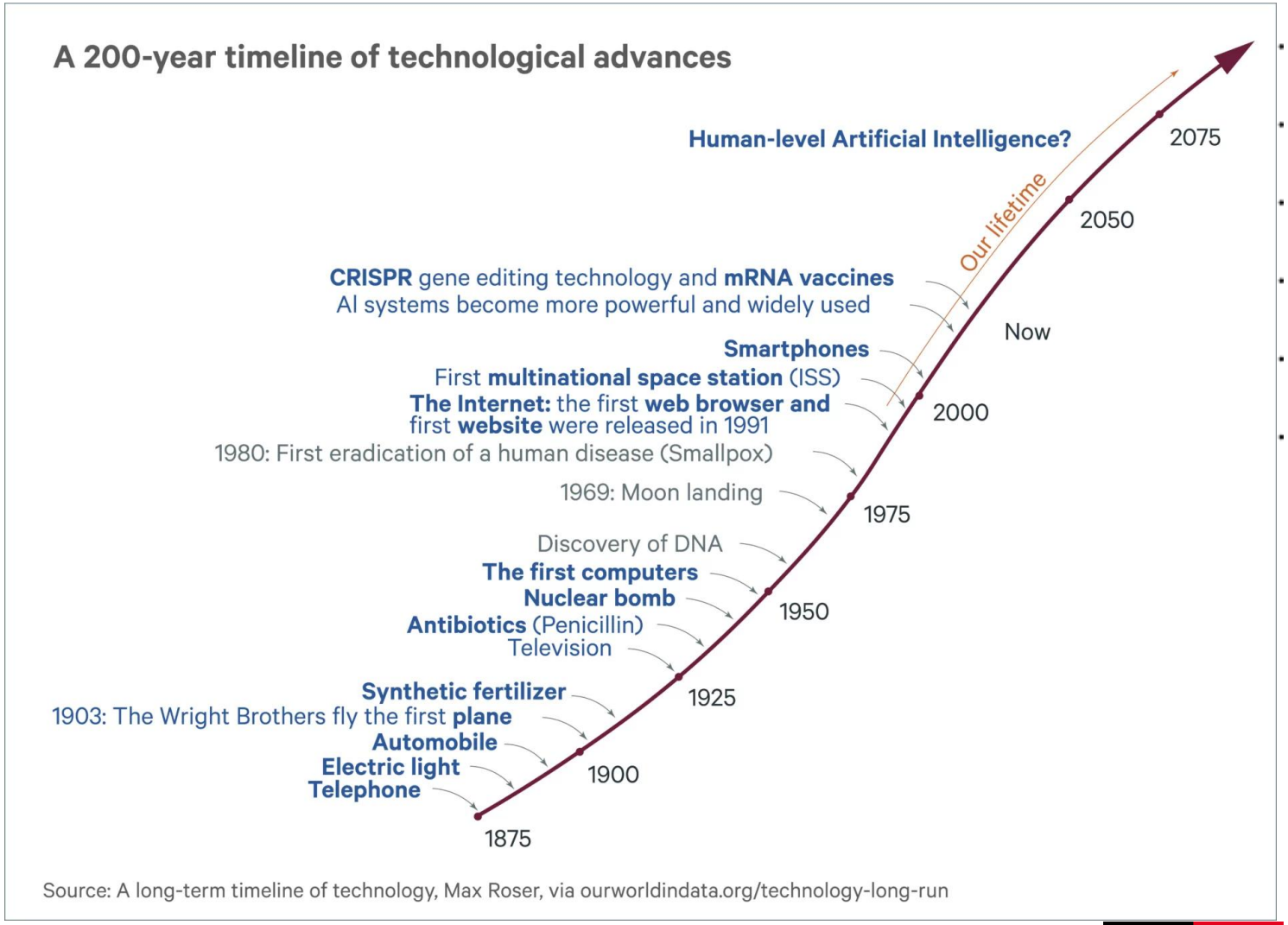
AI-Powered Analysis of ORCID and RLA Data for Technology Landscaping

