Singapore Healthcare Management 2018

Managing Multidimensional Risks Associated with the use of Artificial Intelligence Technologies for the Delivery of Health Services

Lam Shao Wei Sean[1], Zhang Yaoxian, Alwin[2], James Pang Yan[3], Liu Nan[1], Julian Thumboo[1], Ong Eng Hock Marcus[1]

1 SingHealth Health Services Research Centre and Duke-NUS Medical School; 2 National University of Singapore Business Analytics Centre; 3 National University of Singapore Business School and School of Computing

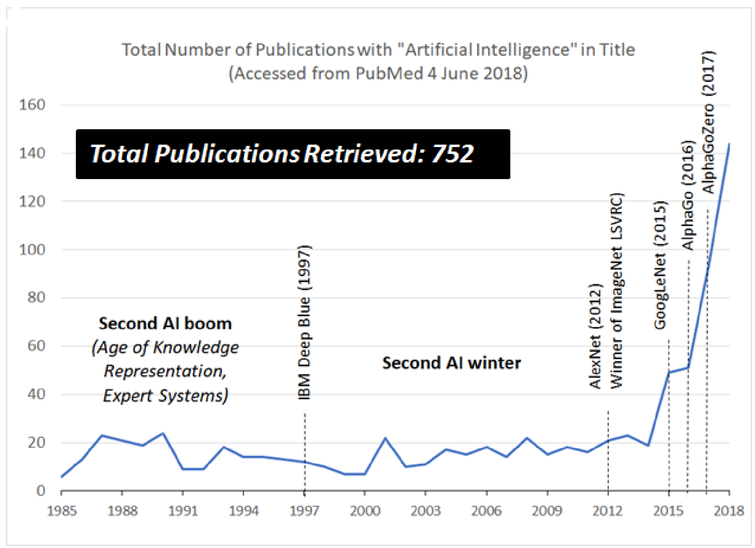
Background: Promise of Artificial Intelligence in Medicine

The ongoing narrative in the use of Artificial Intelligence (AI) in medicine and healthcare suggests that this technology will significantly disrupt the health services delivery process across the entire care delivery chain. There exists a huge potential for the innovative use of AI in medicine and healthcare. This poster outlines an exploratory study into the area of AI governance from the perspective of **Fairness, Accountability, Transparency and Ethics (FATE)**. The focus of the expository study will be in the development of rigorous data scientific approaches for improving AI **Accountability and Transparency** through explainable AI models.

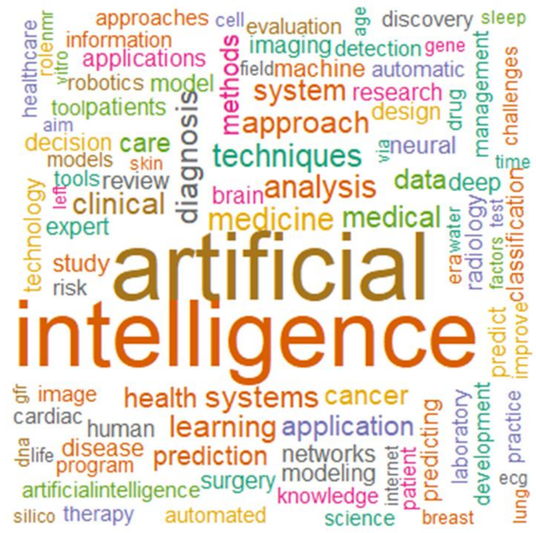
Literature review was conducted in PubMed for "Artificial Intelligence" in the title. A total of 752 medical related publications were identified. The literature review revealed a sharp increase in AI related medical research since 2012. Titles were stemmed and a wordcloud was developed from these records to explore the key topics of interest.

Input → AI Black Box → **Output**

The key questions that we attempt to address is that even when the AI system performs with acceptable behaviour through standard model validation, should I trust your model? If there is a problem with the model, who is accountable for the decision?



