Data Analytics & AI in Health

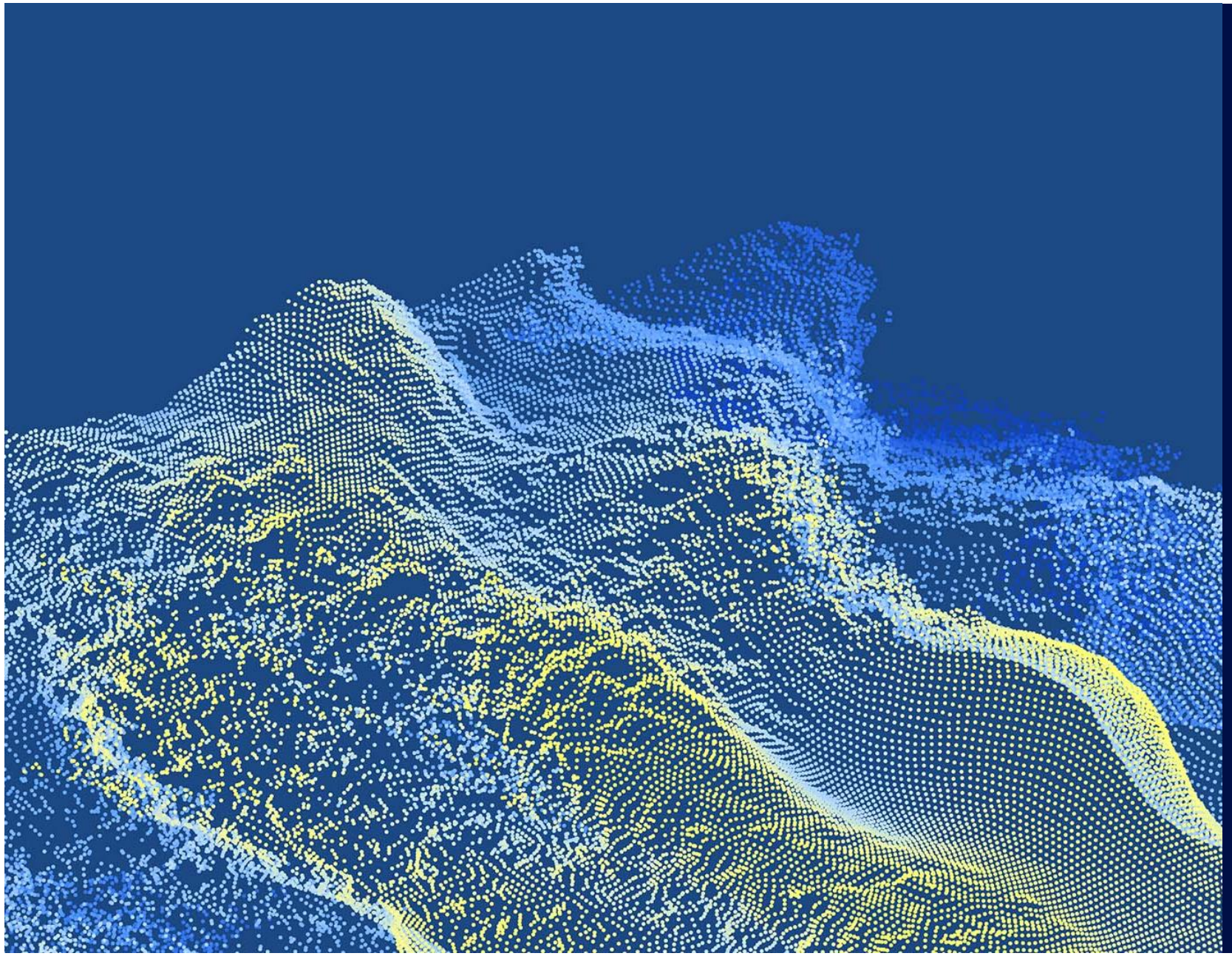






Dr Noel Faux
Ms Priyanka Nair-Turkich
Dr Mel Mistica

Melbourne Data Analytics Platform (MDAP)



MDAP overview

THE UNIVERSITY OF MELBOURNE



Dr Noel Faux

Senior Research Data Specialist

The MDAP team

COLLABORATION
COMPUTING
RESEARCH IMPACT

MDAP is a team of academic specialists enabling data-intensive research across disciplines at UoM.

- Established in 2019 as part of Petascale Campus, funded by Chancellery, Faculties and partial cost recovery.
- Supported more than 160 collaborative research projects across all Faculties, and many more researchers via consultation, events and training.



Research Data Specialist expertise



HIGH PERFORMANCE AND CLOUD COMPUTING



NATURAL LANGUAGE PROCESSING



MACHINE LEARNING



COMPUTING INFRASTRUCTURE



DATA ANALYSIS



DATA STEWARDSHIP

Plus

- Data management, curation and wrangling
- Software development and pipeline management
- Image processing
- Virtual/augmented reality
- Visualisation
- Web development
- HASS and STEM knowledge

and more…

What does working with MDAP look like?

MDAP researchers work collaboratively on projects, are acknowledged in publications or research outputs developed, and where appropriate, share authorship.

Working with us may include, but is not limited to:

- Collaborative research design
- Expertise in computing, data management and processing, computational analysis, visualisation, interpretation and more
- Tool familiarisation and upskilling
- Contribution to grant or funding applications
- Advice on avenues for publishing or sharing research outcomes

We can also redirect you to other platforms and groups who might be better suited to advance your research.

Engage with us



Faculty supported collaborations

- Formal
- LOI open September every year
- Projects lasting 3+ months
- High strategic value
- Merit based allocation
- Independent senior academic decision-making committee

Researcher funded engagements

- Engage individual MDAP staff via a salary cost-recovery model OR
- Engage a team of data specialists at Academic Level C1 rates
- Open year round (subject to team capacity)

Advice and short projects

- Informal
- Open year round (subject to team capacity)
- Includes consultancy-style advice
- Free or at cost-recovery rates

Get in touch at mdap-info@unimelb.edu.au

unimelb.edu.au/mdap/engage-with-us

Building a data-intensive community of practice

Alongside Collaborations, researcher-funded engagements and short projects, the MDAP team play a pivotal role nurturing the vibrant data-intensive research space at UoM. Explore and join our newsletter unimelb.edu.au/mdap/community

The team provide leadership, advocacy, and expertise at the University, as well as across Australia and internationally. Our growing community of practice supports researchers, academics and third-spacers to:

- share knowledge and resources through our regular seminars and panels
- develop skills and research capacity through digital skills training and our intern program
- share career progression, funding and other opportunities
- build awareness and recognition of non-traditional research outcomes
- develop connections and collaborate across disciplines.