Presenter: Luhan Cheng

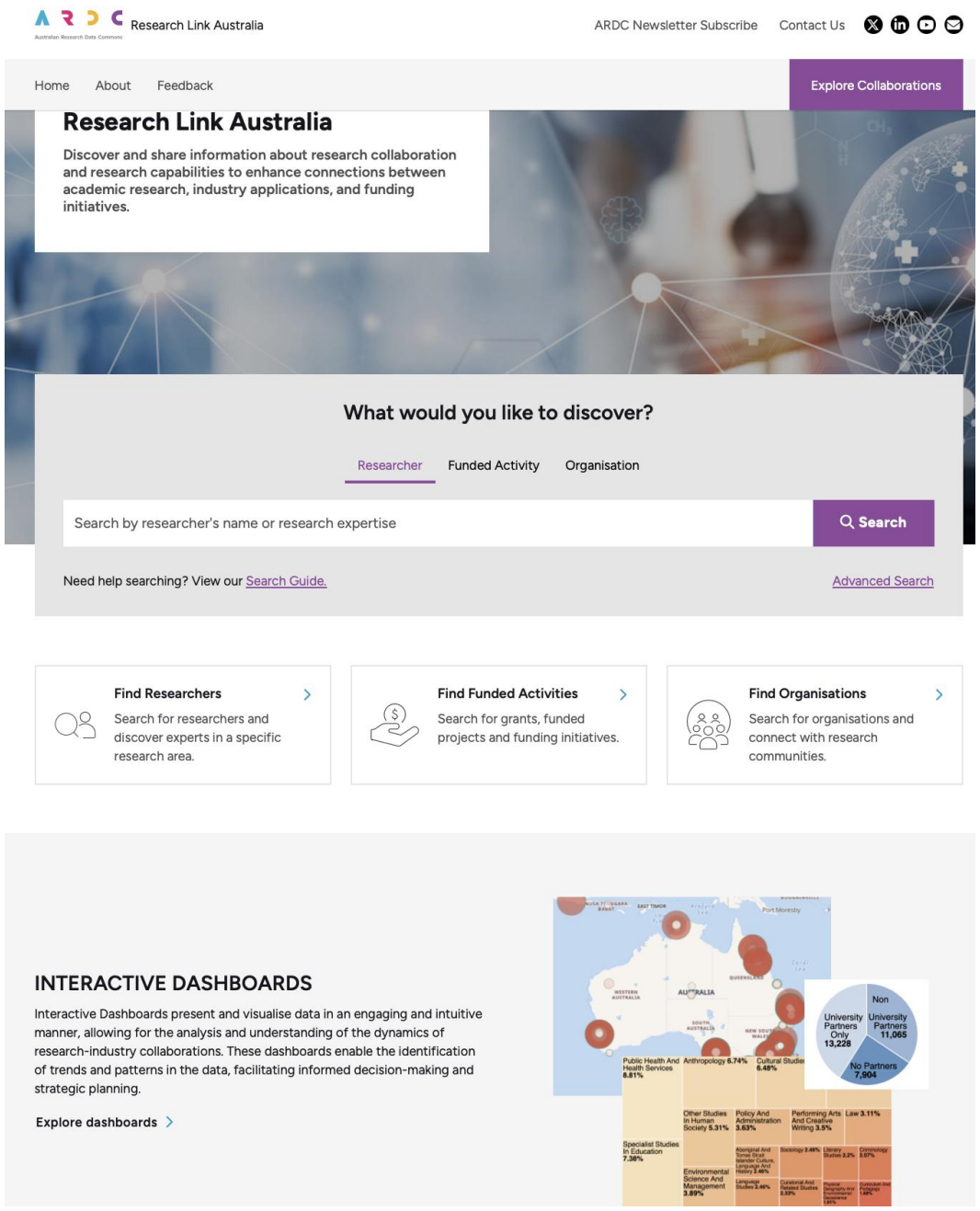
Technology Landscaping

- Using AI to identify key technologies strategically relevant to organisations, research projects and industry.
- Understand the innovation trends, opportunities and risks
- Make rational decisions on how to allocate resources



https://ourworldindata.org/technology-long-run

SWINBURNE UNIVERSITY OF TECHNOLOGY



Connected network of *Researchers*, *Projects* and *Organisations*

Provided by Australian Research Data Commons

805,500+

RESEARCHERS
(ORCID)

1,800,000+

LINKED PUBLICATIONS
(CROSSREF, PUBMED, ORCID)

88,000+

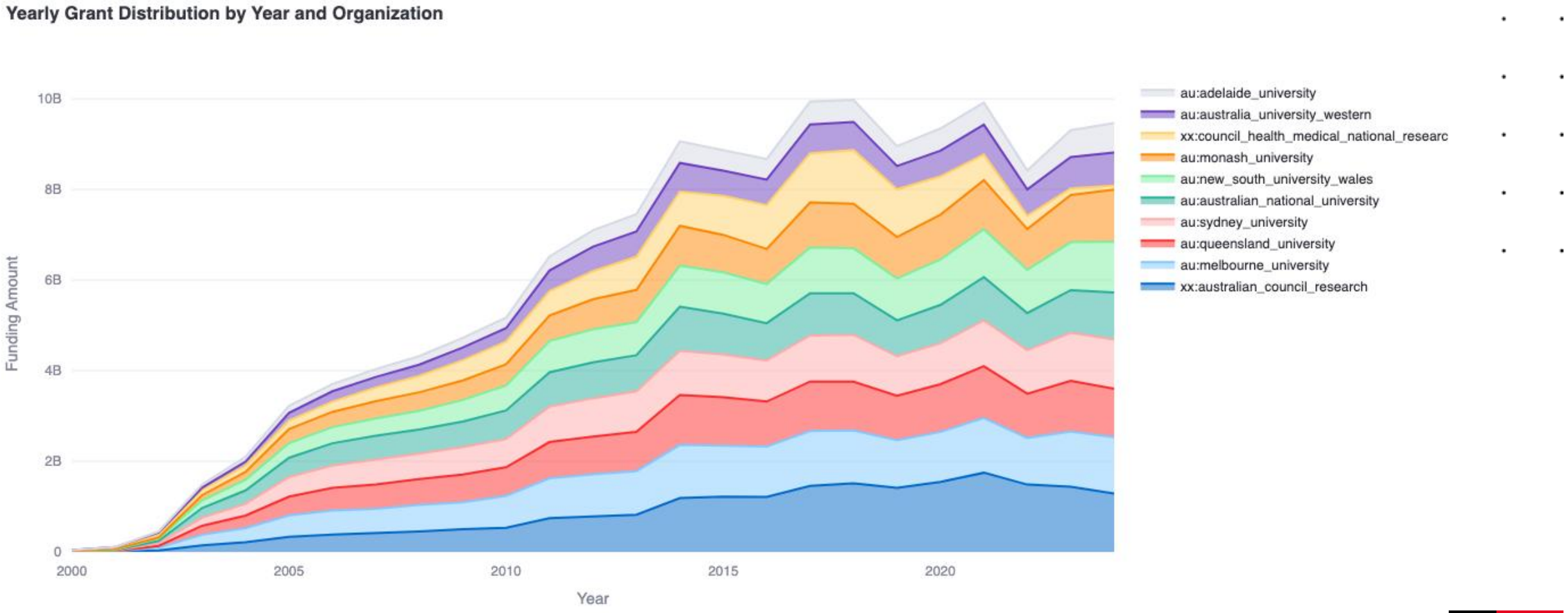
FUNDED ACTIVITIES
(ARC, NHMRC, ORCID)

310,000+

ORGANISATIONS
(CROSSREF, ROR, ABR)

