



RADAR PROJECT - HORIZON SCAN AUTOMATION

Presented by:
Anh Le, Ashley Wang and Sagar Bhagwatkar



Welcome to Our Project Presentation

Good afternoon.

Thank you all for joining today's session. I appreciate your presence and participation.

Agenda

- 01 Background
- 02 Objectives
- 03 Achievements
- 04 Key Limitations
- 05 Show Case
- 06 Q&A

What is today's agenda?

We'll begin by delving into the background of our project, providing context for our discussion.

Next, we'll outline the objectives we aim to achieve through this initiative, clarifying our goals and desired outcomes.

We'll then review the achievements we've made so far, highlighting our progress and successes.

Following that, we'll discuss the key limitations or challenges we've encountered, identifying areas for improvement and growth.

Finally, we'll conclude with an example of our project, showcasing its features, functionalities, and impact.

If you have any questions, please don't hesitate to ask anytime, or you can save them for the Q&A session.

Background

Insights from the WHS Radar are derived from:

- the analysis of existing databases (Australian Business Register), social media data and data reports (Australian Bureau of Statistics),
- a review of the international grey and academic literature, and
- consultations with Australian workers, with senior WHS professionals and WHS inspectors.

As familiar to us, Insights from the WHS Radar report come from

Analyzing existing databases

Evaluating international grey and academic literature.

Engaging in consultations with Australian workers, senior WHS professionals, and inspectors.

With an extensive manual review article process in place, our objective is to automate this task.

Objectives



Our project aims to automate the extraction of up-to-date and pertinent content from a selection of reliable public data sources, cleaning and analyzing the information, and finally delivering valuable summaries and insights.

- 01 Generate all links from Media/News pages
- 02 Scrapping data with Title, Content, Publish Date, Link of Article
- 03 Data cleaning to filter the articles within selected date range and key words
- 04 Summarize each article

What steps can we take to accomplish our project goals?

By extraction of relevant content from trusted public data sources, followed by data cleaning and analysis to produce insightful summaries.

Here's a breakdown of the steps involved:

Collect clickable links from diverse media/news pages.

Extract article details such as title, content, publish date, and article link

Process the data to filter articles based on specific date ranges and keywords.

Summarize each article to distill key insights efficiently.

Achievements

	Headline	Content	Date	Reference URL	Similarity Score
18	Asbestos removal licence suspended after safet...	The contractor also failed to give SafeWork SA...	2024-01-24	https://www.safework.sa.gov.au/news-and-alerts...	0.145463
20	Locking in our major public holidays	The Public Holidays Act 2023 was finalised in ...	2023-11-16	https://www.safework.sa.gov.au/news-and-alerts...	0.058703
37	2023 Augusta Zadow Awards announced	The SafeWork SA awards were presented by the G...	2023-10-20	https://www.safework.sa.gov.au/news-and-alerts...	0.206498
46	Carbon Monoxide: A threat in your workplace	Carbon monoxide: a threat in your workplace/n...	2024-03-25	https://www.safework.sa.gov.au/news-and-alerts...	0.085564
57	Changes to regulations for Amusement devices	On 25 December 2023 amendments relating to amu...	2023-11-30	https://www.safework.sa.gov.au/news-and-alerts...	0.174004
58	Holiday rostering reminder	Section 13B of the Shop Trading Hours Act 1977...	2023-12-15	https://www.safework.sa.gov.au/news-and-alerts...	0.019332
61	Retail work hazards under the microscope	From March, SafeWork SA inspectors will visit ...	2024-03-15	https://www.safework.sa.gov.au/news-and-alerts...	0.180482
62	Building site blitz reaches new heights	Since 3 July, SafeWork SA inspectors have perf...	2023-11-13	https://www.safework.sa.gov.au/news-and-alerts...	0.097568
65	Dangers of bypassing safety devices	Inspectors recently attended a worksite where ...	2023-12-12	https://www.safework.sa.gov.au/news-and-alerts...	0.304078

Data Clean steps

- remove duplicate
- convert to a valid date format
- drop the rows if date is empty
- filter date
- key words similarity check
- find threshold for similarity

Here is the highlight we've achieved thus far

The table demonstrates the results obtained after the data from the safework SA website has been cleaned and filtered.

Achievements - (cont'd)



Next I will be talking about our achievements with the content summarisation with 3 different models.

The first model we try is LLAMA 2. Its advantage is that it can provide a very descriptive summary as you can see here. And above all, its free of charge.

Pros and Cons

Model	Pros	Cons	Recommendations
LLAMA2	<ul style="list-style-type: none"> • Free to use • Detailed summary • No hallucination 	Computational resources required	
Gemini Pro	<ul style="list-style-type: none"> • Concise summary • No hallucination 	Subscription required in the future	If computational resource is an issue, Gemini Pro is recommended
LLAMA Chat	<ul style="list-style-type: none"> • More concise than LLAMA 2 with better statistical details • Free to use 	Computational resources required	If computational resource is not an issue, LLAMA Chat is recommended

In the data cleaning step to filter the articles within selected date range and key words,

As regards to the content summarisation step, with LLAMA 2 and LLAMA Chat, more computational resources is required. So anything above i7 and a decent GPU is recommended.

As for Gemini Pro, So if you think Gemini Pro is a suitable solution, you may consider purchasing the subscription in the future.

