

# Reading and Critiquing Journal Articles

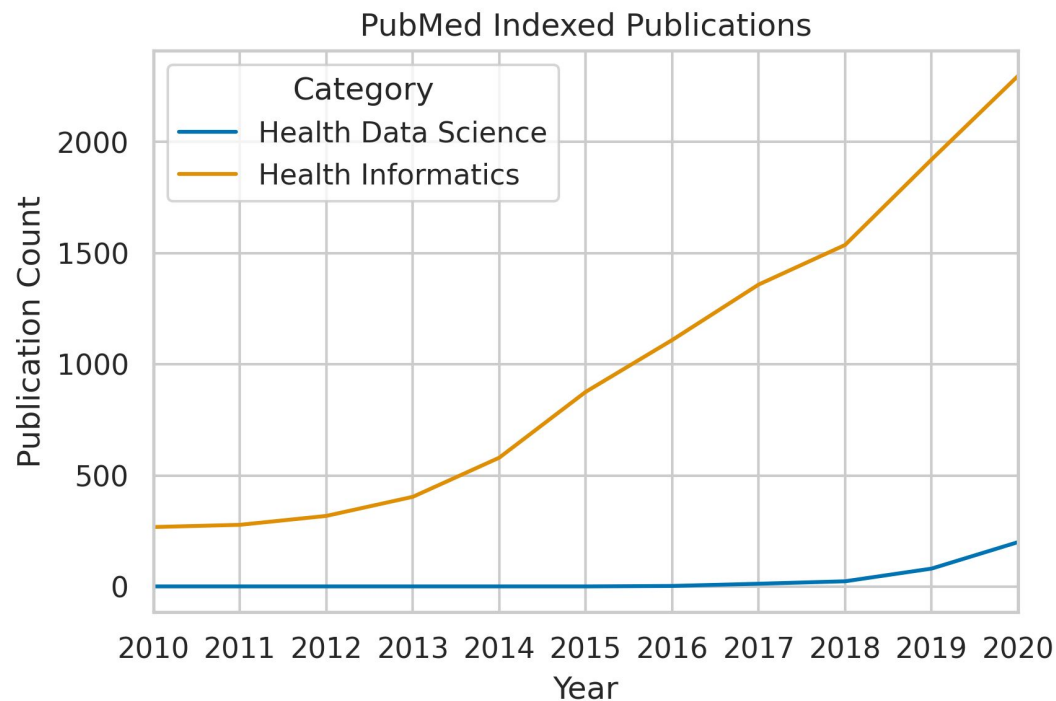
CSCI6410/4148/EPAH6410

Finlay Maguire (finlay.maguire@dal.ca)

**Warning: start thinking about your general area of interest related to health data!**

# Goals of a Journal Club

- Improve understanding of prior research
- Keep up with **newest** research
- Get practice reading research
- Learn how to critically appraise research
- Get practice presenting research



2021: 1k “Health Data Science”  
10k “Data Science”  
20k “Machine Learning”  
48k Health Data Science  
101k Data Science

# What is a research article?

- Standards differ across fields and subfields (conference vs journal)
- Preprint archives
- Peer review process
- “In-Press”
- Open review
- Archaic identifiers

The image shows a screenshot of a research article page from Applied Surface Science. Red arrows point to various elements:

- Publishing house:** Elsevier logo.
- Journal name:** Applied Surface Science.
- Date year:** 2014.
- Page numbers:** 105-108.
- Journal name:** Applied Surface Science (repeated).
- Article name:** Pulsed laser deposition of  $\text{Co}_3\text{O}_4$  nanocatalysts for dye degradation and CO oxidation.
- Author surname:** Edla\*.

Other visible text includes: "Contents lists available at ScienceDirect", "journal homepage: www.elsevier.com/locate/apsusc", "CrossMark", "ARTICLE INFO", "ABSTRACT", and "Article history: Received 17 June 2013, Received in revised form 16 September 2013, Accepted 16 September 2013".

# Parts of a paper

## IMRAD format

- **Abstract** (Summary of key justification, method, results, and take-aways)

~18–68% of medical journal abstracts contain omissions or inaccuracies ([Pitkin, 1999](#)).