Agent-based modelling and simulation in health research



Priyanka Nair-Turkich

Research Data Specialist

Agent-based models: Complex systems with simple rules

"Agent-Based Models (ABMs) are computational models in which autonomous "agents" interact with one another and with their environment, thereby producing aggregate emergent phenomena"

- Computer simulations used to study the interactions between people, things, places, and time. In ABMs, an agent is an encapsulated computer simulation that is situated in some environment.
- The agent could be humans, plants, biological entities, animals, machines, etc. Used in several fields to study the evolution of complex systems from micro-level assumptions

Agent-based models: Complex systems with simple rules

ABMs are *stochastic* models built from bottom up and are assigned rules and attributes that govern –

1) Interactions between agents in the target system 2) the interactions of the agents with the target system

**Stochasticity* - A stochastic process is a type of mathematical model that describes a process that evolves over time and is governed by randomness or uncertainty. The term “stochastic” is used to describe any process or variable that is subject to random fluctuations or uncertainty.

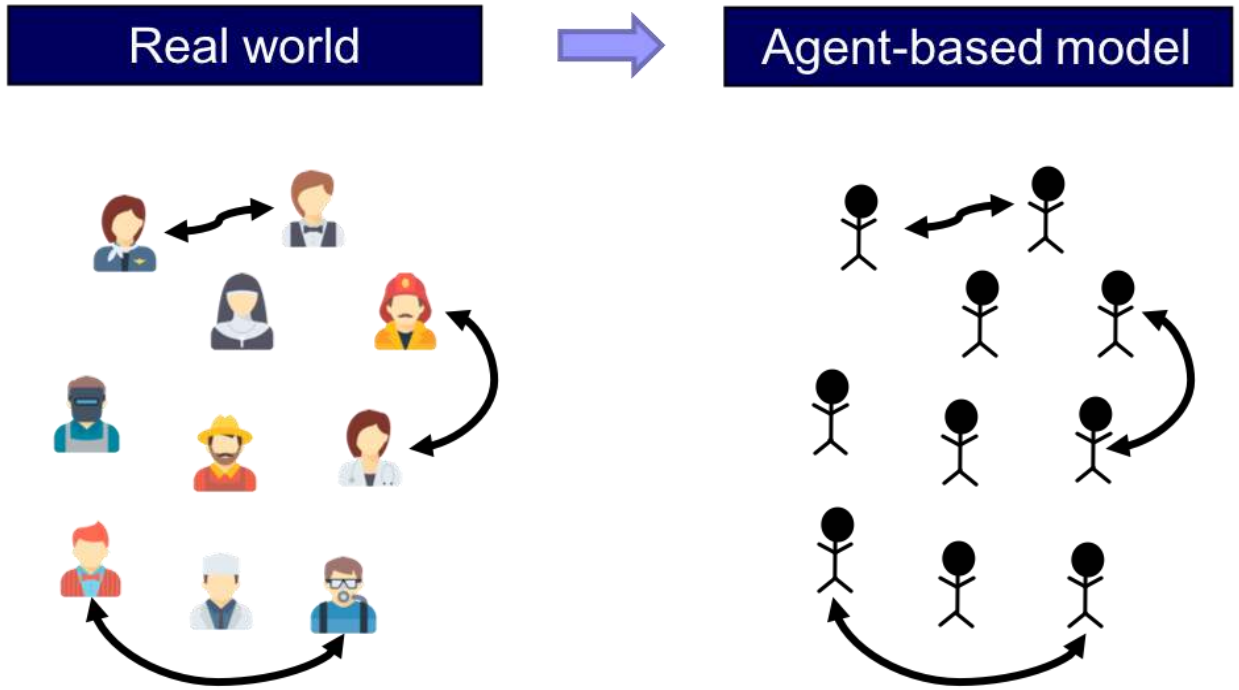


Figure source - LibreTexts - Introduction to agent-based modellling

Core concepts of ABMs

O	1. Purpose and patterns
	2. Entities, state variables and scales
	3. Process overview and scheduling *Submodel A* *Submodel B …*
D	4. Design concepts
D	5. Initialization
	6. Input data
	7. Submodels *Submodel A (Details)* *Submodel B (Details) …*

- Basic principles
- Emergence
- Adaptation
- Objectives
- Learning
- Prediction
- Sensing
- Interaction
- Stochasticity
- Collectives
- Observation

Figure source - The ODD Protocol for Describing Agent-Based and Other Simulation Models https://www.jasss.org/23/2/7.html

When to use agent-based models?

Use statistical models when you need to understand relationships between variables and make inferences about large data/populations based on sample data.

Use ML when you have large datasets and need to predict future outcomes. MLs are best suited for tasks like forecasting, where historical data can inform future decisions.

Use ABM when you need to understand how individual entities interact and the impact of their interaction on the system

Simulate what happens when individual parts interact, providing a deeper understanding of system behaviour under different conditions. ABM are valuable when trying to test “what if” and ”what if x did not happen”

Can be used to model the counterfactual or experiments that may be impossible or unethical to conduct in the real world!

ABM Project – National Disability Insurance Scheme (NDIS)

Collaborators – Dr Yi Yang, Dr Marissa Shields, Dr George Disney, Ms Huixin Chen

MDAP team – Ms Priyanka Nair-Turkich, Dr Simon Mutch

An agent-based model designed to analyse and understand the flow of individuals through Australia's National Disability Insurance Scheme (NDIS). The model traces individuals from initial mental health and employment needs through support services focussed on health and job guarantee.

Features –

- **Modelling multiple care pathways** across support tiers and programs

- **Demographic simulation** with age-structured populations and socioeconomic diversity

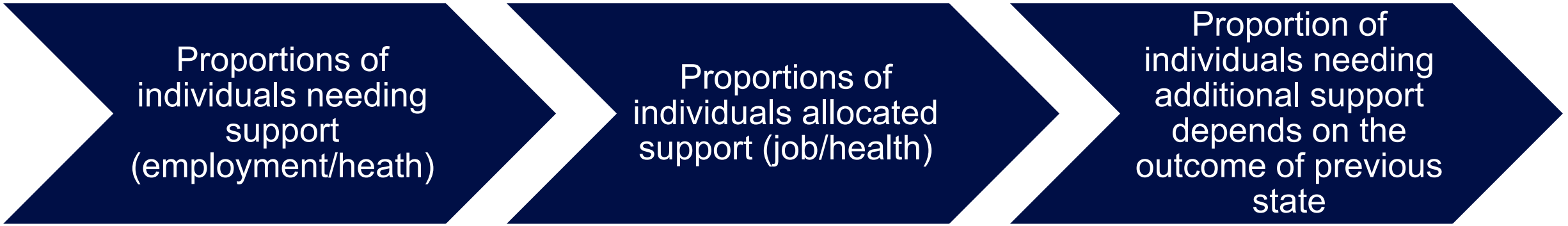
- **Dynamic transitions** between support states based on needs, accessibility, and outcomes

- S**cenario testing** to evaluate effectiveness of programs

ABM Project – National Disability Insurance Scheme (NDIS)

