Impacts of GenAI on Research:

The Good, the Bad, and the Ugly

Professor Karin Verspoor
Executive Dean, School of Computing Technologies
STEM College
RMIT University
Melbourne, Australia

RMIT UNIVERSITY

Acknowledgements

Australian Government
Australian Research Council

ARC

ELSEVIER

NHMRC
BUILDING A HEALTHY AUSTRALIA

BioGrid
Connecting health information

Opyl

Australian Government
Department of Health and Aged Care

AstraZeneca

ARC TRAINING CENTRE
IN COGNITIVE COMPUTING
FOR MEDICAL TECHNOLOGIES

digital health crc

Pfizer

RMIT
UNIVERSITY

The Ugly

RMIT UNIVERSITY

“Salami Slicing”

aka the minimal publishable unit

Knee Surg Sports Traumatol Arthrosc (2013) 21:751–752
DOI 10.1007/s00167-013-2413-3

EDITORIAL

**Legitimate division of large data sets, salar
publication, where does a fraud begin?**

Jon Karlsson · Philippe Beaufils

Salami slicing consists of splitting data from t
units, each of which is submitted—and in ma

In **dual publication**, the authors knowingly su
identical data to two different journals, either

Advice: https://publicationethics.org/case/salami-publication

- A distinction needs to be made between salami and redundant publication: where there is a two thirds overlap, it is redundant publication.
- If the hypotheses were completely separate questions then it is acceptable for them to be posed in two separate papers.
- If they are related questions, or very closely related, then they should be published as a single paper.
- Salami publication is where papers cover the same population, methods, and question.
- Splitting up papers by outcomes is not legitimate.
- It is an editorial decision as to whether to publish or not: there is no ethical problem here.

“Salami Slicing”

aka the minimal publishable unit

GenAI will contribute to

- **rapid generation of minor paper variants**
- **with harder-to-detect (self-) plagiarism**

RMIT UNIVERSITY

Superfluous citations

Some of the most common algorithms include A1[32,33,34], A2[35,36,37], A3[38], A4[39,40], A5[41], A6[42], A7[43], A8[44,45,46], A9[47], A10[48], A11[49], and A12[50].

Ref	Algorithm	Year
[51]	A13	2019
[51]	A14	2019
[52]	A15	2020
[53]	A16	2020
...	...	...
[112]	A85	2024

Table 1. A brief review of X algorithms in the last 5 years.

Highly cited researchers are also suspected of participating in so-called citation cartels or citation rings, whereby a group of researchers dishonestly cite each other's work. In Iran, some even speak of publication and citation mafias, whose members not only trade citations but can even manipulate journals' refereeing processes.

[1,102,103,104,105,106,107,108,109,110,111,112,113,114]

Citation counting is encouraging cheating in Iran

THE Times Higher Education

Ministers' metric-based boasts about the country's scientific prowess are belied by the reality, as a recent incident illustrates, says Roohola Ramezani