- **Introduction** (why the authors decided to do this research)
  - **Methods** (how they did it, and how they analysed their results)
  - **Results** (what they found)
  - **Discussion** (what the results mean).
- 
- **Conclusions** (what they want you to take away)

# Presenting a paper to the class

## Overview of the paper (30 minutes including discussion!):

- **Background:**
  - Describe rationale and importance of paper
  - Highlight the previous research that underlie this paper
- **Methodology:**
  - Describe the dataset/collection (and exclusion criteria)
  - Describe the main analysis methods they've selected and their justification
- **Results:**
  - Summarise the key results/figures
  - What didn't they detect?
- **Discussion/Conclusion:**
  - Summarise their discussion points: what limitations/contextualisation did they highlight?
  - What conclusions did they draw?

# Presenting a paper to the class

## **Critique of the paper:**

- Main question: relevant/interesting?
- Originality?
- Easy to read?
- Conclusions supported by results?
- Unaccounted for Biases?
- Missing contextualisation?

## **Conclusion:**

- Restate take-homes
- What is the broader implication of this paper?
- What follow-up experiments would this work warrant?
- How would you translate these findings into impact

Let's discuss Alston & Rick 2021



# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

- What is the general problem they identify?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

- What is the general problem they identify?

*Replication crisis*

- What is the specific problem?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

- What is the general problem they identify?

*Replication crisis*

- What is the specific problem?

*Analyses are not reproducible*

- Why do they think people should do reproducible research?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

- What is the general problem they identify?

*Replication crisis*

- What is the specific problem?

*Analyses are not reproducible*

- Why do they think people should do reproducible research?

Benefits researchers:

1. *Documentation*
2. *Easier to update analyses*
3. *Re-use*
4. *Rigour*
5. *Citations*

Benefits community:

1. *Accelerates field*
2. *Improves understanding*
3. *Finding mistakes*

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do they think it isn't done?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do they think it isn't done?

1. *Complexity*
2. *Technology change*
3. *Human error*
4. *IP*

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do they think it isn't done?

1. *Complexity*
2. *Technology change*
3. *Human error*
4. *IP*

What is their suggested solution?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do they think it isn't done?

1. *Complexity*
2. *Technology change*
3. *Human error*
4. *IP*

What is their suggested solution?

1. *Before analysis: planning/storage/version control/metadata*
2. *During analysis: comments/automation/containers*
3. *After analysis: dynamic/notebook/full release/DOI*



# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do you think they've missed?

# A Beginner's Guide to Conducting Reproducible Research

[Jesse M. Alston](#), [Jessica A. Rick](#) First published: 15 January 2021 <https://doi.org/10.1002/bes2.1801>

Why do you think they've missed?