SWINBURNE UNIVERSITY OF TECHNOLOGY

The value of ORCID data



SWINBURNE UNIVERSITY OF TECHNOLOGY

Sample Grants

Table 1

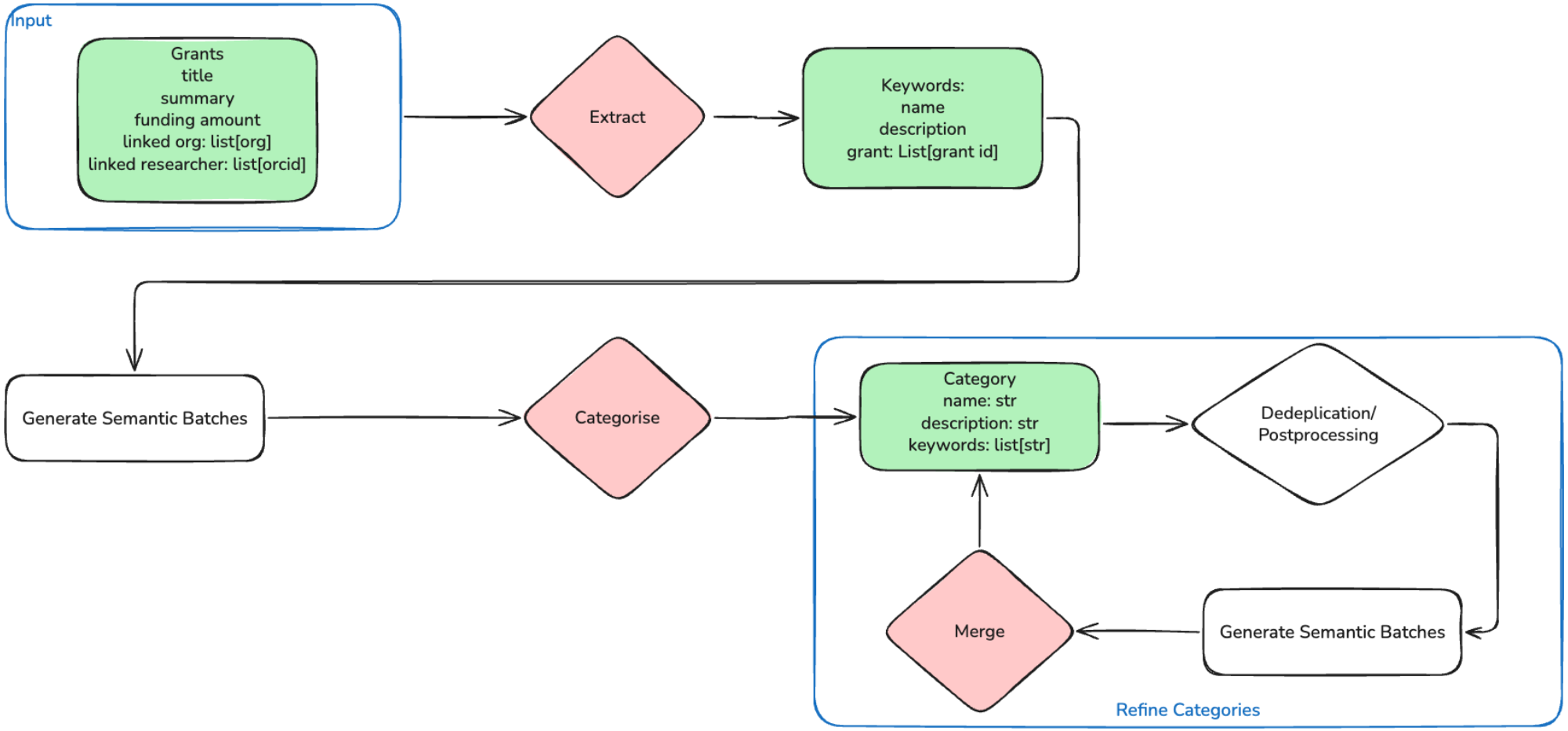
Title	ARC Training Centre for Information Resilience.
Grant_Summary	ARC Training Centre for Information Resilience. The proposed centre aims at building workforce capacity in Australian organisations to create, protect and sustain agile data pipelines, capable of detecting and responding to failures and risks across the information value chain in which the data is sourced, shared, transformed, analysed and consumed...
Funding_amount	4883406.0
Start_year	2021
End_year	2026
Funder	Australian Research Council
Linked_researchers	['0000-0002-5414-8276', '0000-0002-4259-9774', '0000-0002-6920-7072', '0000-0002-7311-3693',...]

SWINBURNE UNIVERSITY OF TECHNOLOGY

Why extract keywords and build categories using LLM

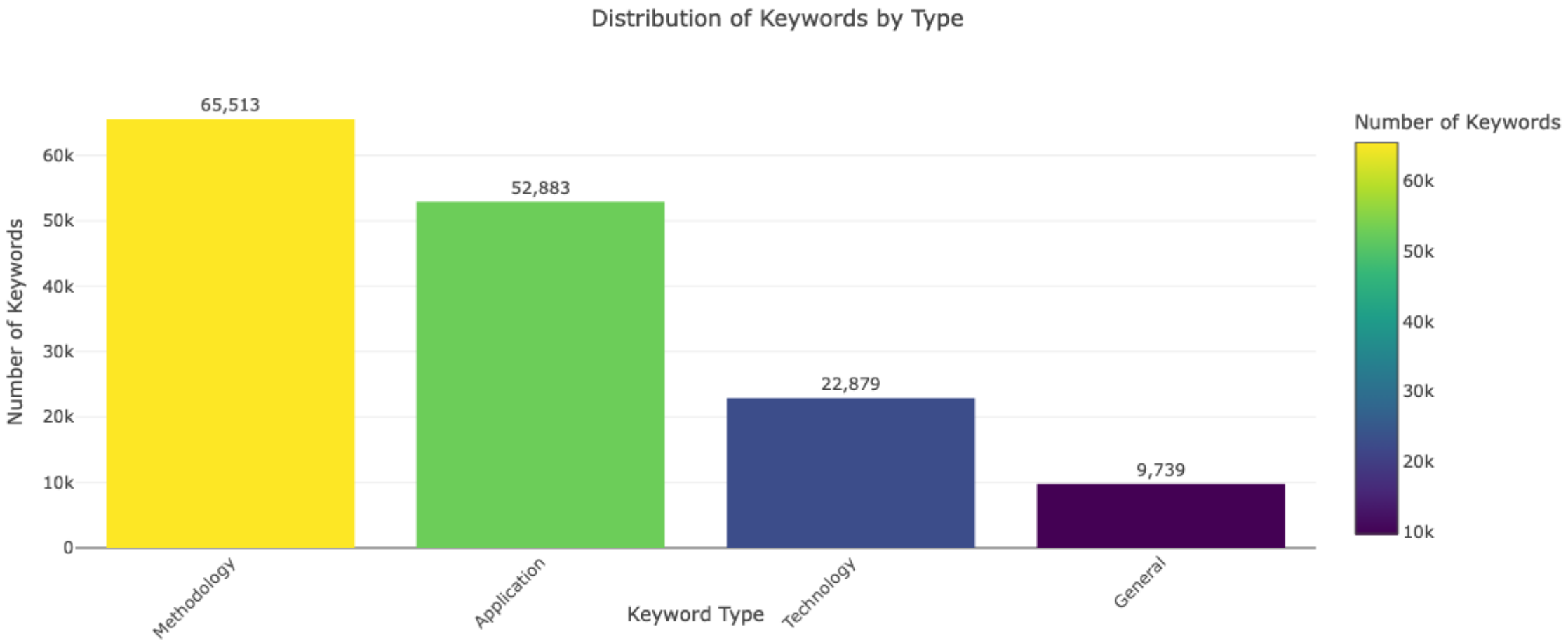
- Existing categories does not capture emerging trends
- Existing categories do not update frequently (10+ year time interval for ANZSRC)
- LLM can approximate domain expert knowledge while being much easier to scale up