June 11, 2023

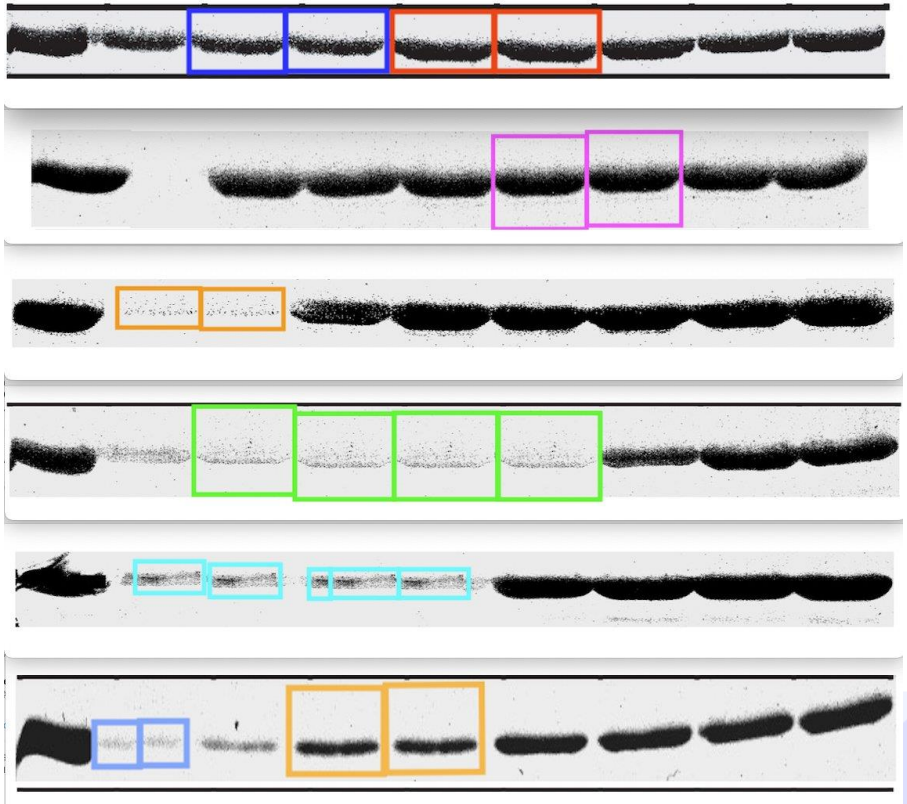
RMIT UNIVERSITY

Superfluous c

GenAI will contribute to

- **collection and manage superfluous reference**
- **automatic generation c**

Image manipulation



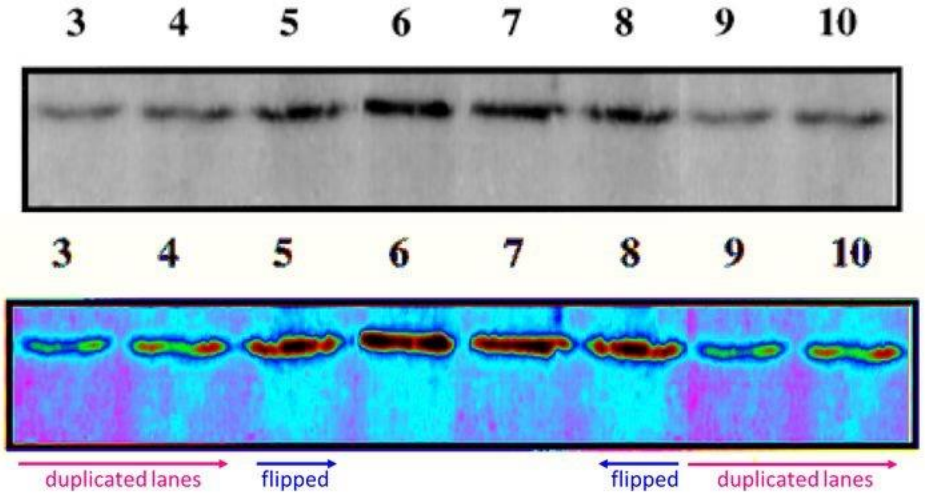
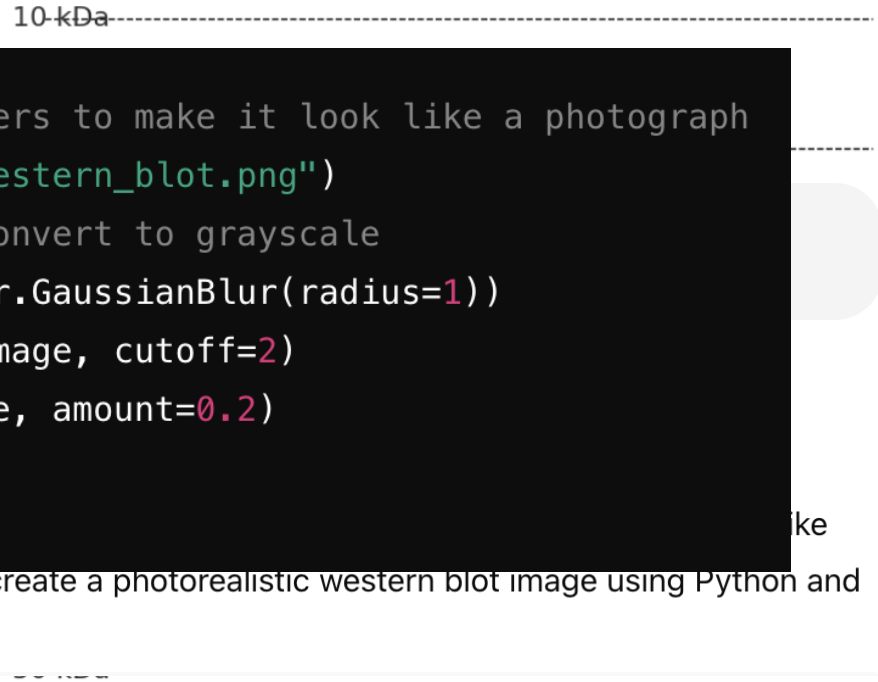


Figure 10 – lower panel only; original and false color
Molecular Plant-Microbe Interactions (2000)
doi: 10.1094/mpmi.2000.13.10.1053

RMIT UNIVERSITY

Image

Realistic Western Blot for Borrelia burgdorferi OspA




```
# Open the image and apply filters to make it look like a photograph
image = Image.open("/mnt/data/western_blot.png")
image = image.convert("L")  # Convert to grayscale
image = image.filter(ImageFilter.GaussianBlur(radius=1))
image = ImageOps.autocontrast(image, cutoff=2)
image = ImageOps.add_noise(image, amount=0.2)

image.show()
```

ike

step-by-step instructions on how to create a photorealistic western blot image using Python and image processing libraries?