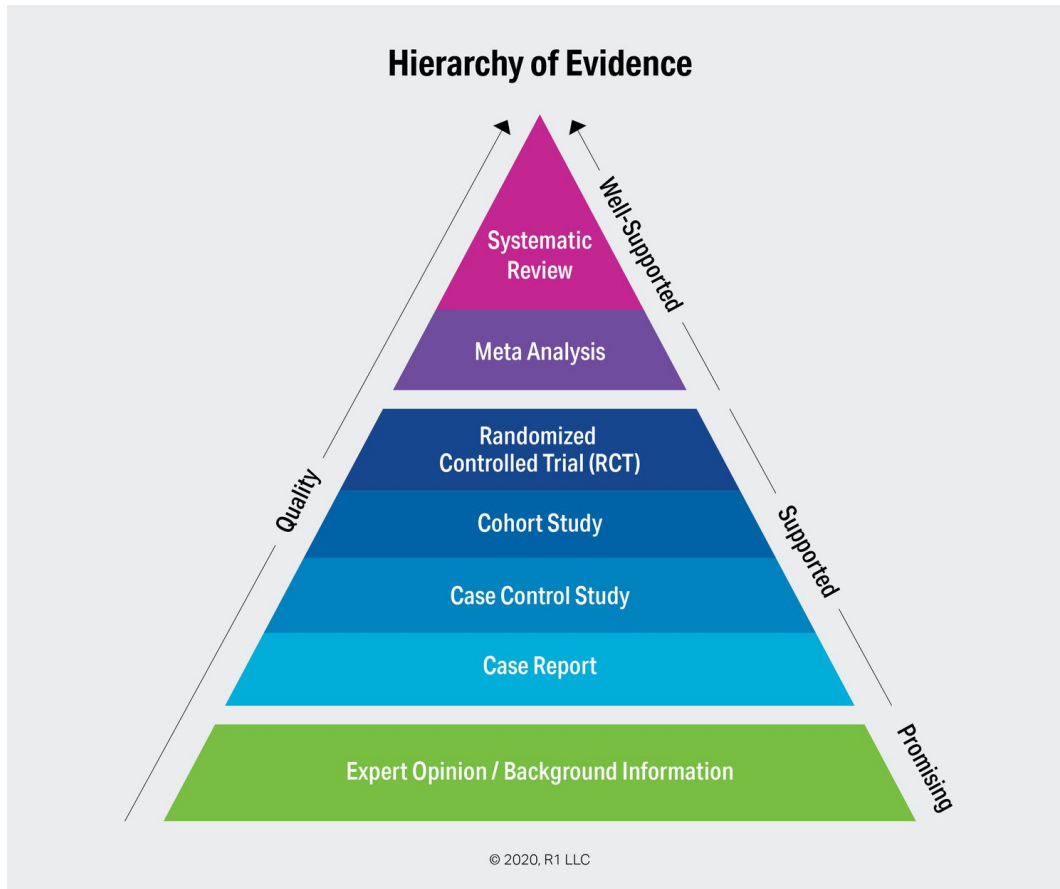
1. *Existence of technical debt suggests stronger incentivisation is needed*
2. *Data ordinality*
3. *Workflow managers (apart from Make)*
4. *Notebooks/markdown*
5. *Typos...*
6. *Gold standard example?*

# Reproducibility in machine learning for health research: Still a ways to go

[Matthew B. A. McDermott](#) [Shirly Wang](#) [Nikki Marinsek](#) [Rajesh Ranganath](#) [Luca Foschini](#) [Marzyeh Ghassemi](#)

Science Translational Medicine • 24 Mar 2021 • Vol 13, Issue 586 • [DOI: 10.1126/scitranslmed.abb1655](https://doi.org/10.1126/scitranslmed.abb1655)

# Evidence in Medicine



# Evidence in Medicine

## Research

### Artificial Intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

BMJ 2020 ; 368 doi: <https://doi.org/10.1136/bmj.m689> (Published 25 March 2020)

Cite this as: *BMJ* 2020;368:m689

#### Linked Editorial

Artificial intelligence versus clinicians


Article

Related content

Metrics

Responses

Peer review

Myura Nagendran , academic clinical fellow<sup>1</sup>, Yang Chen, academic clinical fellow<sup>2</sup>, Christopher A Lovejoy, physician<sup>3</sup>, Anthony C Gordon, professor<sup>1 4</sup>, Matthieu Komorowski, clinical lecturer<sup>5</sup>, Hugh Harvey, director<sup>6</sup>, Eric J Topol, professor<sup>7</sup>, John P A Ioannidis, professor<sup>8</sup>, Gary S Collins, professor<sup>9 10</sup>, Mahiben Maruthappu, chief executive officer<sup>3</sup>