Collaborators – Dr Yi Yang, Dr Marissa Shields, Dr George Disney, Ms Huixin Chen

MDAP team – Ms Priyanka Nair-Turkich, Dr Simon Mutch

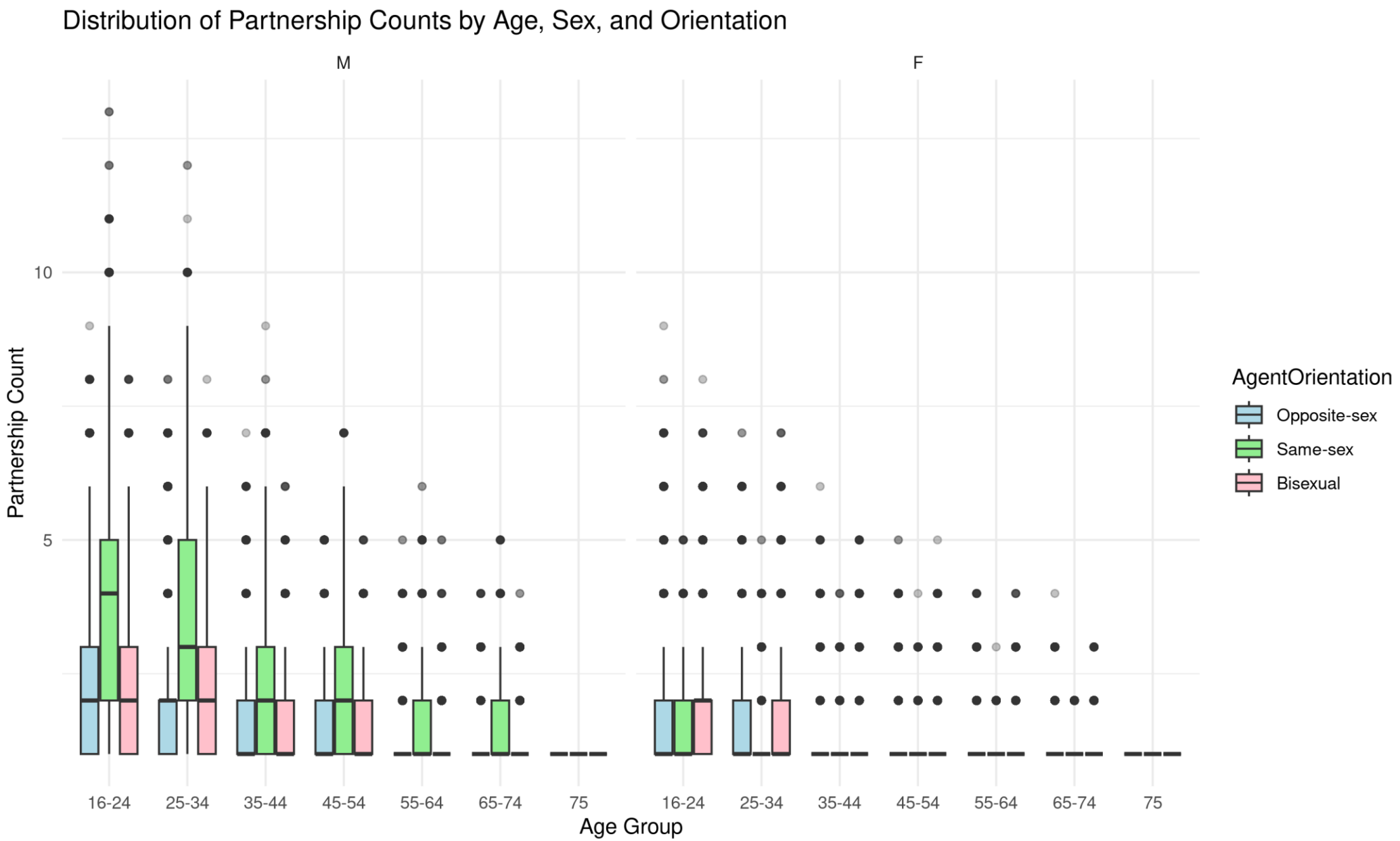


ABM Project – Sexually transmitted Infections (STIs)

Dynamic agent-based network model of sexual partnerships

- Partnership formation and duration-dependent partnership dissolution
- Sex and sexual orientation-specific mixing rules
- Dynamic aging of agents and age-assortative mixing
- Varying levels of concurrency
- Bridging bisexual individuals who form partnerships between same-sex and heterosexual individuals
- Latin Hypercube Sampling (LHS) for model calibration

STI-NET – Dynamic sexual partnership model



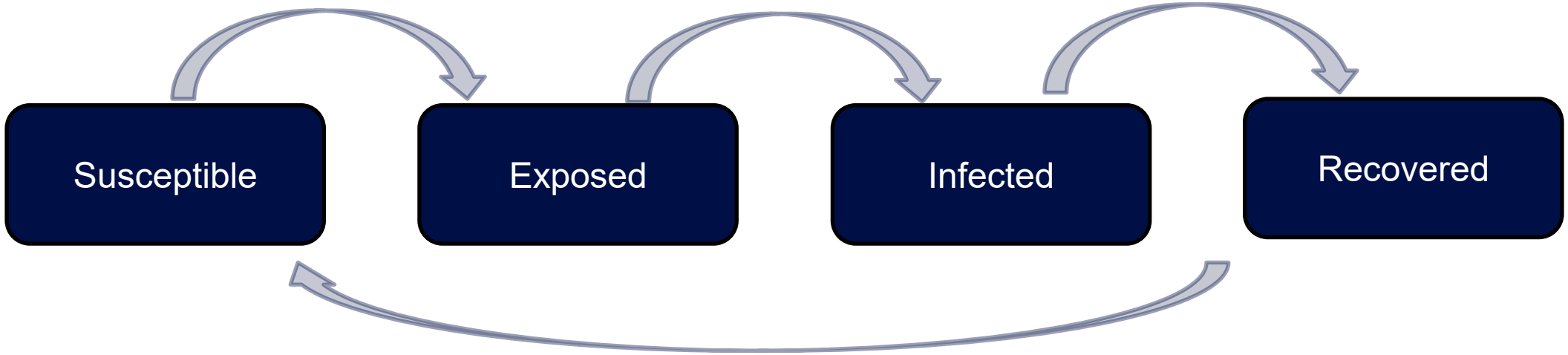
[16]

ABM Project – Sexually transmitted Infections (STIs)