RMIT UNIVERSITY

The role of Generative AI

- Easy generation of text paraphrases
- Easy generation of fabricated images
- Easy generation of fabricated data/tables
- Generation of text describing data
- Instant literature reviews

CAUTION: GenAI doesn’t understand numbers

CAUTION: nor any meta-knowledge about the text

RMIT UNIVERSITY

The Bad (?)

RMIT UNIVERSITY

Literature Reviews

GenAI will contribute to

- automatic synthesis of literature
- with sources via Retrieval-Augmented Generation

CAUTION: synthesis relies on critical reading

CAUTION: reviews require meta-analysis and reasoning

RMIT UNIVERSITY

Writing papers

GenAI will contribute to

- generating summary abstracts
- explaining results
- producing multi-lingual abstracts

CAUTION: hallucinations and confabulations

RMIT UNIVERSITY

Doing science

GenAI will contribute to

- analysing data
- data curation
- via “instruction” prompting

CAUTION: hallucinations and confabulations

CAUTION: experiments suggest weak data curation skills

RMIT UNIVERSITY

Peer review & assessment

GenAI will contribute to

- generating critical summaries
- expanding short notes into longer comments

CAUTION: IP leakage

CAUTION: “generic” reviews

RMIT UNIVERSITY

Proposal Writing

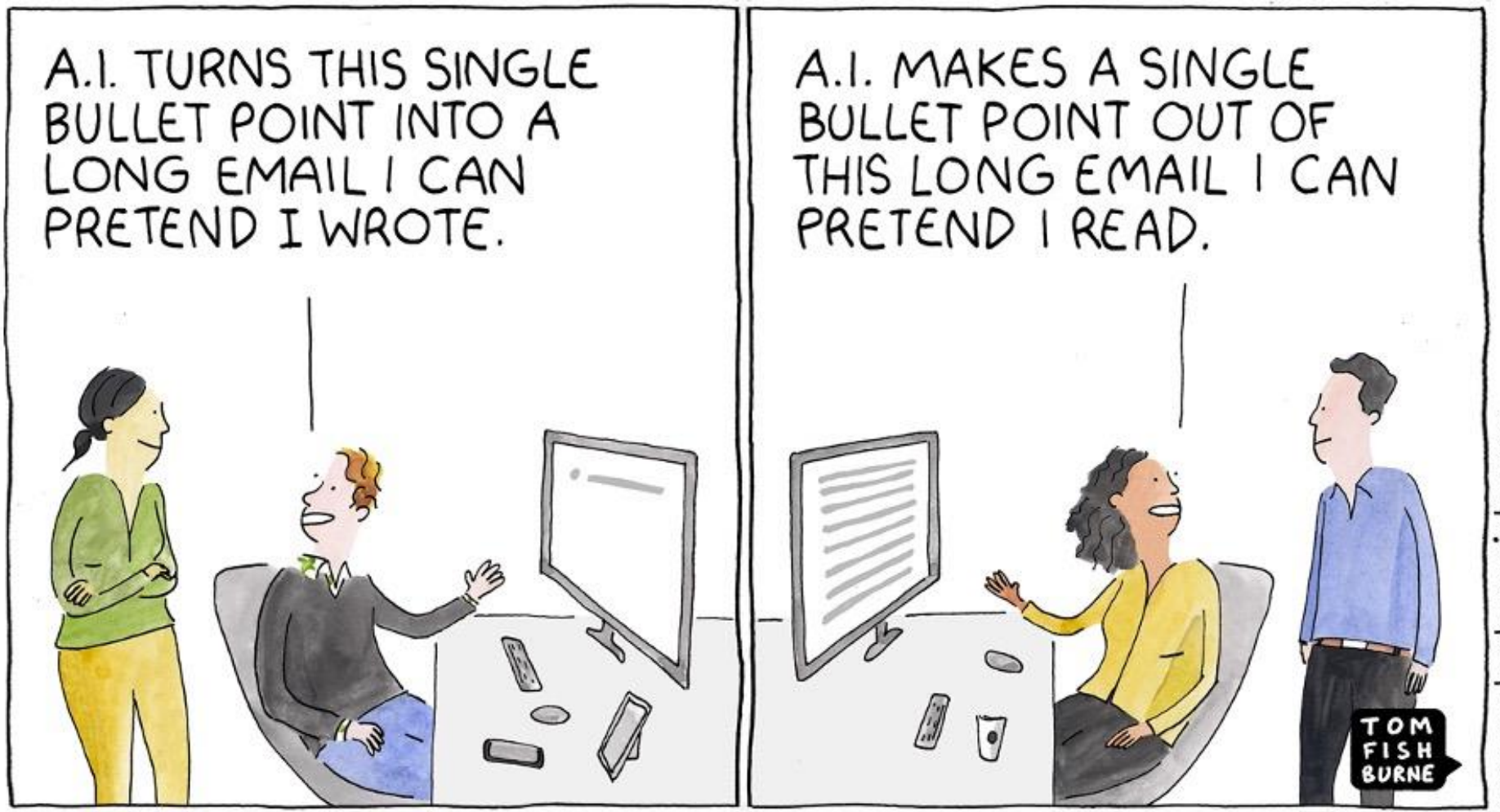
GenAI will contribute to

- Rewriting previous proposals for new funding scheme
  - Adding in constraints (e.g. page length)
  - Addressing specific elements/criteria
- Updating literature review

CAUTION: IP leakage

CAUTION: Novelty?