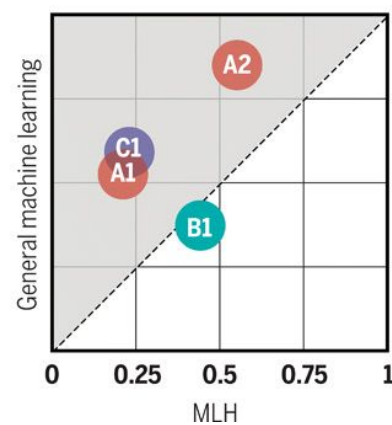
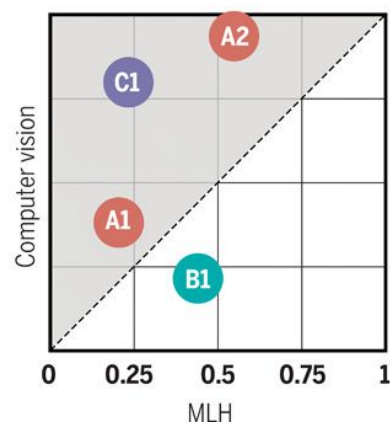
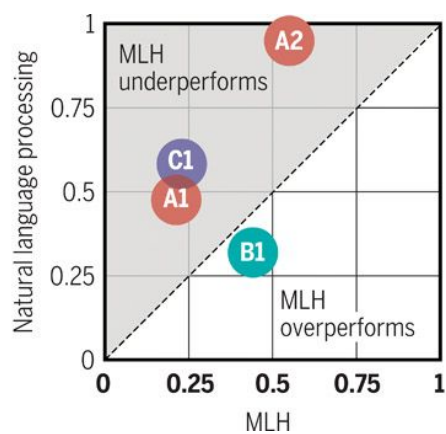
<https://r1learning.com/blog/2020/02/upon-what-evidence-are-evidence-based-practices-based-4pjtt>

Only 10 RCTs (2 published with no blinding), 81 non-randomised (6 actually tested in real clinical setting), median of 4 experts comparison but 61/81 stated comparable to human performance

