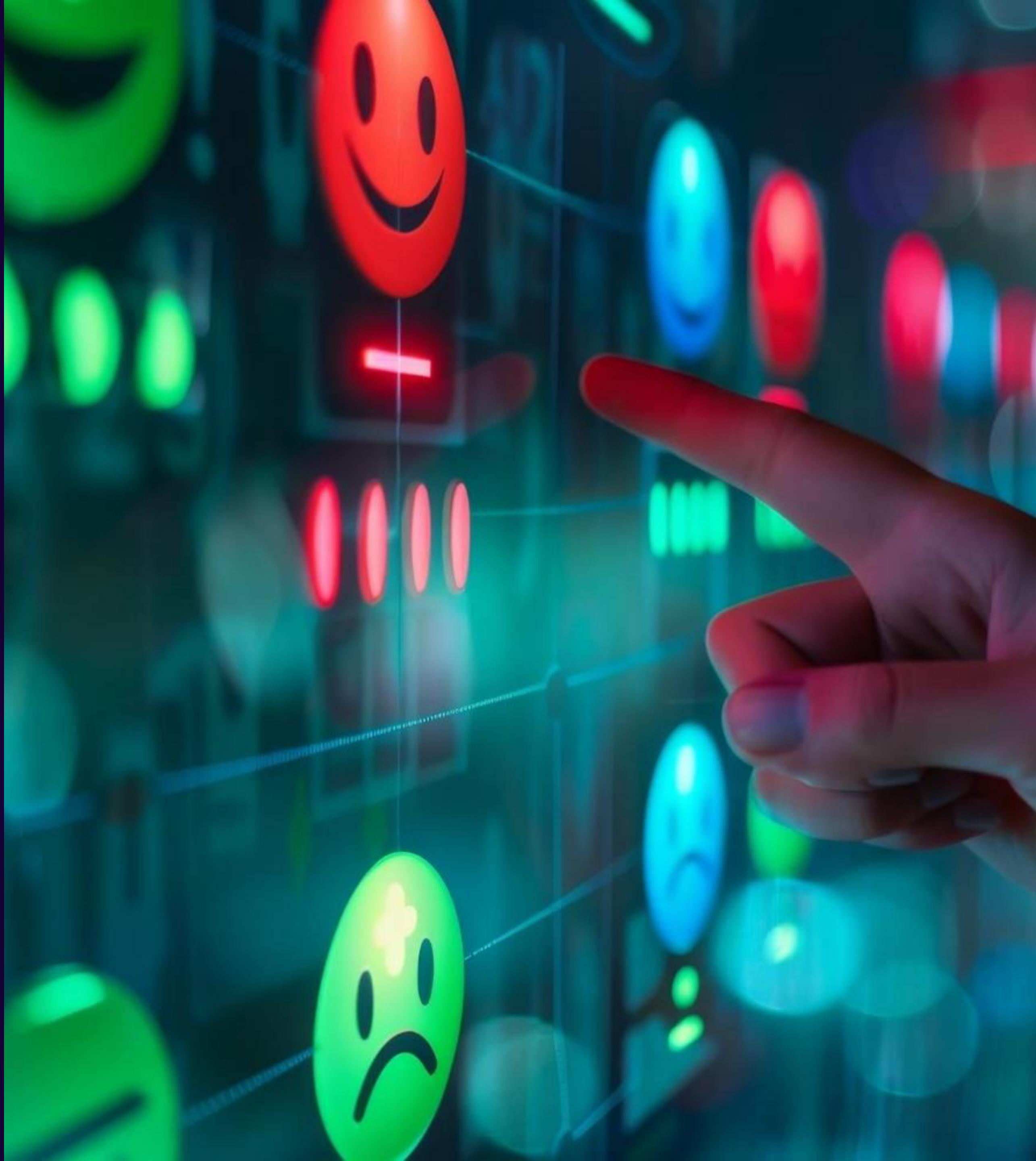
18 September 2025

Floating Beyond the Bubble

Unlocking insights on complex systems change with sentiment analysis

Dr Squirrel Main

ACIL ALLEN



Place-based systems change is messy….