RMIT UNIVERSITY


A.I. TURNS THIS SINGLE BULLET POINT INTO A LONG EMAIL I CAN PRETEND I WROTE.
A.I. MAKES A SINGLE BULLET POINT OUT OF THIS LONG EMAIL I CAN PRETEND I READ.
TOM FISH BURNE
© marketoonist.com

The Good

RMIT UNIVERSITY

AI for fraud detection

GenAI will contribute to combatting fraud

- image duplication detectors
- paraphrase / text similarity measurement
  - better plagiarism detection
  - document similarity / clustering
- (in)appropriate citation classification
- fact and consistency checking

RMIT UNIVERSITY

AI for writing

GenAI will contribute to supporting writing

- **grammar and fluency checking**
- **polishing text**
- **drafting content**

levelling the playing field for non-native English speakers?

RMIT UNIVERSITY

AI for reading

GenAI will contribute to supporting reading

- **machine translation of content between languages**
- **improved querying of literature databases**
- **integration of knowledge and databases with literature**

levelling the playing field for non-native English speakers?

finding more relevant papers

RMIT UNIVERSITY

AI for reviewing

GenAI will contribute to supporting reviewing

- **improved querying of literature databases**
  - **making more of the literature findable**
  - **“conversational” search**
- **screening studies / relevance detection**

RMIT UNIVERSITY

AI for hypothesis generation

GenAI will contribute to supporting new science

- **connecting dots between things in the literature**
- **injecting stochasticity**
  - **putting things together in unexpected ways**

RMIT UNIVERSITY

Beyond Generative AI

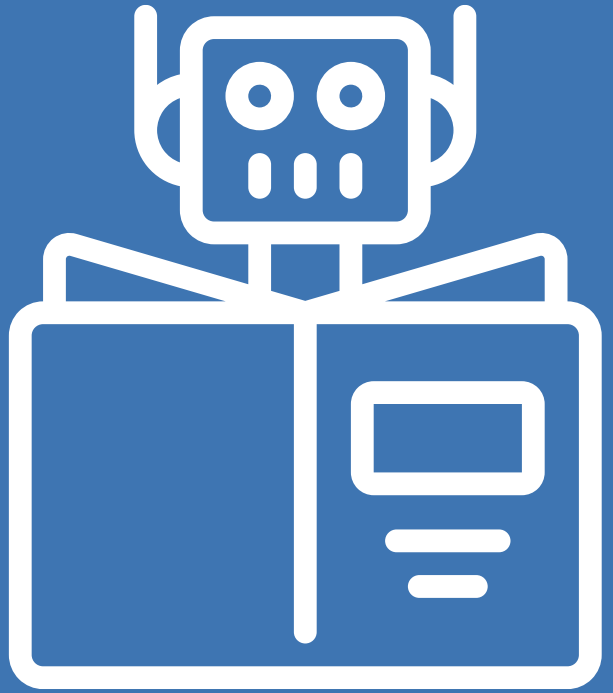
Not all AI is generative!

- **classification**
- **prediction**
- **clustering**
- **data mining/finding patterns**

RMIT UNIVERSITY

AI *supporting* research

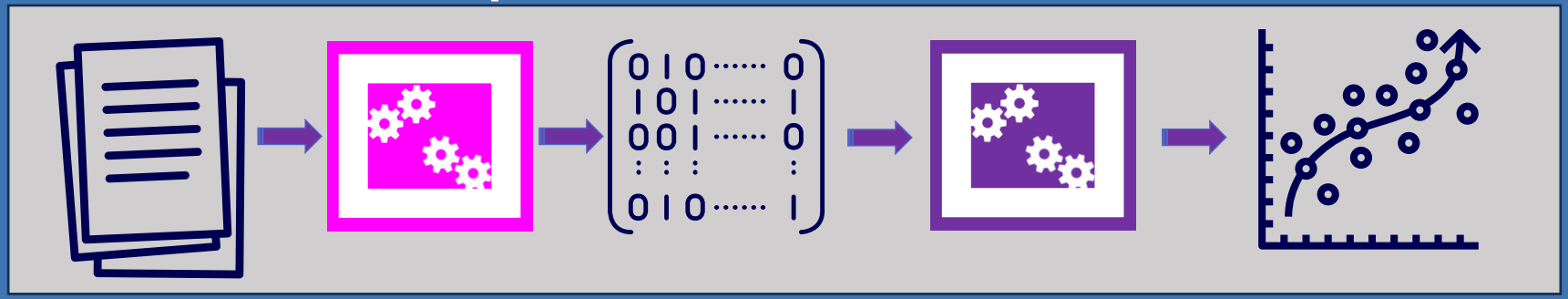
- **Predictive modelling**
- **Evidence detection** enabling *concept* search
- **Evidence exploration** tools allowing more open-ended literature navigation
- **Evidence summarization** and **synthesis**
- **Evidence discovery**