Overview of the workflow



SWINBURNE UNIVERSITY OF TECHNOLOGY

What are the keywords detected by our methods

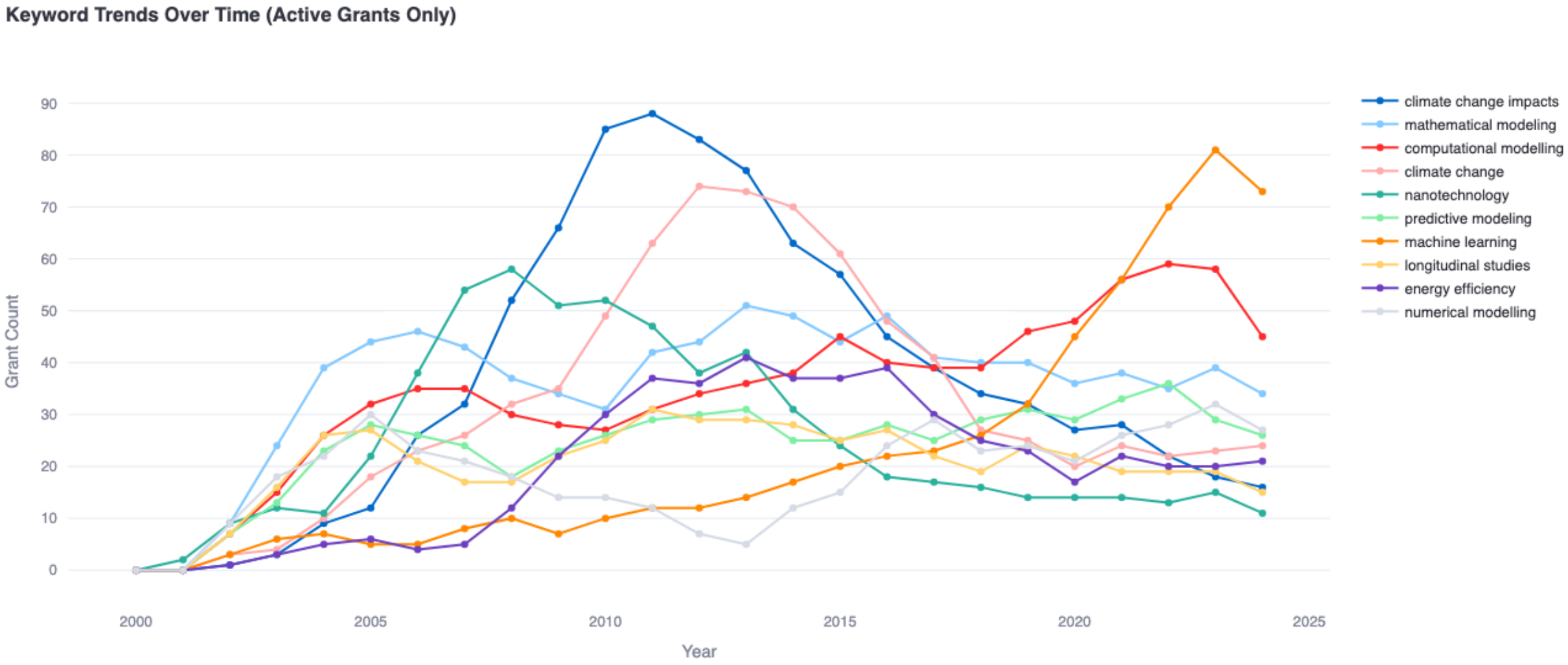


Sample Keywords

Name	laser interferometry
Type	Methodology
Description	A technique that uses laser beams and interferometers to measure extremely small changes in distance, enabling the detection of gravitational wave signals.
Grants	['arc/FT100100613', 'arc/LE0989093', 'arc/DP110103347']
Organisation_name	['Australian National University', 'University of Otago', 'Australian National University', 'University of Western Australia',...]
Organisation_country	['AU', 'NZ', 'AU', 'AU',....]
Researcher_orcid	['0000-0001-6210-5842', '0000-0001-6210-5842', '0000-0001-6210-5842', '0000-0001-6210-5842',...]

