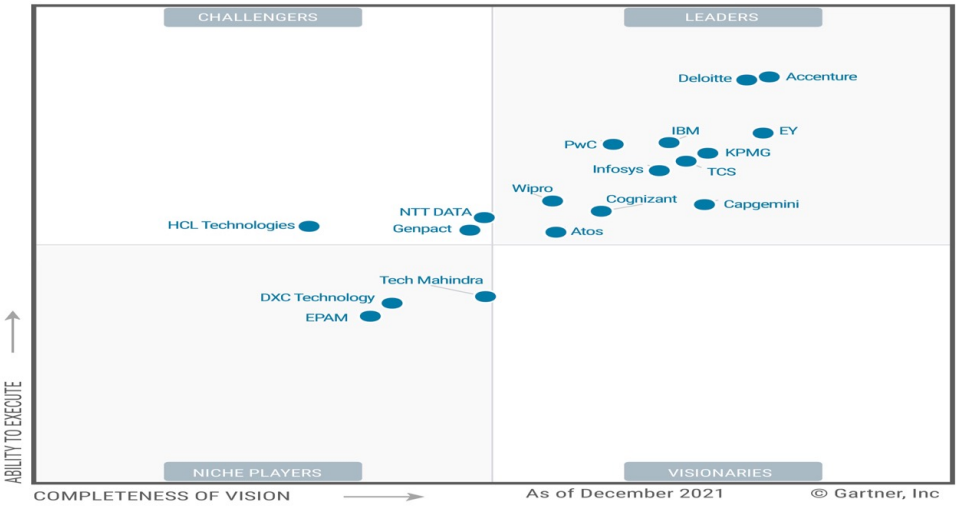
Richard Doherty

Conclusion

- Incorporation of our proposed use of LLMs and most importantly, RLHF, to be further defined and stated within the guidelines specified by governing bodies such as WHO referenced in our problem definition above.
- Progression through a full Data Science Lifecycle with emphasis on the ***Eight Core Principles of AI*** discussed in the Methodology section above should be included as part of any Heathcare AI system development.

- The global consulting firms highlighted in the Gartner Magic Quadrant for Data and Analytics service providers should all be well positioned to help make sure the guidelines from the WHO become best practice and measure up to generally accepted international standards for future AI audit and compliance.

Figure 1: Magic Quadrant for Data and Analytics Service Providers



Source: Gartner (February 2022)

RMIT UNIVERSITY

Bibliography

Ali, N. (2022) EHR interoperability challenges and solutions. Available at: https://www.ehrinpractice.com/ehr-interoperability-challenges-solutions.html#:~:text=The%20most%20common%20approach%20for,and%20health%20information%20technology%20systems. (Accessed: 30 May 2023).

Bresciani, S., and M. J. Eppler. 2008. “The risks of visualization: A classification of disadvantages associated with graphic representations of information.” Institute for Corporate Communication. https://pdfs.semanticscholar.org/23d2/3f5152c9b8b34f104b43d1c862ee62d2edac.pdf.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.
CSIRO (2018) Australia Ai Ethics Framework, CSIRO. Available at: https://www.csiro.au/en/research/technology-space/ai/ai-ethics-framework (Accessed: 31 May 2023).

OpenAI (2022) ChatGPT, Introducing ChatGPT. Available at: https://openai.com/blog/chatgpt (Accessed: 01 June 2023).

Rigby ,Michael J. (2019) Ethical Dimensions of Using Artificial Intelligence in Health Care ,Journal of Ethics | American Medical Association (ama-assn.org) (Assessed 1 May 2023)
https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02

National Health and Medical Research Council, (2018), National Statement on Ethical Conduct in Human Research (2007) - Updated 2018 . accessed 1 May 2023.
https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018

Role of Artificial Intelligence (AI) in Health Insurance (acko.com)
Team Acko 2023, Understanding the Role of AI in Health Insurance, Team Acko, accessed 1 May 2023. https://www.acko.com/health-insurance/ai-artificial-intelligence-in-health-insurance/

Chen I, Szolovits P, and Ghassemi M (2019) Can AI Help Reduce Disparities in General Medical and Mental Health Care? Journal of Ethics | American Medical Association (ama-assn.org) (Assessed 1 May 2023)
https://journalofethics.ama-assn.org/article/can-ai-help-reduce-disparities-general-medical-and-mental-health-care/2019-02

Morgan M (2023) Why Artificial Intelligence Is Becoming A Cybersecurity Imperative And How To Implement It (forbes.com) , accessed 1 May 2023
https://www.forbes.com/sites/forbestechcouncil/2023/03/15/why-artificial-intelligence-is-becoming-a-cybersecurity-imperative-and-how-to-implement-it/?sh=5172d1d9610d

La Vivien Post (2023), ‘How ChatGPT works – Architecture illustrated (viewed 1 June 2023). https://www.lavivienpost.com/how-chatgpt-works-architecture-illustrated/

RMIT UNIVERSITY

Bibliography

Dawson D and Schleiger E* , Horton J, McLaughlin J, Robinson C∞, Quezada G, Scowcroft J, and Hajkowicz S† (2019) Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia. Assessed 1 May 2023
https://www.csiro.au/-/media/D61/Reports/Artificial-Intelligence-ethics-framework.pdf

Chen I Y., Szolovits P,, and Ghassemi M, 2019 Can AI Help Reduce Disparities in General Medical and Mental Health Care? | Journal of Ethics | American Medical Association (ama-assn.org) , assessed 1 May 2023. https://journalofethics.ama-assn.org/article/can-ai-help-reduce-disparities-general-medical-and-mental-health-care/2019-02

Office of Information Security & Health Sector Cybersecurity Coordination Center, (2023), Data Exfiltration Trends in Healthcare. accessed 1 May 2023.
https://www.hhs.gov/sites/default/files/data-exfiltration-in-healthcare-tlpclear.pdf

Ross C and Swetlitz I (2018), IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show, STAT, accessed 1 May 2023.
https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/
AstraZeneca (2020) Lung Cancer in Asia , accessed 1 May 2023 (https://www.astrazeneca.com/content/dam/az/our-focus-areas/Oncology/2020/lungcancer/Lung%20Cancer%20in%20Asia%20Backgrounder_APPROVED_2020.pdf)

Appen (2020) *What does interoperability mean for the future of machine learning?*, Appen website, accessed 15 May 2023. https://appen.com/blog/what-does-interoperability-mean-for-the-future-of-machine-learning/?amp.

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.

healthdirect (2021) *CT scan*, healthdirect website, accessed 15 May 2023. https://www.healthdirect.gov.au/ct-scan.

Lehne M, Sass J, Essenwanger A, Schepers J and Thun S (2019) 'Why digital medicine depends on interoperability', *npc Digital Medicine*, 2(79). https://doi.org/10.1038/s41746-019-0158-1.

Panch T, Mattie H and Celi L (2019) 'The "inconvenient truth" about AI in healthcare', *npj Digital Medicine*, 2(77). https://doi.org/10.1038/s41746-019-0155-4.

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.

RMIT UNIVERSITY