Predictive modelling

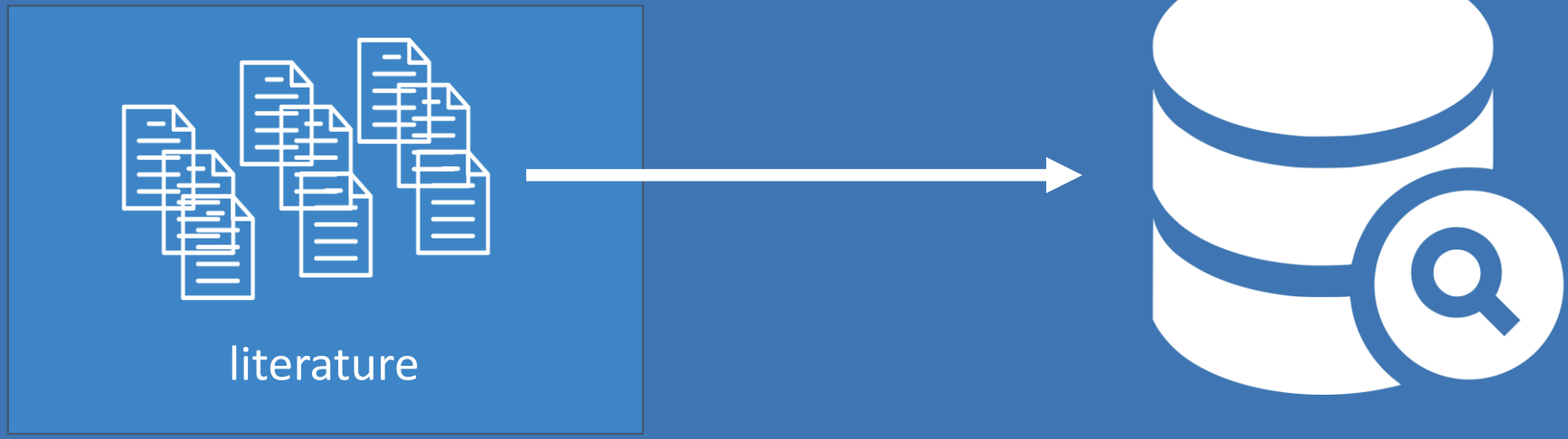
- **AlphaFold**: 3D protein structure prediction at scale
- **Drug Discovery**: protein binding prediction
- **GNoME**: predicting new material stability
- **FunSearch**: search for new mathematical and algorithmic solutions

LLMs for text *representation*
in downstream *prediction* tasks



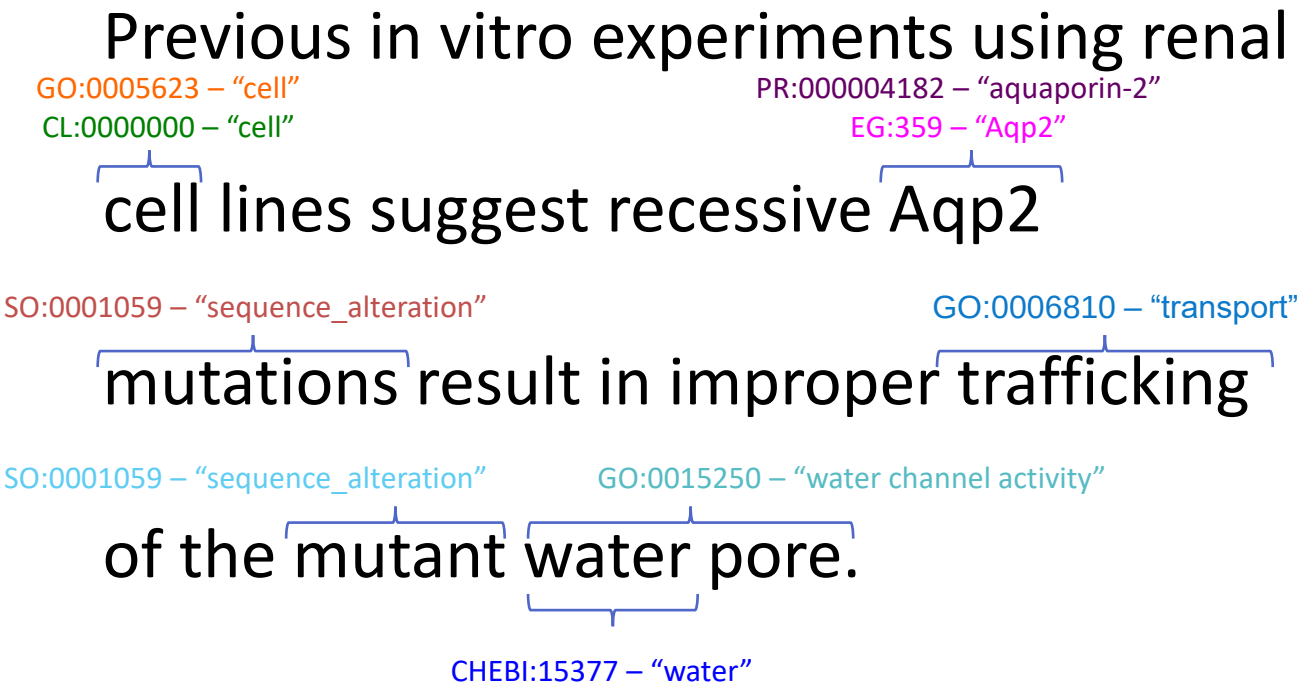
- document classification (histopathology report → infection Y/N?)
- sentence classification (does this sentence describe a disease?)
- outcome prediction (will patient survive 30 days?)
- event detection
- (has patient experienced an adverse drug reaction?)
- (does patient have disease X?)
- information extraction (what is the smoking status of the patient?)