ACIL ALLEN

GIGO concept – some solutions

Aggregators	Google Scholar Microsoft Academic Science.gov JSTOR Directory of Open Access Journals
Methodological Solutions	Database selection (sampling / comparisons) Imputation
LLMs & AI	Research rabbit Litmaps SciSpace Newspaper Navigator (early to mid-1900s) ChatGPT & other chatbots Meltwater and in-built AI
Government	Hansard GovConnex Council records
Search Options	Nexis ProQuest Elicit Public library databases School library databases

How about coding 20,000 articles?

How about not…

Latent Dirichlet Allocation Topic Modeler (LDA)

- **Imagine organising a messy library:** You have many books but no clear labels, and you want to group similar books together.
- **LDA finds themes (topics):** LDA scans through lots of text (like books, articles, or tweets) and tries to guess what themes (or topics) are inside, based on how often certain words appear together.
- **It's a bit like sorting puzzle pieces:** LDA figures out which words go together across many documents to form meaningful clusters or themes.

Credit: Seth Tucker, TCC Group, 2024 AEA

LDA Topic Modeler

Text 1

I love throwing snags on the barbie and baking lamingtons when the family gets together once a year.

Text 2

Australia is investing in nuclear-powered submarines, with debates about defence and technology growing every year.

Text 3

Baking pavlova and Anzac biscuits, and roasting lamb, is fun, especially for big family gatherings every year.

Tokenising

Text 1

I love throwing snags on the barbie and baking lamingtons when the family gets together once a year.

Token #	Token
1	I
1	love
1	throwing
1	snags
1	on
1	the
1	barbie
1	and
1	baking
etc	etc

Get rid of stop words

Text 1

I love throwing snags on the barbie and baking lamingtons when the family gets together once a year.

Token #	Token
~~1~~	~~I~~
1	love
1	throwing
1	snags
~~1~~	~~on~~
~~1~~	~~the~~
1	barbie
~~1~~	~~and~~
1	baking
etc	etc

Tokenise the lists of words

WORD	TEXT
love	1
throw	1
snag	1
barbie	1
bake	1
lamington	1
family	1
get	1
together	1
year	1

WORD	TEXT
australia	2
invest	2
nuclear	2
submarine	2
debate	2
defence	2
technology	2
grow	2
year	2

WORD	TEXT
bake	3
pavlova	3
anzac	3
biscuit	3
roast	3
lamb	3
fun	3
big	3
family	3
gathering	3
year	3

Apply LDA (Illustrative)

Text 1	Text 2	Text 3
love throw snags barbie bake lamington family get together year	australia invest nuclear submarine debate defence technology grow year	bake pavlova anzac biscuit roast lamb fun big family gathering year

Play! (1 – how many groups, 2 - what shall we name them)