Simulating chlamydia transmission on dynamic networks

Includes individual level condom use probability and frequency of sex acts





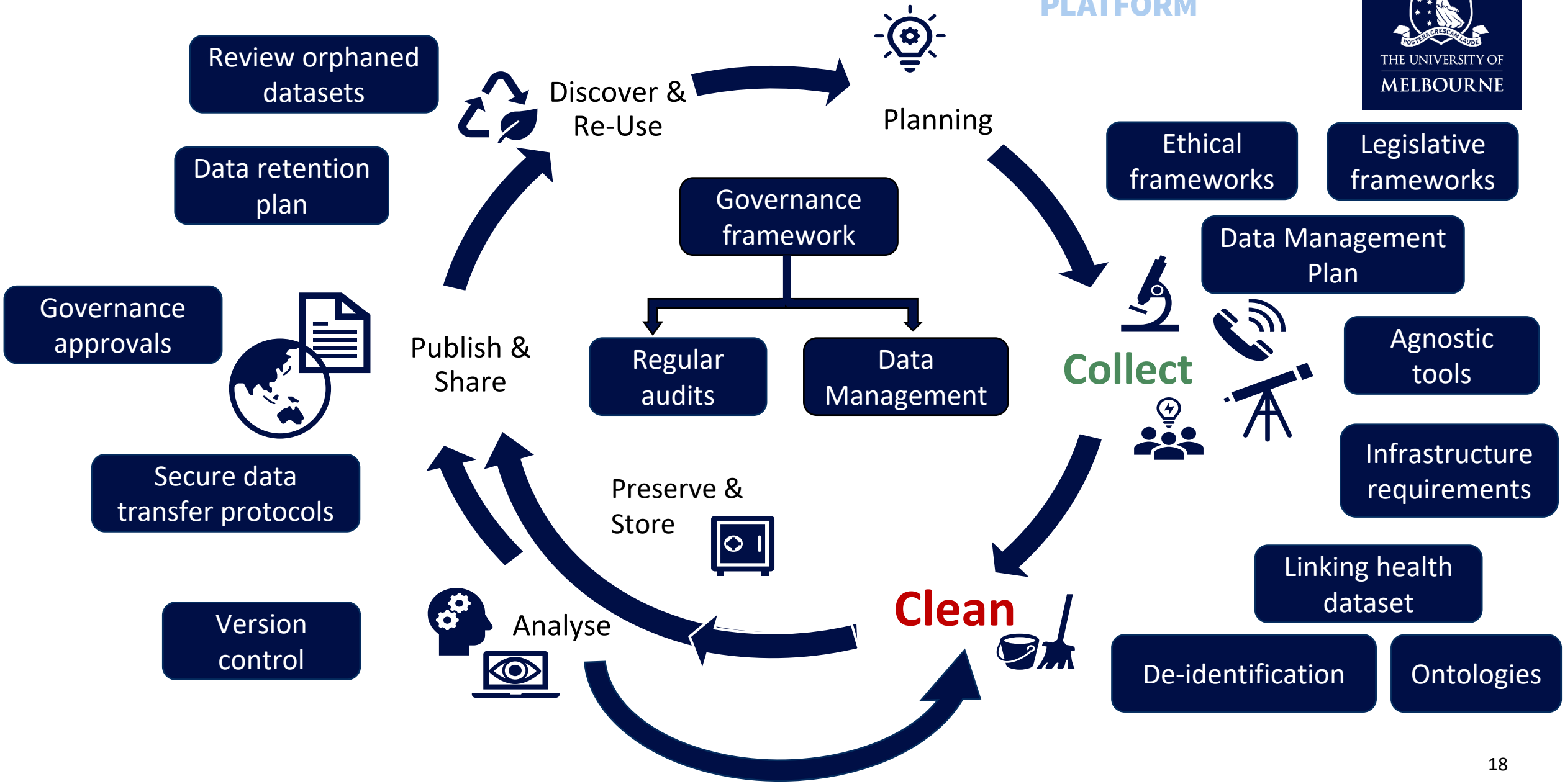
Transmission depends on age, sex, orientation, and condom use probability

The model calculates proportion of individuals infected

Health Data Stewardship

MELBOURNE DATA ANALYTICS PLATFORM

THE UNIVERSITY OF MELBOURNE





Prediction Post - Operative Complications PRECAST

THE UNIVERSITY OF MELBOURNE



Dr Noel Faux

Senior Research Data Specialist

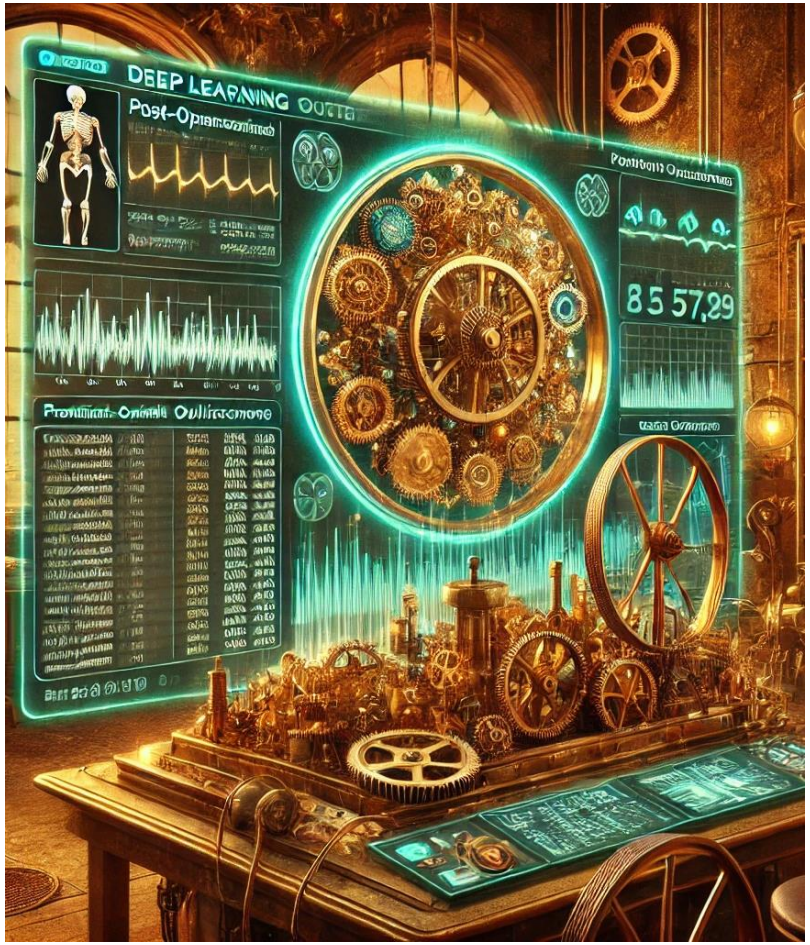
A/Prof Hilmy Ismail – Peter Mac Cancer Centre

Hidden pandemic – Postoperative Complications

- **A 3rd of the global burden of health requires surgery**
- **2/3 of the cancer patients undergo surgery**
- **Death after 30days of surgery is the third leading cause of death globally**
  - Driven by postoperative complications
- 1:4 patients suffer postoperative complications
  - Increase length of stay
- ~1.6M patients undergo major surgery for cancer
  - ~1M suffer postoperative complications
- Fortunately many postoperative complications are preventable
  - Requiring timely identification and risk stratification