Growth of Relevant Publications in PubMed



AI Topical Terms in PubMed Review

AI Value Chain and Key Governance Issues

Data
- Entity Resolution
- Relationship and Feature Extraction and Transformation
- Annotation and Tokenization
- Internal (e.g., EMR, Disease Registries) and external data sources

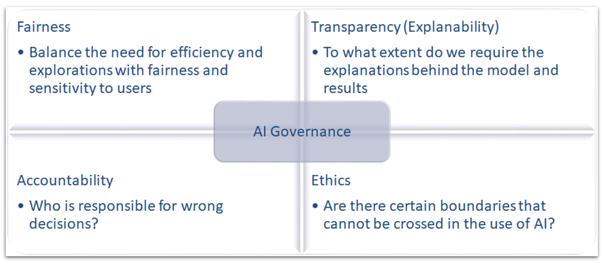
Descriptive
- Query, Filter, Drill-down
- Report Generation
- Exploratory Data Analysis
- Dashboard Development

Predictive
- Machine Learning, Association Studies, Causality
- Forecasting and risk scoring models
- Anomaly detection and alerts
- Deep Learning

Prescriptive
- Optimization under uncertainty and ambiguity
- Optimization, decision complexity and computational efficiency
- Reinforcement Learning

Cognitive
- Adaptive analysis – responding to context
- Continual learning – Learning Healthcare System, responding to dynamic changes and feedback

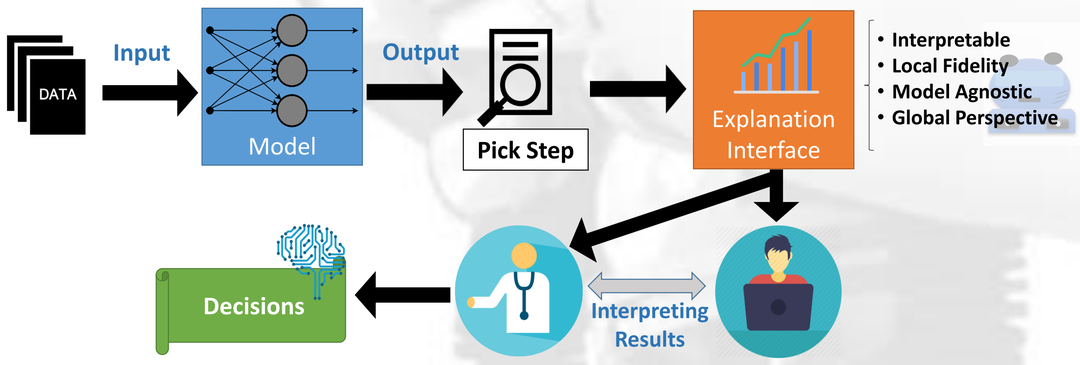
AI Governance Matrix



Mathematical Framework for Explanable AI

Desired Characteristics of Explanable AI models:
- **Interpretability** that is able to account for user's limitations in understanding the machine explanations
- **Local fidelity**. Usually it is impossible to explain a complex AI model faithfully, but a meaningful explanation should be available which is locally faithful
- **Model Agnostic**. Explanation should be able to explain any model
- **Global perspective**. Explanation should also be globally consistent.

Explanability & Accountability

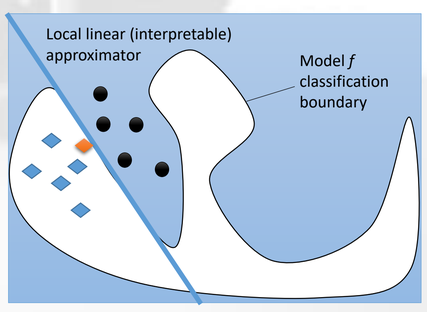


Optimal Trade-offs Between Interpretability and Local Fidelity

Optimal Explanation: $\xi(x) = \underset{g \in \mathcal{G}}{\operatorname{argmin}} \{\mathcal{L}(f, g, \pi_x) + \Omega(g)\}$

$g \in \mathcal{G}$: Class of possible locally interpretable models (e.g., linear models, decision trees)
$\Omega(g)$: Measure of complexity (**Interpretability**)
$\mathcal{L}(f, g, \pi_x)$: Measure of "infidelity" that local interpretable model g is in approximating model f in the locality define by π_x. Measure of **Fidelity**

Given the class of linear locally interpretable models, $\mathcal{L}$ can be defined on a suitable distance function (e.g., ℓ_2- norm) with locality π_x defined with a suitable kernel (e.g., exponential).