SWINBURNE UNIVERSITY OF TECHNOLOGY

Keywords Trends

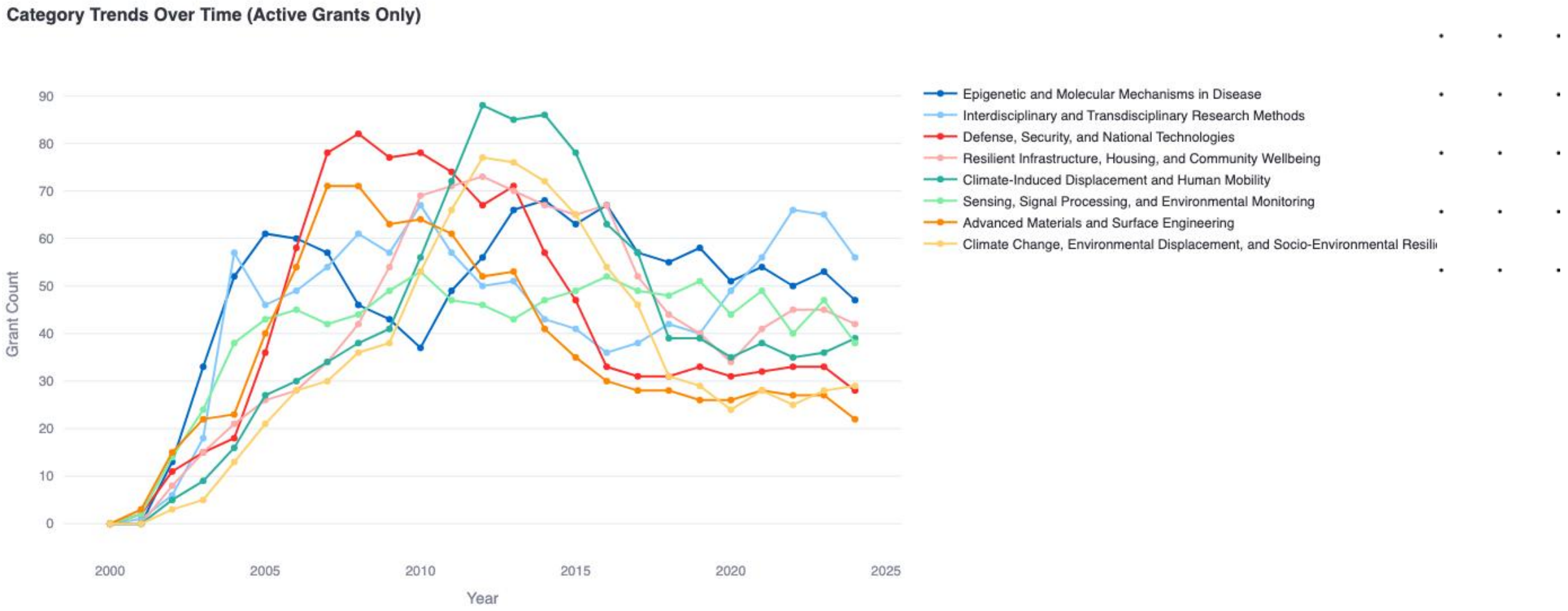


Category Sample

name	description	keywords
Climate and Paleoenvironmental Reconstruction	A unified domain integrating climate variability, paleoclimate reconstructions, and ecological responses to assess long-term environmental change. This includes studies of past climate events (e.g., Eocene-Oligocene transition, Pleistocene climatic events), sea-level changes, atmospheric circulation, and climate-reanalysis using geological, biological, and historical proxies such as coral records, pollen, tree rings, and sediment cores…..	["Australia's flooding history", 'Australian climate history', 'Cenozoic landscape evolution', 'Eocene-Oligocene transition', 'Pliocene warm period', 'Quaternary climate modelling', 'Quaternary glacial-interglacial transitions', 'South Polar greenhouse', 'abrupt warming', 'ancient air bubbles', 'ancient analogue', …]
Evolutionary and Paleobiological Foundations	This category explores the deep evolutionary history of life, particularly in Australia, including early animal evolution, Ediacaran origins, and the Cambrian explosion. It integrates fossil records, faunal succession, and evolutionary transitions to understand the origins of animal phyla and the biogeographic context of Gondwanan life. This research provides foundational insights into the evolutionary pathways of modern biodiversity and the development of complex life forms in prehistoric Earth.	["Australia's oldest complex organisms", 'Cambrian Explosion', 'Cambrian fossils', 'Cambrian radiation', 'Cambrian records', 'Ediacaran faunal succession', 'Ediacaran origins', 'Ediacaran roots', 'Ediacaran–Cambrian periods', 'Precambrian fossils', 'diversification events', 'early Cambrian fossils', 'evolutionary thought', 'fossil record', 'vertebrate fauna']