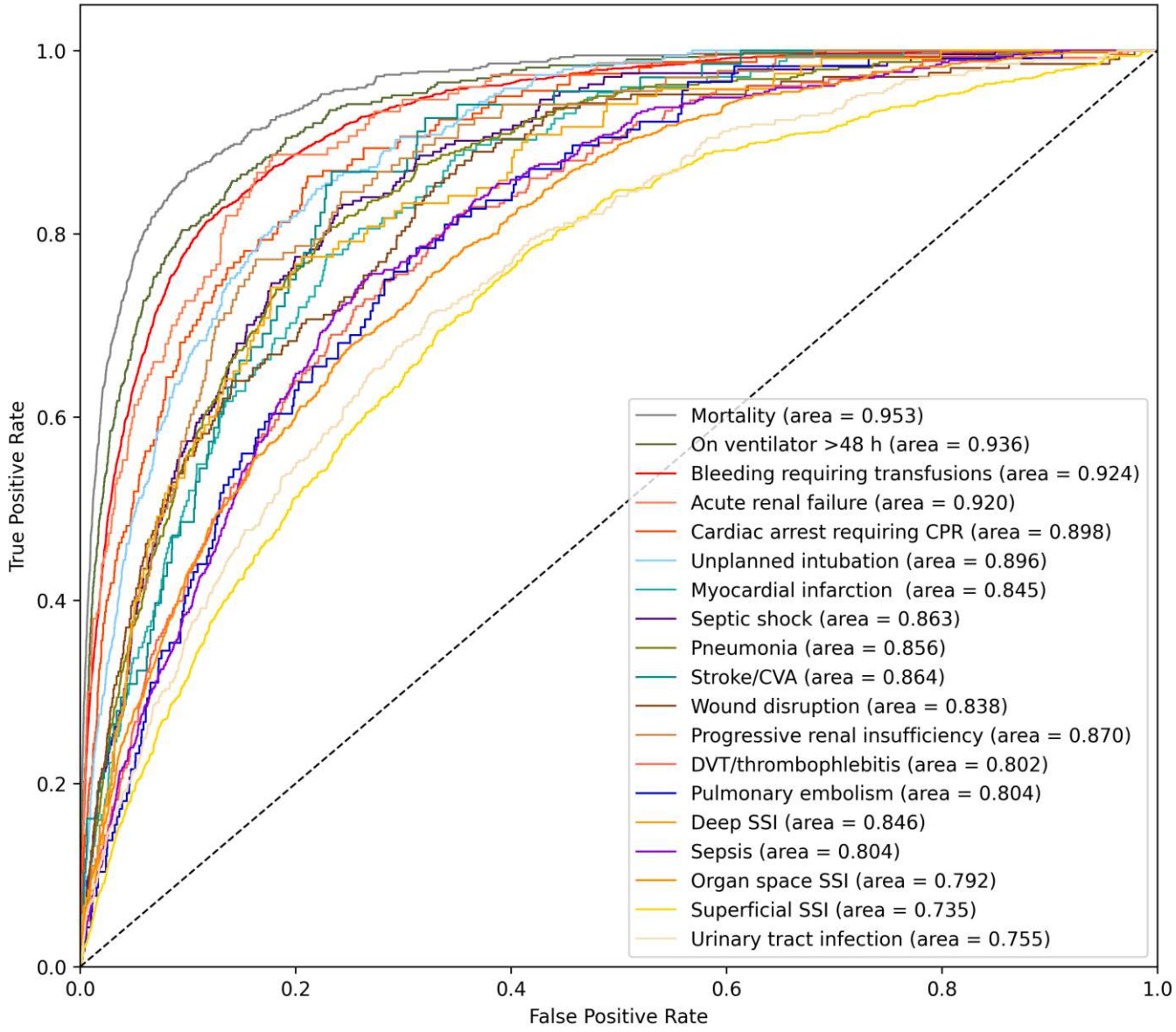
Prediction of postoperative complications

- Development of a Deep Learning model for prediction of POO-CS and identification of key predictors
- Used large international dataset of surgical outcomes - National Surgical Quality Improvement Program (NSQIP)
  - **Initially >2M records**
  - Trim down to only cancer surgery procedures
  - **757,468 records**
  - **19 post operative outcomes**, including morality
    - Superficial surgical site infection (ssi), deep ssi, Organ space ssi, wound disruptions, pneumonia, unplanned intubation, pulmonary embolism, On ventilator >48hr, progressive renal insufficiency, acute renal failure, urinary tract infection, stroke, cardiac arrest requiring CPR, myocardial infection, dvt, sepsis, septic shock, bleed, morality
  - **314** features/variable/predictors