Organising knowledge



- Find key concepts, entities and events
- Map to standard identifiers and/or ontology terms
- Support indexing and retrieval

Recognising biological ontology concepts



Funk C, Baumgartner WA, Garcia B, Roeder C, Bada M, Cohen KB, Hunter LE, Verspoor K. (2014)
Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters. BMC Bioinformatics 15(1):59.

Gene Ontology vs Natural Language

- Variation in PMID: 12925238

[Term]

id: GO:0006900

name: membrane budding

…

def: "The evagination of a membrane, resulting in formation of a vesicle."

…

synonym: "membrane evagination"

synonym: "nonselective vesicle assembly"

synonym: "vesicle biosynthesis"

synonym: "vesicle formation"

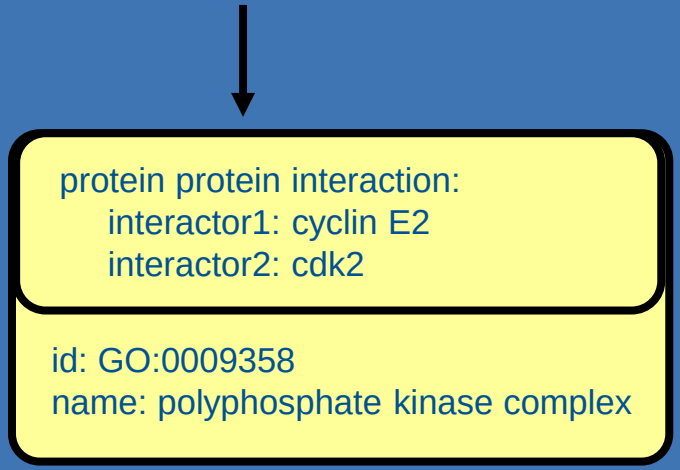
…

- Lipid rafts play a key role in **membrane budding**…
- …involvement of annexin A7 in **budding of vesicles**…
- …Ca2+-mediated **vesiculation process** was not impared.
- Red blood cells which lack the ability to **vesiculate** cause…
- Having excluded a direct role in **vesicle formation**…

Structuring relations

- Capturing entities and relations
  - “PROTEIN interacts with PROTEIN”
  - “CHEMICAL treats DISEASE”
  - “MUTATION causes DISEASE”
- Incorporating knowledge
  - cf. “ACE inhibitor treats hypertension”
  - \+ benazepril *–isa-* ACE inhibitor

Cyclin E2 interacts with Cdk2 in a functional kinase complex.