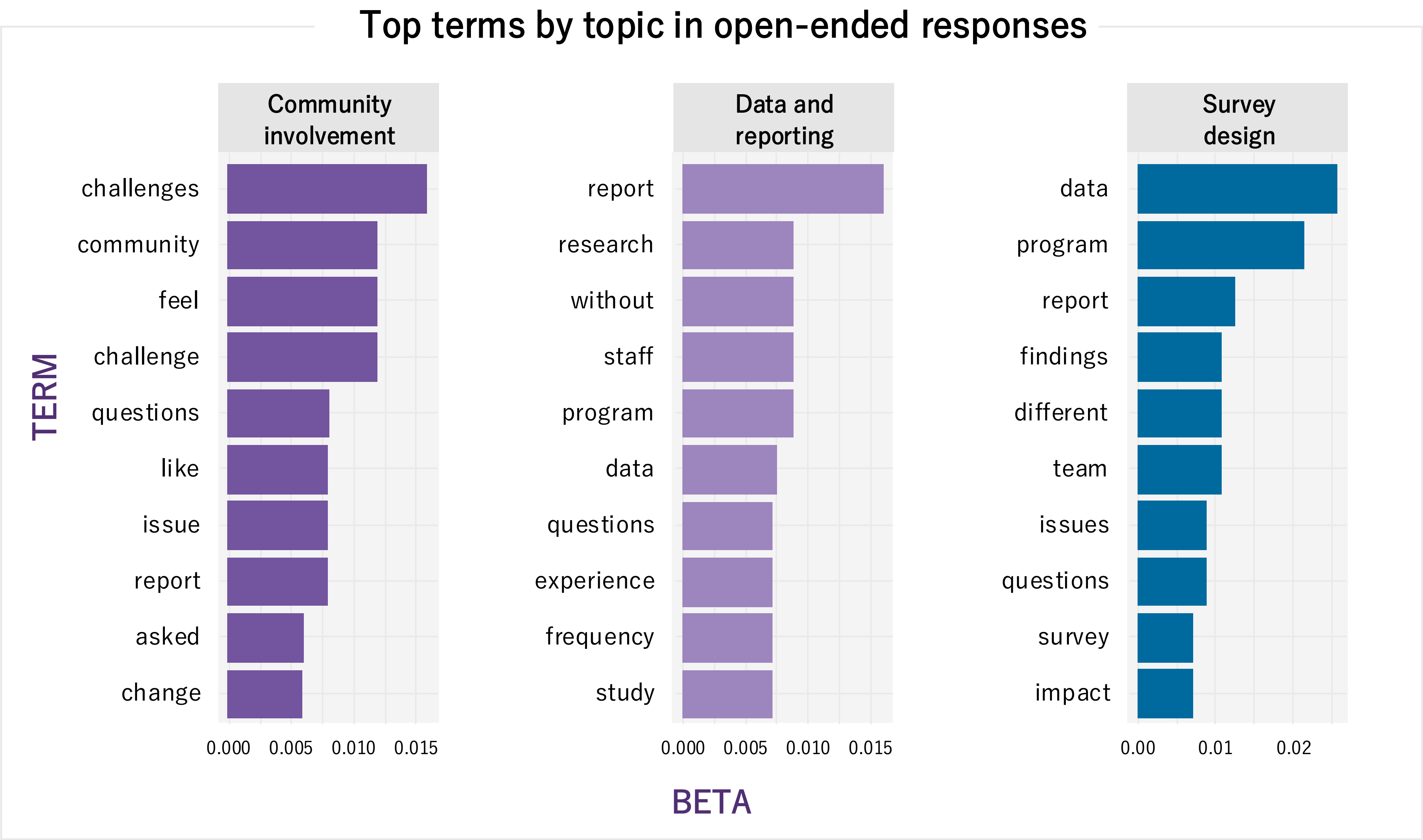
Topic 1 – Preparing food for family	
love	biscuit
snags	roast
barbie	lamb
bake	family
lamington	gathering
pavlova	year
anzac	

Topic 2 – Nuclear defence	
australia	defence
invest	technology
nuclear	grow
submarine	year
debate	

LDA – Let's practise

Topic A		Topic B	
oval	tackle	community	funding
goal	umpire	incarceration	partnership
behind	fans	prevention	young people
kick	premiership	diversion	crime
mark	grand final	rehabilitation	prison

LDA – Everyday ethics example



Limitations – LDA

- Doesn’t allow you to quantify
- Struggles with overlapping topics
- Needs a lot of data
- Sometimes hard to interpret results

We now have the “what”, but what about the “how”?

Sentiment Analysis

- **Imagine listening to many conversations:** You hear lots of comments but can't immediately tell if people are happy or upset. Sentiment analysis helps with that.
- **It identifies feelings (sentiments):** Sentiment analysis analyses words in text and tries to understand the emotions behind them.
- **It uses a lexicon, or sentiment dictionary:** It figures sentiments out on its own, like understanding that words like "love," "happy," and "pleased" are positive, while "hate," "angry," and "disappointed" are negative.

Credit: Seth Tucker, TCC Group, 2024 AEA

Data

Text 1

This café's food is the best! I loved the meat pie and the hot chips.

Text 2

The customer service was shocking; the staff are unhelpful and always leave me waiting.

Tokenising

Text 1

This café’s food is the best! I loved the meat pie and the hot chips..

Token #	Token
1	This
1	café’s
1	food
1	is
1	the
1	best
1	I
1	loved
1	the
etc	etc

The customer service was shocking; the staff are unhelpful and always leave me waiting.

Word	Category
the	customer service
customer	customer service
service	customer service
was	customer service
shocking	customer service
the	customer service
staff	customer service
are	customer service

Etc.

This café's food is the best! I loved the meat pie and the hot chips.

Word	Category
this	food
cafe	food
food	food
is	food
the	food
best	food
I	food
loved	food

Etc.