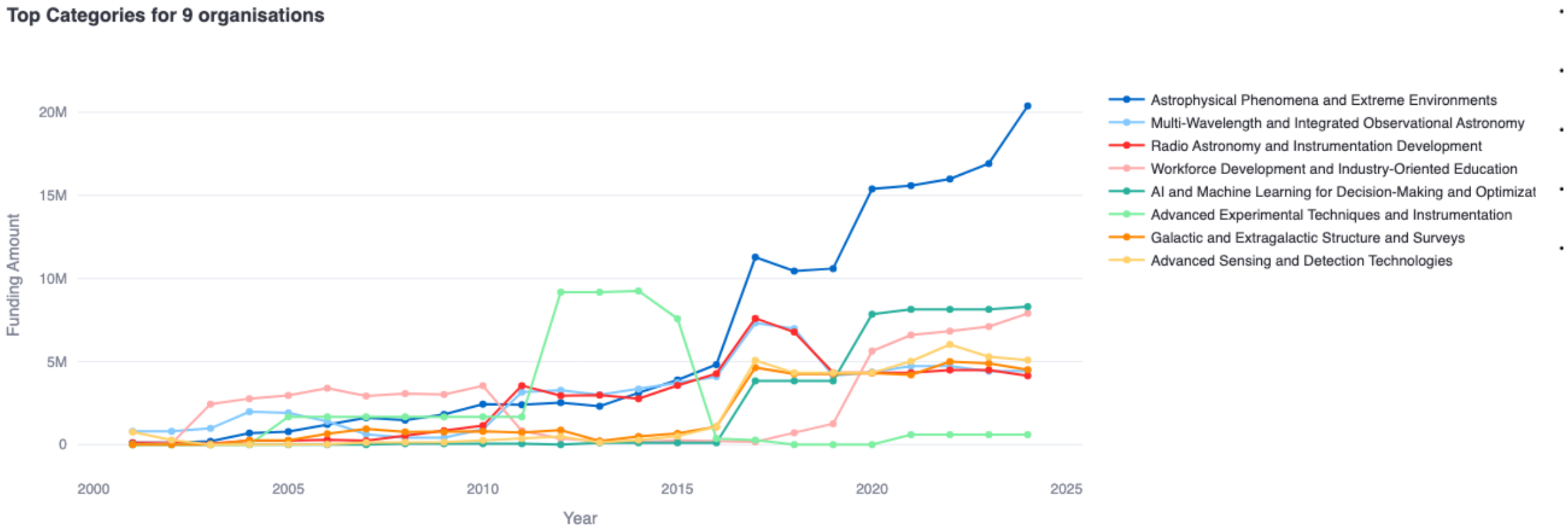
SWINBURNE UNIVERSITY OF TECHNOLOGY

Category Overview

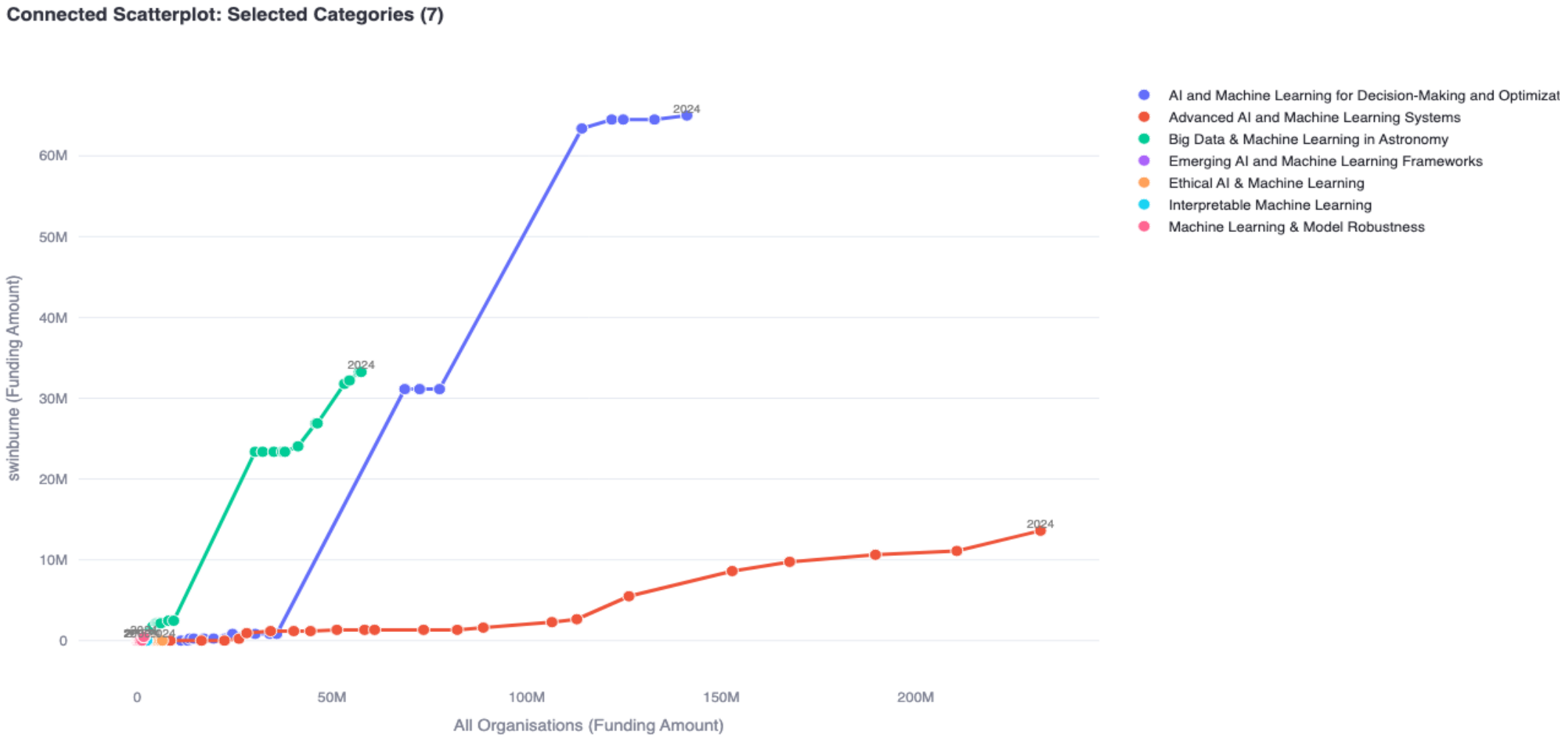


SWINBURNE UNIVERSITY OF TECHNOLOGY

Swinburne


Top Categories for 9 organisations
Funding Amount
20M
15M
10M
5M
0
2000
2005
2010
2015
2020
2025
Year
Astrophysical Phenomena and Extreme Environments
Multi-Wavelength and Integrated Observational Astronomy
Radio Astronomy and Instrumentation Development
Workforce Development and Industry-Oriented Education
AI and Machine Learning for Decision-Making and Optimizat
Advanced Experimental Techniques and Instrumentation
Galactic and Extragalactic Structure and Surveys
Advanced Sensing and Detection Technologies

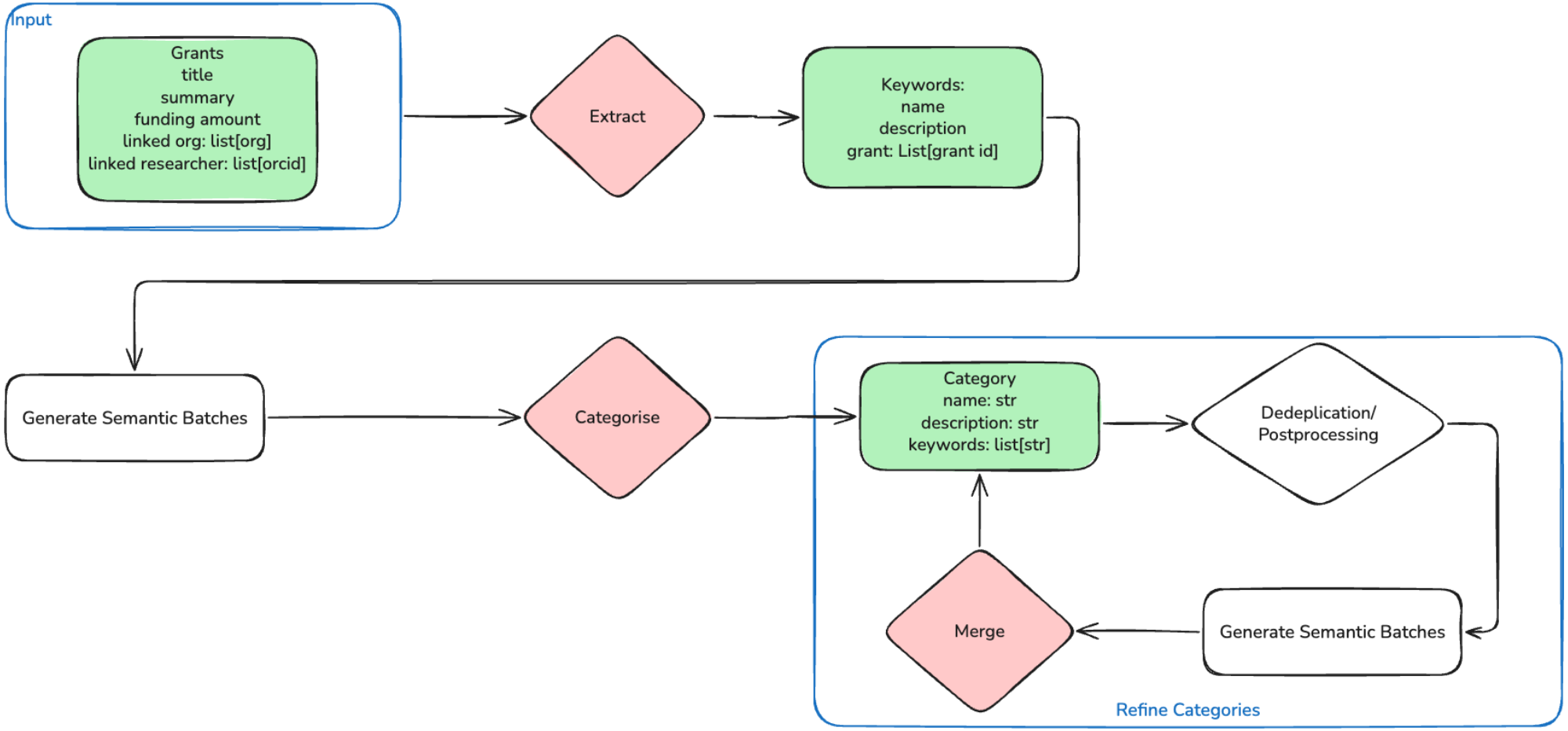
Compare Swinburne with rest of Australia in machine learning