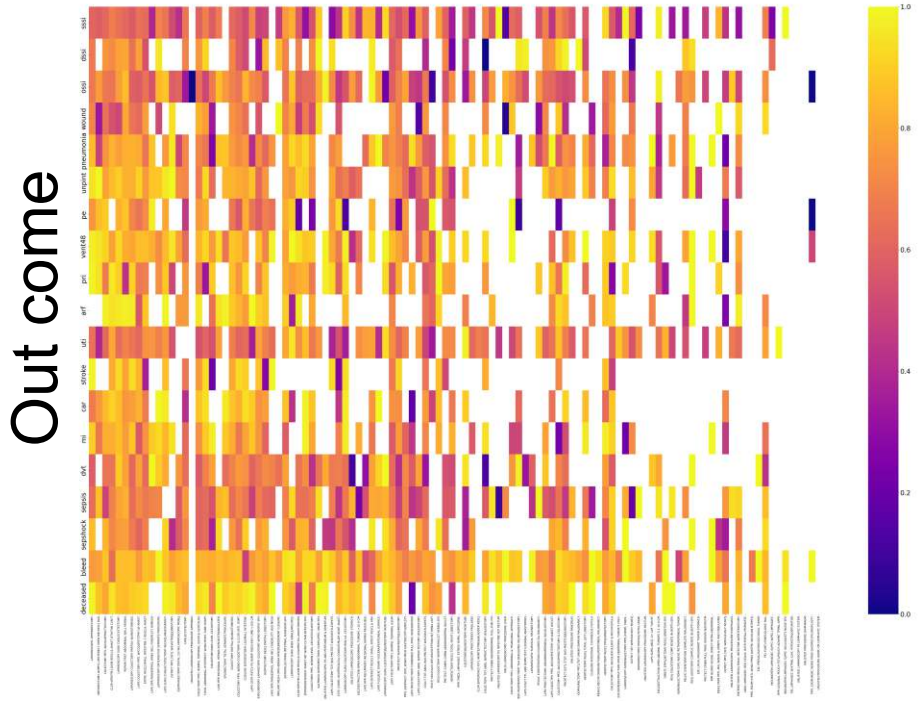


DALL.E 3

Results

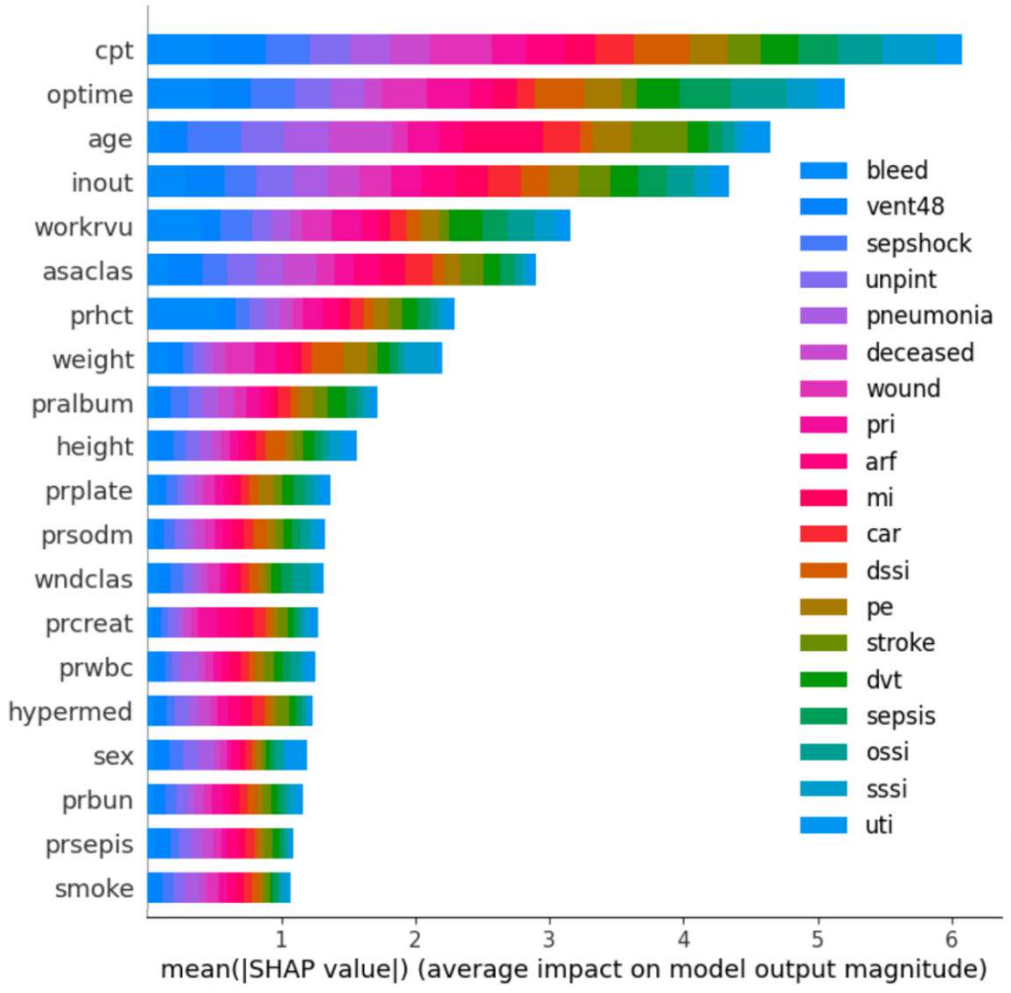

True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Mortality (area = 0.953)
On ventilator >48 h (area = 0.936)
Bleeding requiring transfusions (area = 0.924)
Acute renal failure (area = 0.920)
Cardiac arrest requiring CPR (area = 0.898)
Unplanned intubation (area = 0.896)
Myocardial infarction (area = 0.845)
Septic shock (area = 0.863)
Pneumonia (area = 0.856)
Stroke/CVA (area = 0.864)
Wound disruption (area = 0.838)
Progressive renal insufficiency (area = 0.870)
DVT/thrombophlebitis (area = 0.802)
Pulmonary embolism (area = 0.804)
Deep SSI (area = 0.846)
Sepsis (area = 0.804)
Organ space SSI (area = 0.792)
Superficial SSI (area = 0.735)
Urinary tract infection (area = 0.755)

AUC


Out come
Procedure

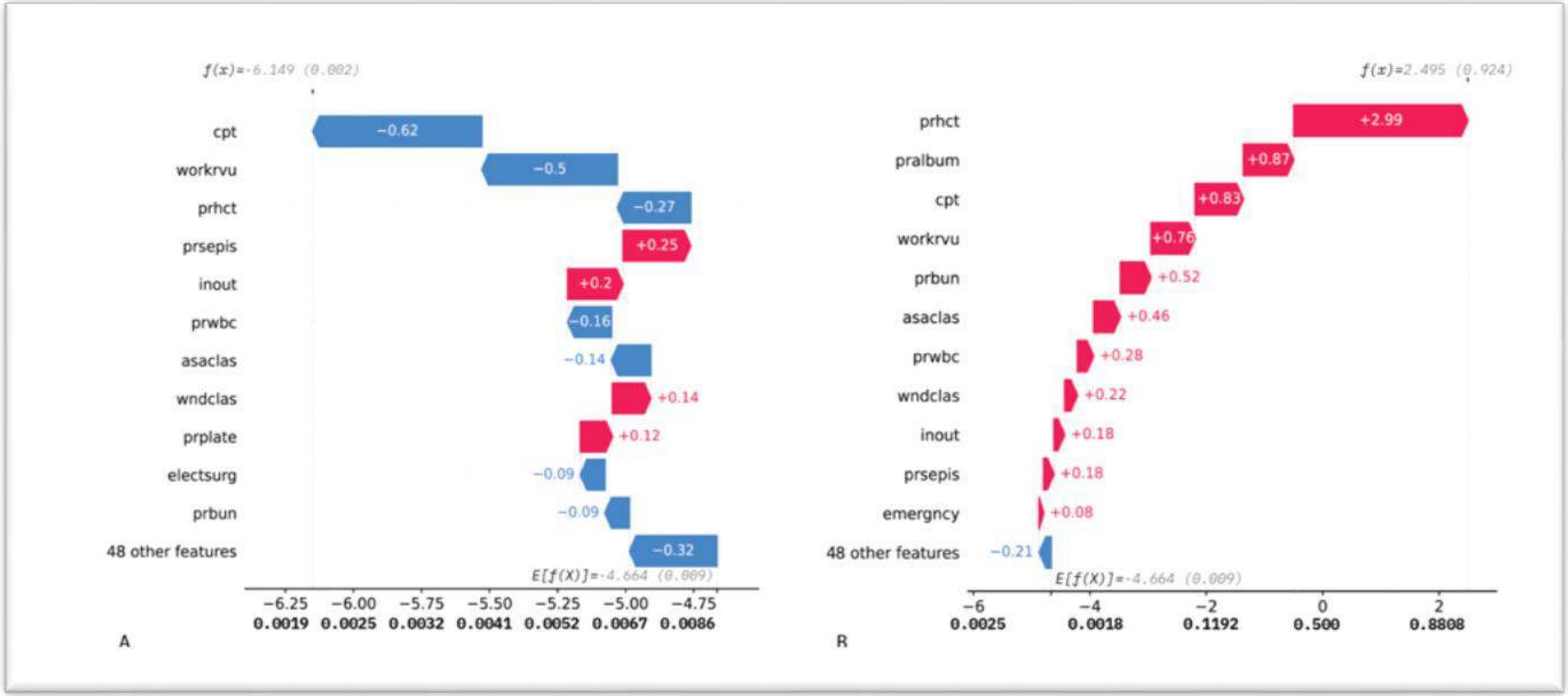
Predictive Features

- Top 20 predictive variables
- Number of preoperative blood markers
  - Platelets, haematocrit, albumin, sodium, white blood count
- Immunological focus
- Suggesting possible preoperative interventions