Organising knowledge enables semantic search

keywords → concepts

AND / OR co-occurrence → relations

literature

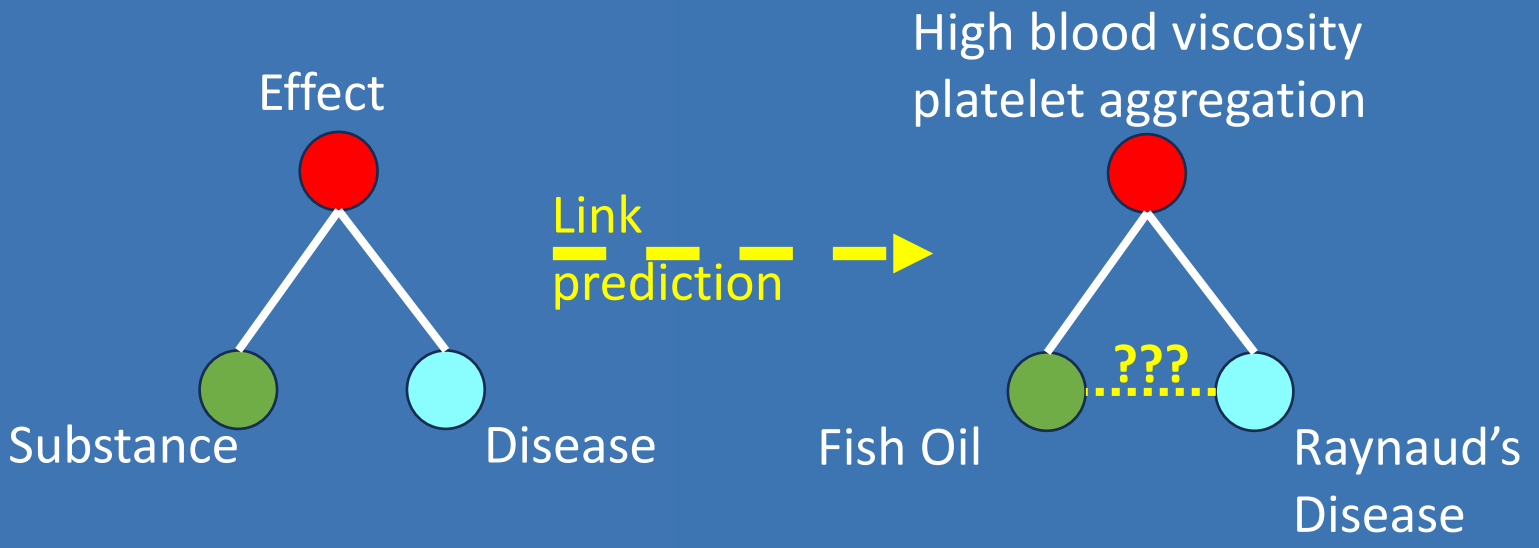
literature
augmented with concepts and relationships

Find papers on

[bariatric surgery]
[type 2 diabetes]
[remission]

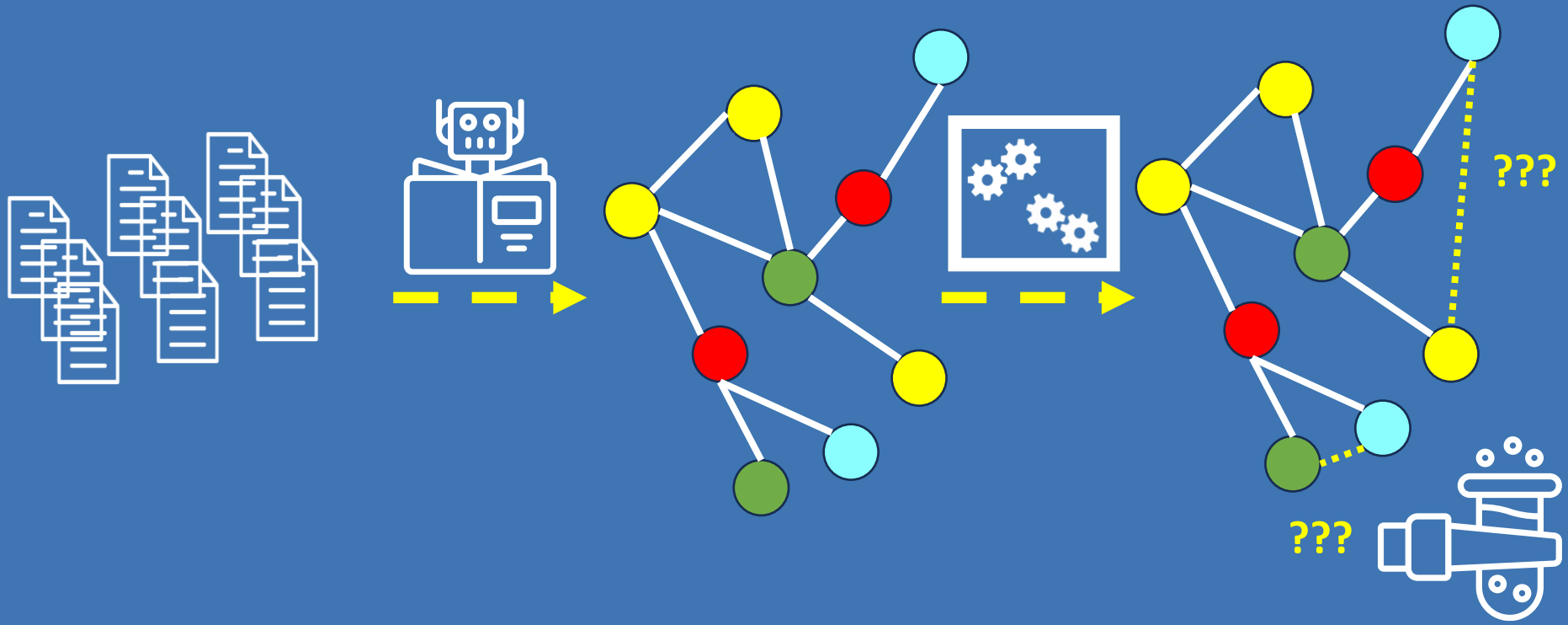
[Flurbiprofen]
[metabolized-by]
[CYP2C9]

Literature-based Discovery



→ Hypothesis generation

Literature-based discovery at scale



graph with thousands of nodes,
representing 20 years of research

→ lots of new hypotheses

What’s next for AI in Research?