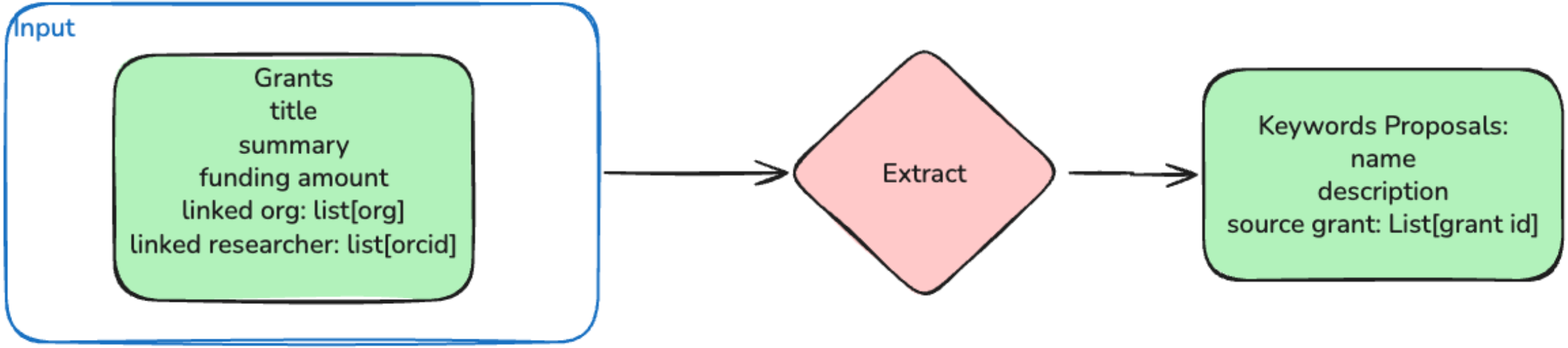
Methodology

SWINBURNE UNIVERSITY OF TECHNOLOGY

Overview of the workflow



Keywords Extraction



Keywords Extraction

System prompt:

You are an expert research analyst with deep knowledge across multiple academic disciplines and a keen eye for emerging research trends.

Your task is to extract meaningful keywords from research grant information that would be useful for:
- Identifying emerging research domains and interdisciplinary areas
- Discovering novel methodologies and cutting-edge approaches
...

User input:

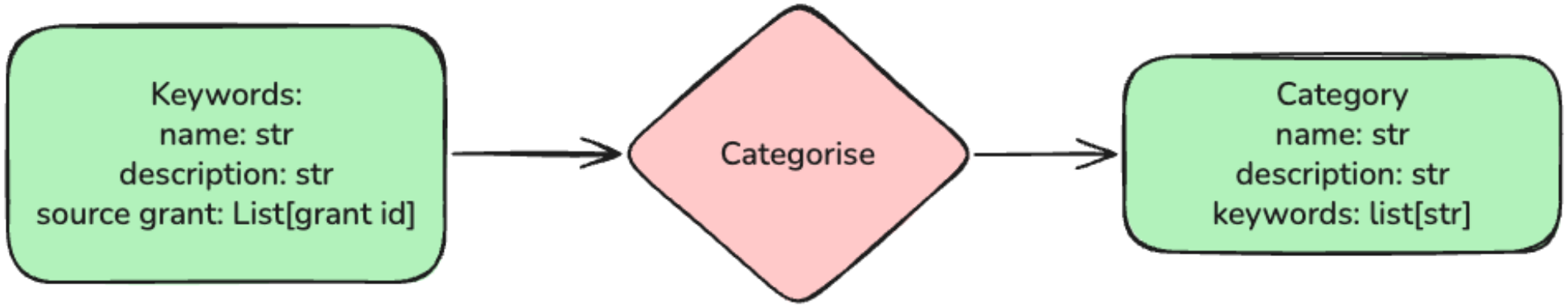
Title: Investigation and Improvement of Live Blue Swimmer Crab Handling in NSW
Description: As the NSW fishing industry moves forward into the future with quota managed fisheries, there is a need for industry to develop procedures...

Output:

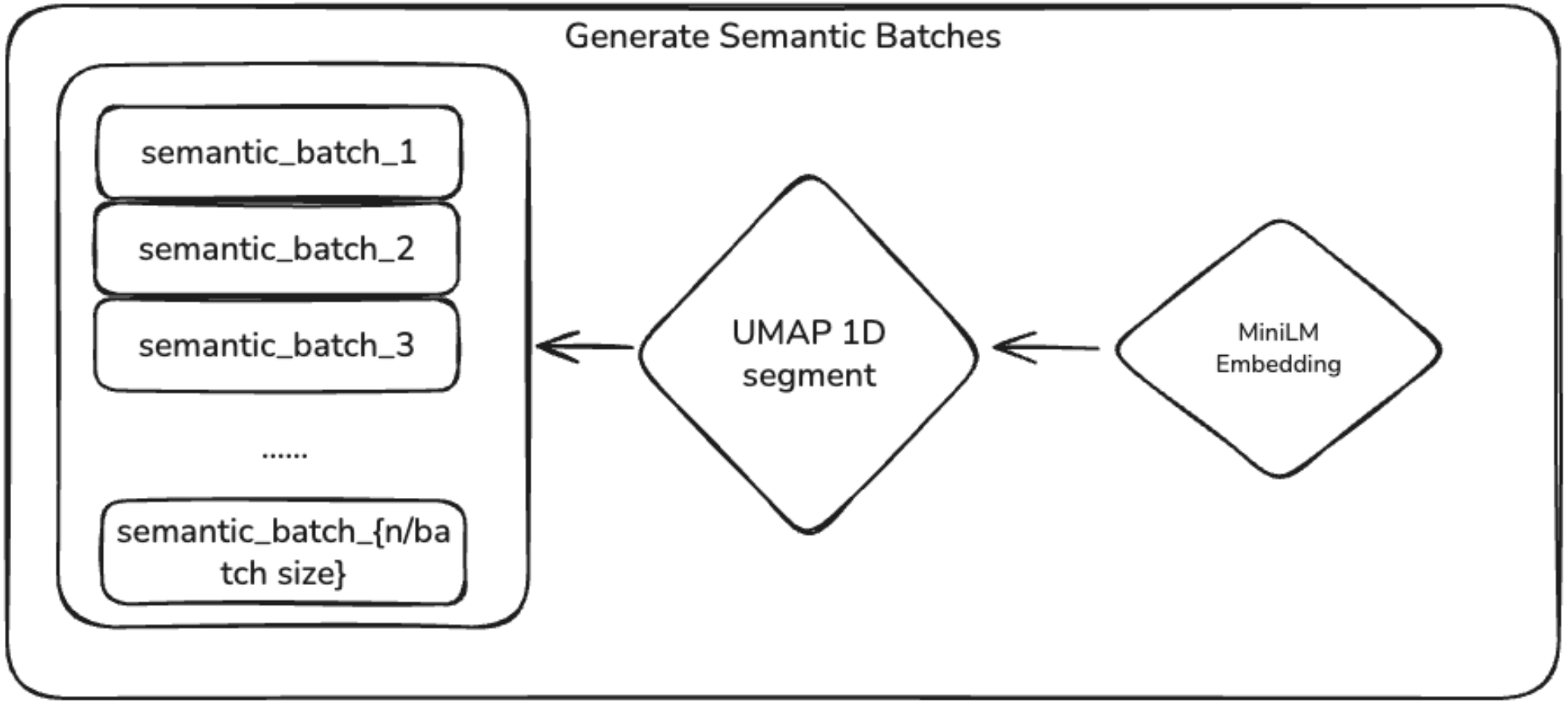
```
{   "keywords": [
      {"name": "quota managed fisheries", "description": ..., "type": "methodology"},
      {"name": "value addition to wild harvest", "description": ..., "type": "application"},
      ...
   ]
}
```