Personalised risk profile

Risk of postoperative bleeding



Localisation for clinical implementation

MRFF Research Data infrastructure Grant – PRECAST

- Project Governance structure
  - Project Control Group
    - Data Working Group
    - Analytics Working Group
    - Hybrid Trial Working. Group
    - Data and Safety Monitoring Board
- Multidisciplinary
  - Propel Health – commercial partner
  - Data Science
  - Implementation Science
  - RCT

Team:
Prof. Bernhard Riedel (CI A), A/Prof Hilmy Ismail, Prof Kate Burbury, Dr Helen Mohan, Prof. Linda Denehy, Dr Emma Tyson, Prof. Alexander Heriot, A/Prof. Stephanie Best, Dr Lisa Guccione, Dr Jason Li, Erminia Schiavone, Matthew Ha, Huw Thomas, Dr Thomas Poulton, Dr Noel Faux, A/Professor Tim Spelman, Dr Gareth Jones, Dr Julia Dubovitch, Sharmaine McGlin and Propel Health



THE UNIVERSITY OF MELBOURNE



Dr Mel Mistica

Senior Research Data Specialist

Behavior Research Methods
https://doi.org/10.3758/s13428-024-02385-5

ORIGINAL MANUSCRIPT

A natural language model to automate scoring of autobiographical memories

Meladel Mistica[1] · Patrick Haylock[2] · Aleksandra Michalewicz[1] · Steph Raad[2] · Emily Fitzgerald[1] · Caitlin Hitchcock[2]

Accepted: 26 February 2024
© The Author(s) 2024

Abstract

Biases in the retrieval of personal, autobiographical memories are a core feature of multiple mental health disorders, and are associated with poor clinical prognosis. However, current assessments of memory bias are either reliant on human scoring, restricting their administration in clinical settings, or when computerized, are only able to identify one memory type. Here, we developed a natural language model able to classify text-based memories as one of five different autobiographical memory types (specific, categoric, extended, semantic associate, omission), allowing easy assessment of a wider range of memory biases, including reduced memory specificity and impaired memory flexibility. Our model was trained on 17,632 text-based, human-scored memories obtained from individuals with and without experience of memory bias and mental health challenges, which was then tested on a dataset of 5880 memories. We used 20-fold cross-validation setup, and the model was fine-tuned over BERT. Relative to benchmarking and an existing support vector model, our model achieved high accuracy (95.7%) and precision (91.0%). We provide an open-source version of the model which is able to be used without further coding, by those with no coding experience, to facilitate the assessment of autobiographical memory bias in clinical settings, and aid implementation of memory-based interventions within treatment services.

Autobiographical Memory

Retrieval of stored memories about your own self

- Formulation of self-identity
- Autobiographical knowledge is stored in a fluid manner
- There are different classifications of memories that are characteristic of their type, for example
  - 'specific': single incident events, isolated in space and time, detailed, eg:
    "Last year, I went to the zoo with my family and as I was standing in front of lion enclosure for a photo, one of the lions gave a big roar, which made me scream involuntarily."
  - 'categoric': characterise a type of event, eg
    "An excursion to the zoo as a child is pretty exciting ."
  - 'extended': an event over an extended period of time, eg
    "Having to live with my grandparents."

Autobiographical Memory

The ability (or inability) to retrieve 'specific' memories has been implicated in mental health disorders, such as:

- the onset of symptoms related to depression – in pre-symptomatic individuals (Askelund et al, 2019)
- being a marker of higher chronicity of prolonged symptoms – with the reduced ability to retrieve specific memories, (Williams & Broadbent, 1986)
- increased frequency of depressive episodes and attempts at self-harm (Williams & Broadbent, 1986)

Testing Autobiographical Memory

Crucial for preventative and early intervention

Important for ascertaining treatment plans

Autobiographical Memory Task (AMT)

- AMT is a cued recall task
- Participants are asked to verbally report or type up a memory that comes to mind in response to the cue word
- The cue words are categorised as 'positive', 'neutral', or 'negative'
  - 'positive': happy, excited, relieved
  - 'neutral': train, chair, mug
  - 'negative': embarrassing, sad, frightening
- Scoring of the AMT is traditionally reliant on human interpretation
- Manual, time-consuming process
- Trained scorers but fatigue and error can occur and so there can be inter-rater variability

Previous Work on Automating the AMT

- Wardell et al (2021): automatically transcribing participant responses and arriving at heuristics to assist human scorers
- Takano et al (2017, 2018): SVM to automate the scoring task
  - English-speaking children
  - Dutch-speaking adults
  - Japanese-speaking adults
- Takano et al (2019): SVM to automate the scoring task for English-speaking adults trained on 10K memories (p=0.80; r=0.82; f1=0.81)
- van Genugten & Schachter (2022): DistilBert model that only identified specific elements in an overall memory

Previous Work on Automating the AMT

- Wardell et al (2021): automatically transcribing participant responses and arriving at heuristics to assist human scorers
- Takano et al (2017, 2018): SVM to automate the scoring task
  - English-speaking children
  - Dutch-speaking adults
  - Japanese-speaking adults
- Takano et al (2019): SVM to automate the scoring task for English-speaking adults trained on 10K memories (p=0.80; r=0.82; f1=0.81)
- van Genugten & Schachter (2022): DistilBert model that only identified specific elements in an overall memory