Better AI tools!

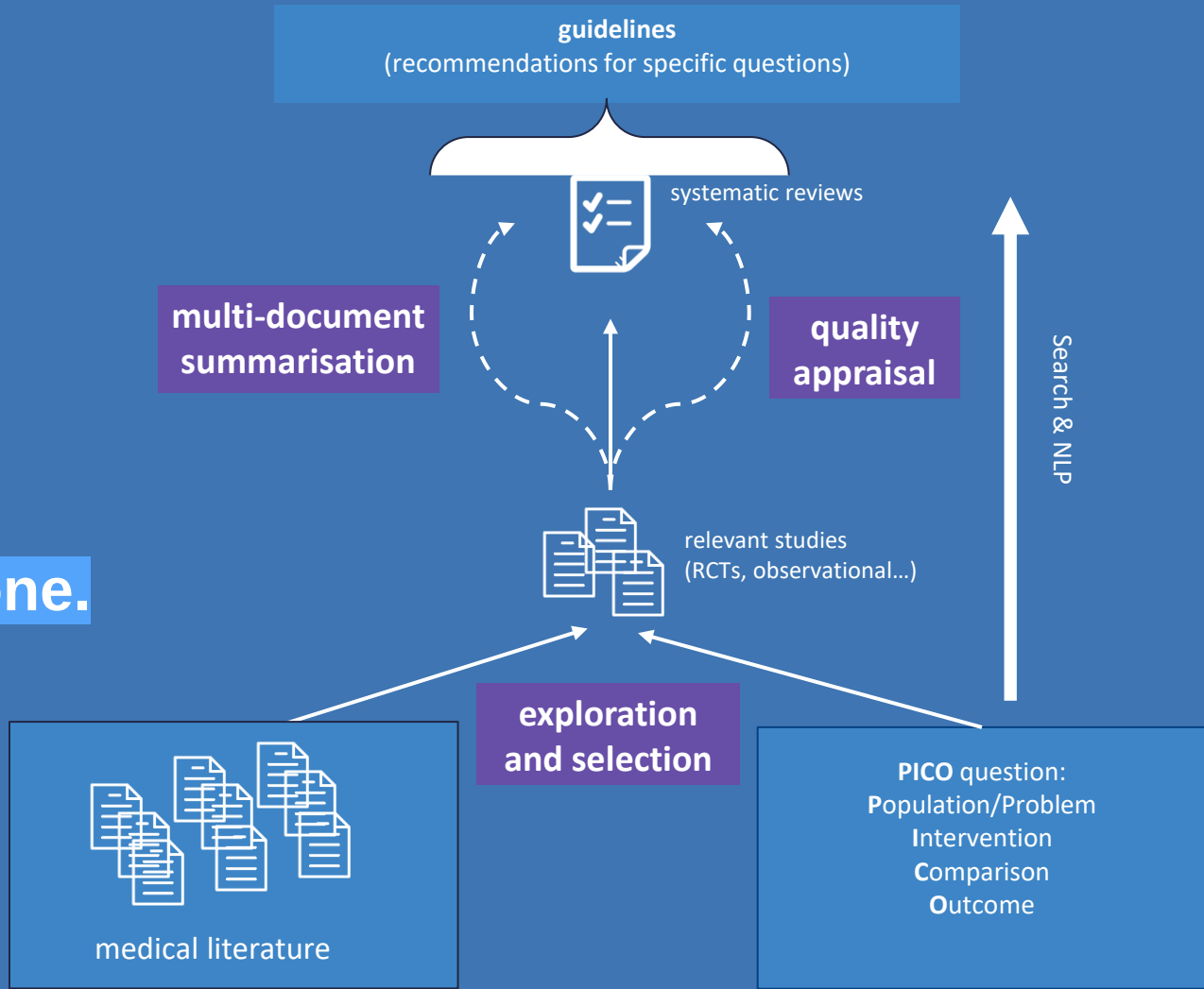
- **TARGETED to specific tasks**
- **multi-modal applications**
  - **text, images, sequences, structured data**
- **privacy-preserving modelling at scale**
- **systematic review automation**

RMIT UNIVERSITY

Systematic Review Automation

Making progress.

More work to be done.



What’s next for GenAI?

More robust GenAI

- rigorous evaluation of particular uses and applications
- knowledge-based methods
- watermarking & detecting GenAI outputs
- more energy efficient uses
- controlling confabulations
- grounding generations in sources

RMIT UNIVERSITY

AI for Good

Thank you!

RMIT UNIVERSITY