SWINBURNE UNIVERSITY OF TECHNOLOGY

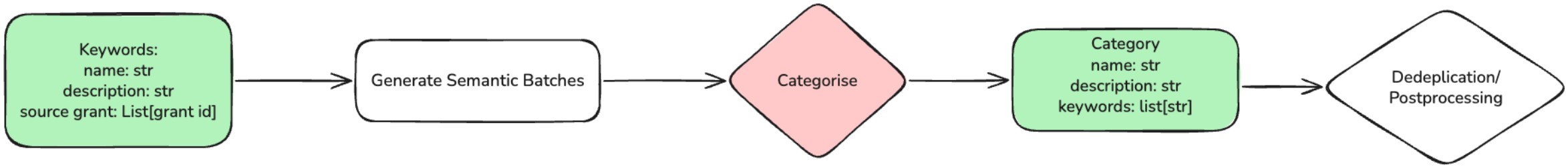
Categorisation



Semantic Batching



Categorisation



SWINBURNE UNIVERSITY OF TECHNOLOGY

Categorisation

System prompt:

You are an expert research analyst evaluating the quality of extracted keywords from a research grant.

Original Grant Input:
{input_text}

Extracted Keywords:
{keywords_text}

CRITICAL REQUIREMENT - Keywords Must Exist in Grant Text:
All keywords MUST be directly derivable from or explicitly mentioned in the original grant input text above (including both the grant title and description).
...

User input:

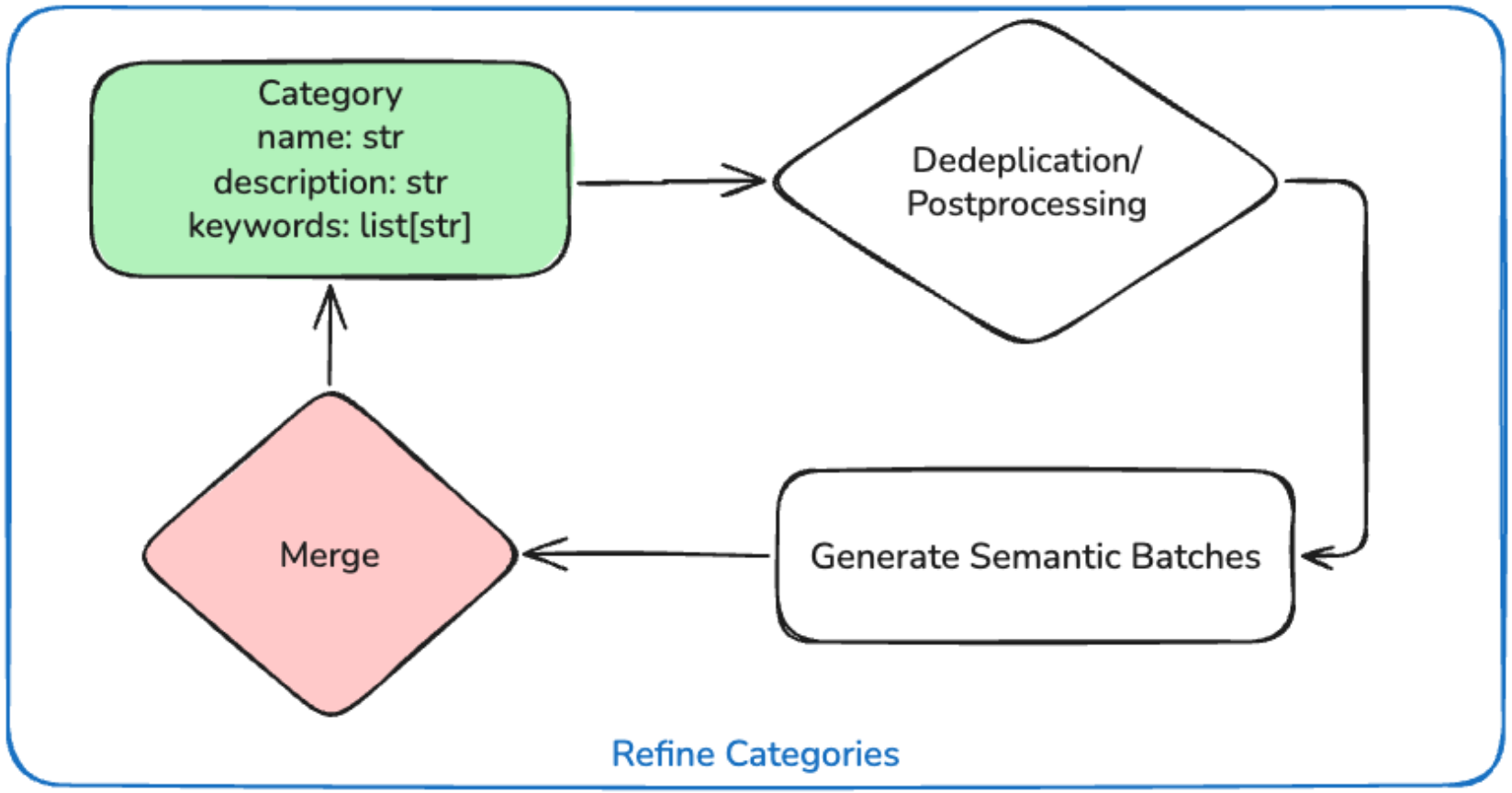
Title: Investigation and Improvement of Live Blue Swimmer Crab Handling in NSW
Description: As the NSW fishing industry moves forward into the future with quota managed fisheries, there is a need for industry to develop procedures...

Output:

```
{   "keywords": [
      {"name": "quota managed fisheries", "description": ..., "type": "methodology"},
      {"name": "value addition to wild harvest", "description": ..., "type": "application"},
      ...
   ]
}
```

SWINBURNE UNIVERSITY OF TECHNOLOGY

Merge



Merge

System prompt:
You are an expert Taxonomy Specialist and Research Category Harmonization Expert, specializing in identifying and merging identical research categories to create unified, coherent taxonomies for strategic analysis...

User input:

<category><name>...</name><description>...</description><keywords>...<keywords>
</category>

<category><name>...</name><description>...</description><keywords>...<keywords>
</category>

Output:

```
{
   "categories": [
      {"name": "quota managed fisheries", "description": "", "source categories": [""]},
      ...
   ]
}
```

Future works

Quality control for the categorisation and merge process

Further examination of different clustering/embedding algorithms

SWINBURNE UNIVERSITY OF TECHNOLOGY

Thank you

- Link to our tool