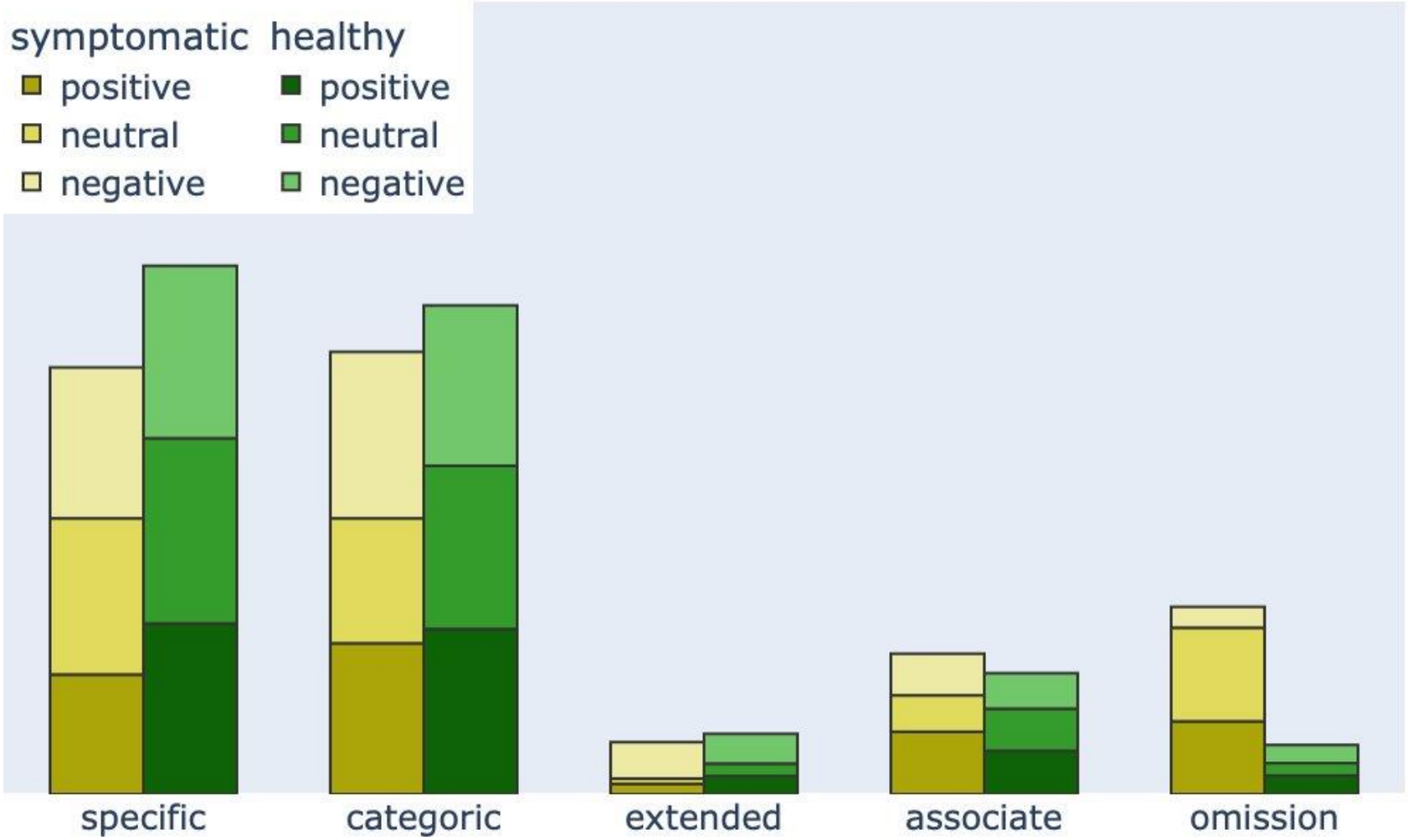
Autobiographical Memory Task (AMT)

- Takano et al (2019): SVM to automate the scoring task for English-speaking adults trained on 10K memories (p=0.80; r=0.82; f1=0.81)
- Has become the default tool to assist in automating the scoring process
- Data had a 5 categories
  - **specific, categoric, extended, associate, omission**
- 1000 participants, each were asked to recall 10 specific memories
- They created a binarised version of the data, collapsing the 4 non-specific memory types:
- specific vs non-specific memory
- However, clinical psychologists and researchers would like to have the more fine-grained non-specific categories ability

Autobiographical Memory Task (AMT)

- We set out to build a 5-way classifier and a binary classifier
- We gathered 3 datasets:
  - Dataset A: 10K memories from Takano et al (2019) – takano
  - Dataset B: 6.2K memories, hand scored at UniMelb – amt-ai
  - Dataset C: 1.5K memories from Cambridge University – hitchcock
- Fine-tuned the above datasets on BERT to see if we could improve upon Takano (2019)

Autobiographical Memory Task (AMT)



Autobiographical Memory Task (AMT)

- We set out to build a 5-way classifier and a binary classifier
- We gathered 3 datasets:
  - Dataset A: 10K memories from Takano et al (2019) – takano
  - Dataset B: 6.2K memories, hand scored at UniMelb – amt-ai
  - Dataset C: 1.5K memories from Cambridge University – hitchcock
- Fine-tuned the above datasets on BERT to see if we could improve upon Takano (2019)

Binarised Task

DATASET SIZE	DatasetA: *takano* 10.0K			DatasetB: *marsh* 1.5K			DatasetC: *amt-ai* 6.1K			All: *combined* 17.7K		
	P	R	F	P	R	F	P	R	F	P	R	F
BERT 0.45	.91	.91	.91	.92	.92	.92	.90	.90	.90	.89	.89	.89
BERT 0.50	.91	.91	.91	.92	.92	.92	.89	.89	.89	.89	.89	.89
BERT 0.55	.91	.91	.91	.92	.92	.92	.89	.89	.89	.89	.89	.89
TAKANO	.80	.82	.81	.72	.71	.70	.72	.72	.71	–	–	–
SVM	.69	.71	.69	.61	.66	.63	.69	.68	.69	.78	.78	.78
MAJORITY	.55	.55	.55	.50	.50	.50	.59	.59	.59	.50	.50	.50

5-way Results

DATASET SIZE	DatasetA: *takano* 10.0K			DatasetB: *marsh* 1.5K			DatasetC: *amt-ai* 6.1K			All: *combined* 17.7K		
	P	R	F	P	R	F	P	R	F	P	R	F
BERT 0.45	.91	.89	.90	.80	.83	.82	.94	.93	.93	.87	.87	.87
BERT 0.50	.92	.88	.89	.83	.81	.82	.94	.92	.93	.88	.85	.87
BERT 0.55	.92	.86	.88	.84	.79	.81	.95	.92	.93	**.91**	.86	**.89**
TAKANO	–	–	–	–	–	–	–	–	–	–	–	–
SVM	.78	.78	.78	.76	.76	.76	.78	.78	.78	.68	.68	.68
MAJORITY	.55	.55	.55	.50	.50	.50	.41	.41	.41	.50	.50	.50

Breakdown of classes

	P	R	F	DATA
Specific	.95	.92	.93	8802
Extended	.89	.77	.83	2852
Categoric	.86	.88	.87	4165
Omission	.99	.80	.88	687
Associate	.84	.66	.74	1113

Future Work in AMT

- More testing has been done create an even larger dataset
- Approached by mental health researchers to incorporate our model into their systems and to assist in building their own for other languages
- Plans to employ our updated model to assess interventions

Spin-off study

- The theory that how we express past events is a "window" into the "organisation" of our memories is further explored with the PTSD project
- Examine linguistic/language usage that are indicative of PTSD
- Create models for clinical practitioners to explore language usage in patients that have experienced difficult events to identify features of PTSD

Thank you

Melbourne Data Analytics Platform (MDAP)



THE UNIVERSITY OF MELBOURNE

MELBOURNE DATA ANALYTICS PLATFORM

www.unimelb.edu.au/mdap
mdap-info@unimelb.edu.au
@MDAP_Unimelb

The University of Melbourne (Australian University) PRV12150/CRICOS 00116K