

Case Studies in Data Science – Week 6

Damiano Spina

Johanne Trippas



Acknowledgement of Country

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

RMIT University respectfully acknowledges their Ancestors and Elders, past and present.

RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.

Artwork 'Sentient' by Hollie Johnson

Hollie is a Gunaikurnai and Monero Ngarigo woman from Gippsland who graduated from RMIT with a BA in Photography in 2016.



Acknowledgement of Country

1. Name Country
2. Acknowledge Ancestors and Elders
3. Acknowledge Aboriginal and Torres Strait Islander peoples in the room

[Optional: positionality/personalise it/share a story...]





Upcoming Assignments

Week 6 reflective portfolio due tomorrow

Mid-semester Break!

Week 8: Individual Task 2 due on Sep 17

Week 8 reflective portfolio



Individual Task 1





Individual Task 1: Most common feedback

- Use **in-text citations** to support your statements and introduce concepts.
- Task: clearly describe what the problem you are trying to address is.
- Lack of evidence of experimentation
- Always include informative captions in figures and tables (vector format, no screenshots from notebooks/terminal).
- Choice of evaluation metrics: informed by the requirements/properties of your problem (and the resources you have to address it, e.g., class imbalance).
- References: consider the authoritativeness of the sources (e.g., blog vs. peer-reviewed publication)
- Statement about the use (or not use) of Generative AI is **mandatory**.
- Video: use results and literature to support your statements
- Pitch: why -> what -> [how] -> so what?






In-text Citations

- Indicate the relevant reference in the text/content.
- In LaTeX, use `\cite{bibkey}` or `\citet{bibkey}` to do so. In ACM format, each reference number should be enclosed in square brackets on the same line as the text, before any punctuation, with a space before the bracket [1].
- Use a single `\cite` or `\citet` command for multiple references in the same sentence (e.g., `\cite{spina2025responsible, trippas2025adapting}`).
- You should only include references that are directly cited in your content/body.

[1] RMIT University. *Easy Cite referencing guide – IEEE – General rules for in-text citations*. [online] URL: <https://www.lib.rmit.edu.au/easy-cite/?styleguide=styleguide-5#stn-0#subtype-1> (Accessed: 26 Aug 2025)



Individual Task 2





Individual Task 2 – Week 8

Part 1: Executive Summary of a Position Paper

Part 2: Deliberation on Task 1 (ethics + bias)

Part 3: Peer-review of video in Task 1

Submission (all parts): 1 PDF (4 pages + 1 page for refs.)

Overleaf: Adapt the template accordingly!

Remember 48-hour silent policy!

Engage on Canvas discussion forum to ask your questions in time. Requests for extensions can only be processed before the silent policy starts.





Part 1: Executive Summary

- We expect you engage with related literature
- Your list of references should include the paper you are discussing (of course), but also others that you consulted to critically engage in the discussion
- At least 5-10 references in addition to the paper you chose
- Hint: quickly skim all the options and choose the position paper that is of most interest for you (work on the topic that it excite you the most!)





Part 2: Deliberation of Task 1

- Remember Arwen's talk in Week 1? 😊
- Read carefully the description of the activities you need to carry out
- Most likely, it will consulting additional resources (not just ChatGPT) to become familiar with concepts
- Fairness-aware analysis: again, explore the options before you commit to one
- Be specific: if you use GenAI to discuss security, privacy, and ethical risks, it's unlikely you will meet the expectations without a critical discussion that goes beyond trivial/generic answers





Part 3

- Read the instructions carefully!
- The review you include in your submitted PDF is the one that is assessed
- OPTIONAL: You can opt to share that review with the author of the video. But we won't check/assess this
- If you only submit the review on Canvas, but you don't include the review in your PDF submission, you don't get marks.





Weeks 7-9

Our colleagues Falk Scholer, Chris John, and Danula Hettiachchi will be hosting our speakers



Weeks 7-9

Our colleagues Falk Scholer, Chris John, and Danula Hettiachchi will be hosting our speakers



CLEF 2025 Lab Co-Chair
EXIST Lab at CLEF 2025
(Madrid, Spain)
<https://clef2025.clef-initiative.eu/>



Weeks 7-9

Our colleagues Falk Scholer, Chris John, and Danula Hettiachchi will be hosting our speakers



CLEF 2025 Lab Co-Chair EXIST Lab at CLEF 2025

9.7 IR With Low Resource Languages

IR research traditionally focuses on languages that are widely spoken, have well-documented linguistic resources, and hold strong societal representation. In contrast, low-resource languages, which have limited data, processing tools, or institutional support, are often overlooked in IR research. This imbalance leads to barriers to information access for these languages.

Conducting IR research with low-resource languages presents numerous challenges. These include limited digital and linguistic resources, the necessity of collaborating with subject-matter experts or protected groups such as Indigenous communities in remote areas, or a general lack of computational resources and digital infrastructure for users. Together, these barriers create a cycle where limited resources hinder research progress, reducing the visibility and usability of these languages within information systems. Despite these obstacles, IR has the potential to break this cycle by improving access, visibility and inclusion for low-resource languages in the digital sphere. Initiatives such as CLEF, NTCIR, and FIRE have demonstrated that the IR community can broaden its scope beyond TREC's historical focus on English. These efforts have expanded research to include European, Asian, Indian, and African languages, but most of the world's spoken languages remain unrepresented in IR research. This is often due to limited involvement from speaker communities in shaping IR systems and benchmarks. Rather than unintentionally contributing to language erosion, IR research should actively support the safeguarding and revitalising of low-resource languages. This can be achieved by pursuing new initiatives, such as the Low Resource Environments Track at ACM SIGIR 2025, and embracing participatory research alongside co-design practices with language groups who benefit from our expertise.

**Visit to Prof. Josiane Mothe
(Toulouse, France)**



Weeks 7-9

Our colleagues Falk Scholer, Chris John, and Danula Hettiachchi will be hosting our speakers



CLEF 2025 Lab Co-Chair EXIST Lab at CLEF 2025

9.7 IR With Low Resource Languages

IR research traditionally focuses on languages that are widely spoken, have well-documented linguistic resources, and hold strong societal representation. In contrast, low-resource languages, which have limited data, processing tools, or institutional support, are often overlooked in IR research. This imbalance leads to barriers to information access for these languages.

Conducting IR research with low-resource languages presents numerous challenges. These include limited digital and linguistic resources, the necessity of collaborating with subject-matter experts or protected groups such as Indigenous communities in remote areas, or a general lack of computational resources and digital infrastructure for users. Together, these barriers create a cycle where limited resources hinder research progress, reducing the visibility and usability of



Dagstuhl, Germany

Visit to Prof. Josiane Mothe (Toulouse, France)

Dagstuhl Seminar 25391

Retrieval-Augmented Generation – The Future of Search?

(Sep 21 – Sep 26, 2025)

Please use the following short url to reference this page: <https://www.dagstuhl.de/25391>

- Matthias Hagen (Friedrich-Schiller-Universität Jena, DE)
- Josiane Mothe (Toulouse University, FR)
- Smaranda Muresan (Barnard College, Columbia University - New York, US)
- Martin Potthast (Universität Kassel, DE)
- Min Zhang (Tsinghua University - Beijing, CN)

Enjoy Mid-Semester Break!