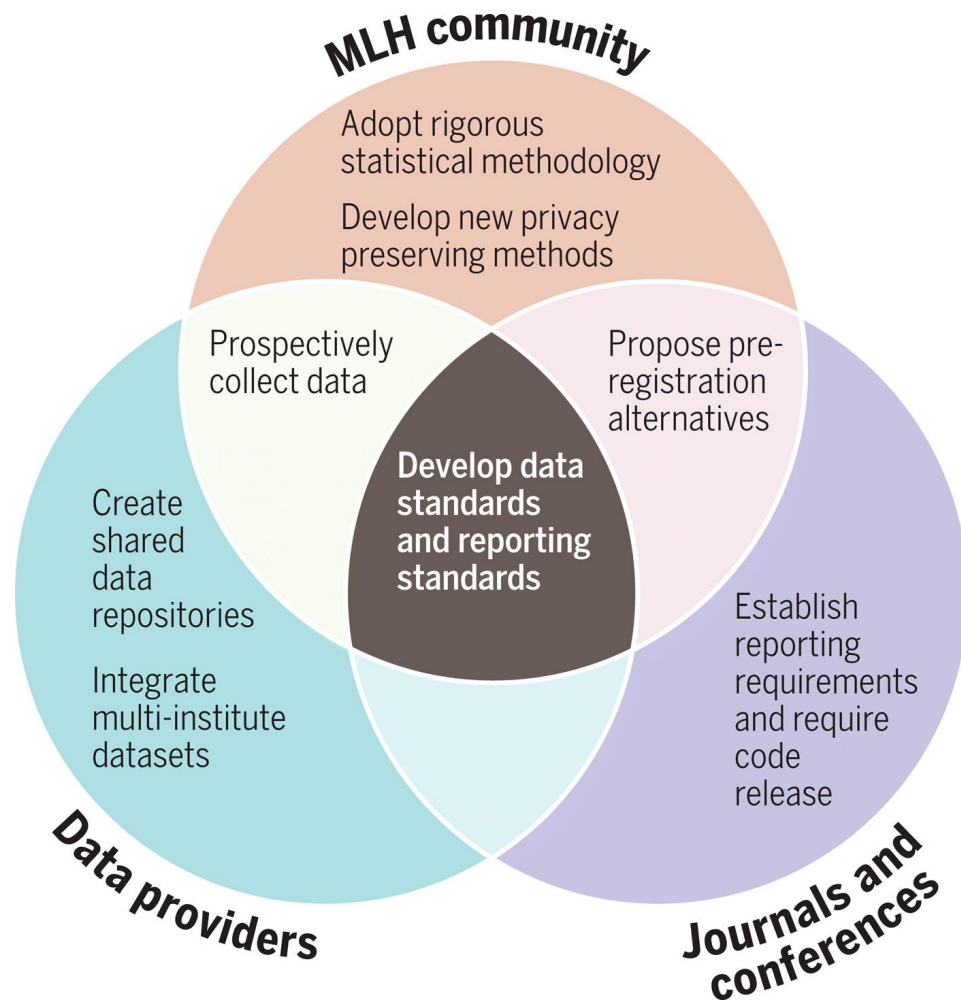
# Figure 1

## Evaluation metrics

- A** Technical reproducibility
  - 1 Code available
  - 2 Public dataset
- B** Statistical reproducibility
  - 1 Variance reported
- C** Conceptual reproducibility (replicability)
  - 1 Multiple datasets



# Figure 2



# Proposal Class

Developing a question



# Learning objectives

1. Understand the components of a research proposal
2. List the basic criteria in selection a health data science research question
3. Describe methods for developing a research question
4. List potential sources for research questions

What goes into a research proposal?

# Components of a written research proposal

- **Research Question:** *clearly defined research question related to solving an important problem*

# Components of a written research proposal

- **Research Question:** *clearly defined research question related to solving an important problem*
- **Abstract:** *concise & informative expert overview*
- **Lay Summary:** *clear general public summary of problem, solution, and relevance*
- **Introduction:** *problem/knowledge gap justification/explanation of relevant methods*
- **Literature Review:** *critical appraisal of broad relevant literature that supports method and question*
- **Methodology:** *appropriate method, data gathering/access, that solves the research question and is justified by literature review*
- **Budget:** *reasonable/appropriate timeline and cost estimates*
- **Ethics:** *explores hurdles/risks/benefits and impact of question, method and KT*
- **Discussion:** *addresses limitations, implications, and future directions/extensions.*
- **Knowledge Translation:** *robust/impactful plan to mobile results across a range of settings.*

# Components of a written research proposal

- **Research Question:** *clearly defined research question related to solving an important problem*
- **Abstract:** *concise & informative expert overview*
- **Lay Summary:** *clear general public summary of problem, solution, and relevance*
- **Introduction:** *problem/knowledge gap justification/explanation of relevant methods*
- **Literature Review:** *critical appraisal of broad relevant literature that supports method and question*
- **Methodology:** *appropriate method, data gathering/access, that solves the research question and is justified by literature review* May 17th
- **Budget:** *reasonable/appropriate timeline and cost estimates*
- **Ethics:** *explores hurdles/risks/benefits and impact of question, method and KT* May 26th
- **Discussion:** *addresses limitations, implications, and future directions/extensions.*
- **Knowledge Translation:** *robust/impactful plan to mobile results across a range of settings.* June 9th

June 2nd: Intermediate Check-in and Troubleshooting!

# Proposal Presentation Structure (June 14th)

## 20 minutes + 10 minutes Q&A

- Title Slide (1)
- Team Background and Conflicts of Interest (1)
- Background/Literature Review (3)
- Research Objectives/Question/Hypothesis (1)
- Methodology (3)
- Budgeting (1)
- Knowledge Translation Plan (1)
- Future Work (1)
- Q&A
  
- **Chance for feedback -> incorporate into final submission**

What makes a good research question?

# Components of a good research question

- **Focused:** single problem or issue
- **Novel:** hasn't already been done
- **Answerable:** ideally quantitatively
- **Feasible:** to answer within the timeframe and practical constraints
- **Specific:** can be thoroughly addressed
- **Interesting:** to you & your collaborators
- **Relevant:** implications for broader field/society (KT)

Additional:

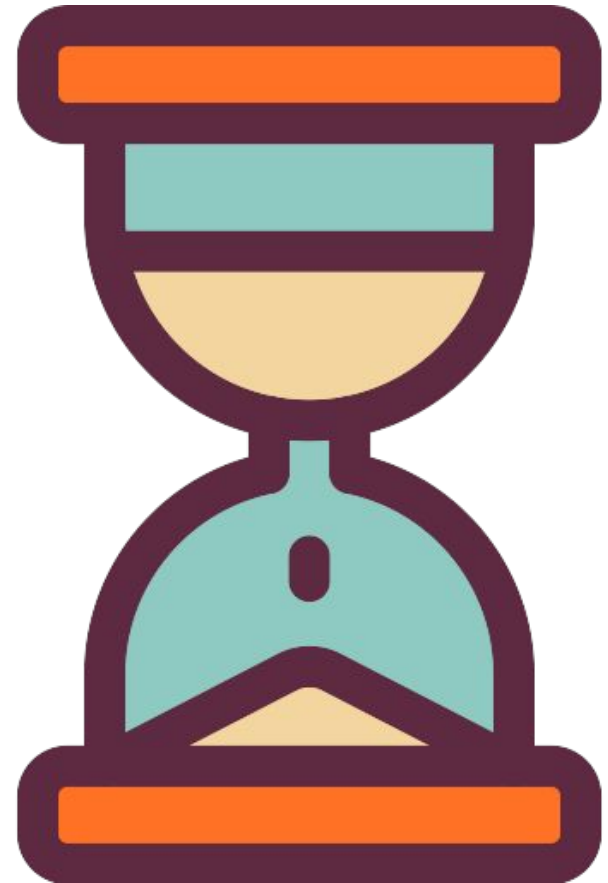
- **(?)Complex:** not too trivial





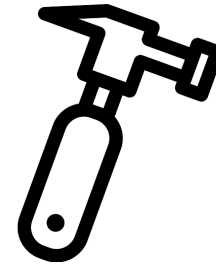
# The Research “Hourglass”

- General question (broad)
- Specific question (narrow)
- Data (collection/curation/gathering)
- Analyse Data
- Contextualise Results (discuss limitations/differences in outcome to other studies)
- Generalise Conclusions (broader relevance)



# 3 ways to identify a general area

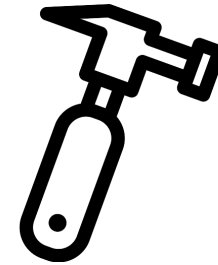
- **Method**
  - new algorithm, new statistical approach, new ML workflow



# 3 ways to identify a general area

- **Method**

- new algorithm, new statistical approach, new ML workflow



- **Data**

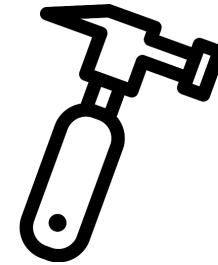
- new/expanded/linked dataset e.g., controlled workplace proximity infectivity data



# 3 ways to identify a general area

- **Method**

- new algorithm, new statistical approach, new ML workflow



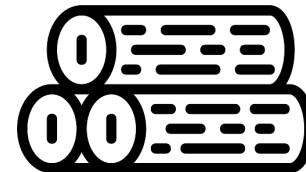
- **Data**

- new/expanded/linked dataset e.g., controlled workplace proximity infectivity data



- **Problem**

- Emergent disease/health/social crisis, long existing/unresolved crisis



# Formulating a research question

Research aims	Research question formulations
Describing and exploring	<ul style="list-style-type: none"><li>• What are the characteristics of <i>X</i>?</li><li>• How has <i>X</i> changed over time?</li><li>• What are the main factors in <i>X</i>?</li><li>• How does <i>X</i> experience <i>Y</i>?</li><li>• How has <i>X</i> dealt with <i>Y</i>?</li></ul>
Explaining and testing	<ul style="list-style-type: none"><li>• What is the relationship between <i>X</i> and <i>Y</i>?</li><li>• What is the role of <i>X</i> in <i>Y</i>?</li><li>• What is the impact of <i>X</i> on <i>Y</i>?</li><li>• How does <i>X</i> influence <i>Y</i>?</li><li>• What are the causes of <i>X</i>?</li></ul>
Evaluating and acting	<ul style="list-style-type: none"><li>• What are the advantages and disadvantages of <i>X</i>?</li><li>• How effective is <i>X</i>?</li><li>• How can <i>X</i> be achieved?</li><li>• What are the most effective strategies to improve <i>X</i>?</li><li>• How can <i>X</i> be used in <i>Y</i>?</li></ul>

**Let's brainstorm some research ideas!**