Schematic of locally interpretable approximator for the classification and explanation of instance

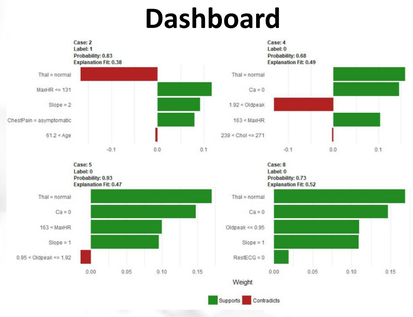
Case Study – AHD Diagnosis

Atherosclerosis Heart Disease (AHD) Explanable Prediction

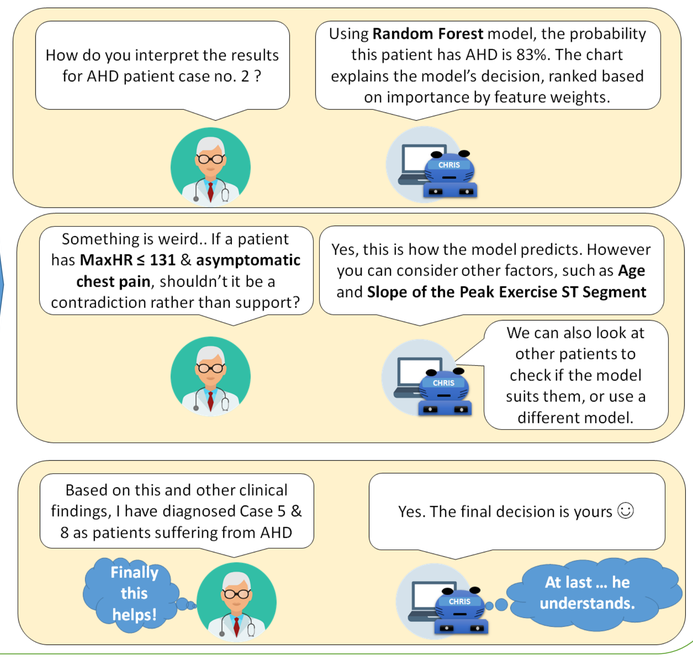
Doctors deciding if they should diagnose a patient with **Atherosclerosis Heart Disease** (AHD) using Explainable AI for augmenting clinical decisions. Data obtained from: http://www-bcf.usc.edu/~gareth/ISL/Heart.csv

Key factors distilled from "Pick Step" which maximizes the coverage across all instances subjected to a finite budget.

Sample Decision Support Dashboard



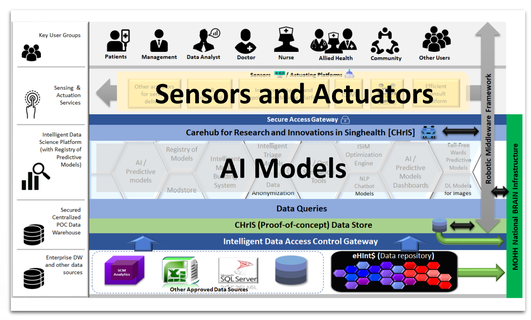
 # Decision support system can be deployed for guiding/ training of junior doctors/residents as decision processes can be captured and reviewed



Infrastructure for Data Science Governance

Proposed development of the Care-Hub for Innovation in SingHealth (CHrIS)

1. Address governance issues under the HBRA.
2. Reproducibility and objective tuning of complex data science models
2. Sustainable monitoring and maintenance of AI models
3. Secured, traceable and sustainable knowledge repository for AI Models
5. Secured, Traceable and Efficient Data Access and Model Development



** CHrIS is interfaced with the SingHealth Datawarehouse, eHIntS, and can be the data science enabling engine for eHIntS*

References:
[1] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 1135-1144. DOI: https://doi.org/10.1145/2939672.2939778