Remove stop words

The customer service was shocking; the staff are unhelpful and always leave me waiting.

Word	Category
~~the~~	~~customer service~~
customer	customer service
service	customer service
~~was~~	~~customer service~~
shocking	customer service
~~the~~	~~customer service~~
staff	customer service
~~are~~	~~customer service~~

Etc.

This café's food is the best! I loved the meat pie and the hot chips.

Word	Category
~~this~~	~~food~~
cafe	food
food	food
~~is~~	~~food~~
~~the~~	~~food~~
best	food
~~I~~	~~food~~
loved	food

Etc.

Join words with lexicon

The customer service was shocking; the staff are unhelpful and always leave me waiting.

This café's food is the best! I loved the meat pie and the hot chips.

Word	Category	Sentiment
customer	customer service	
service	customer service	
shocking	customer service	negative
staff	customer service	

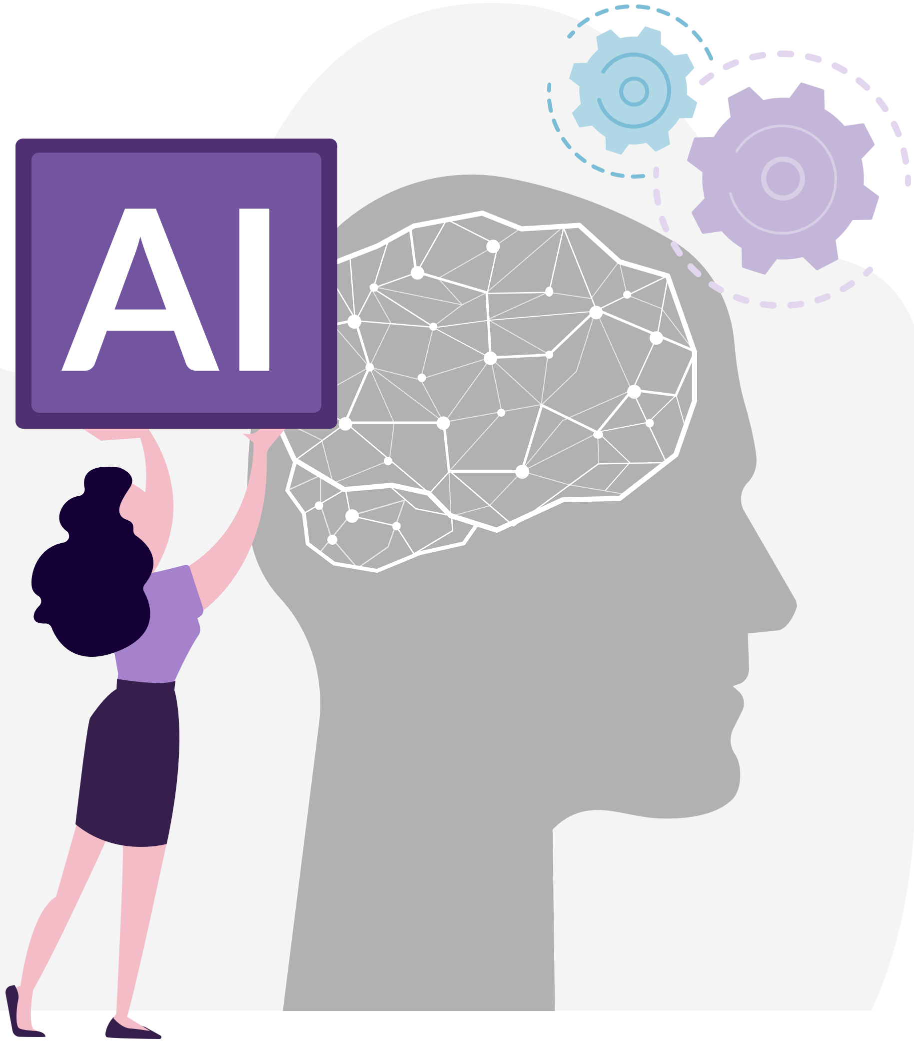
Word	Category	Sentiment
cafe	food	
food	food	
best	food	positive
loved	food	positive

Example sentiments table

Category	Positive words	Negative words	Net positive
Food	53%	22%	31%
Customer service	14%	44%	-30%

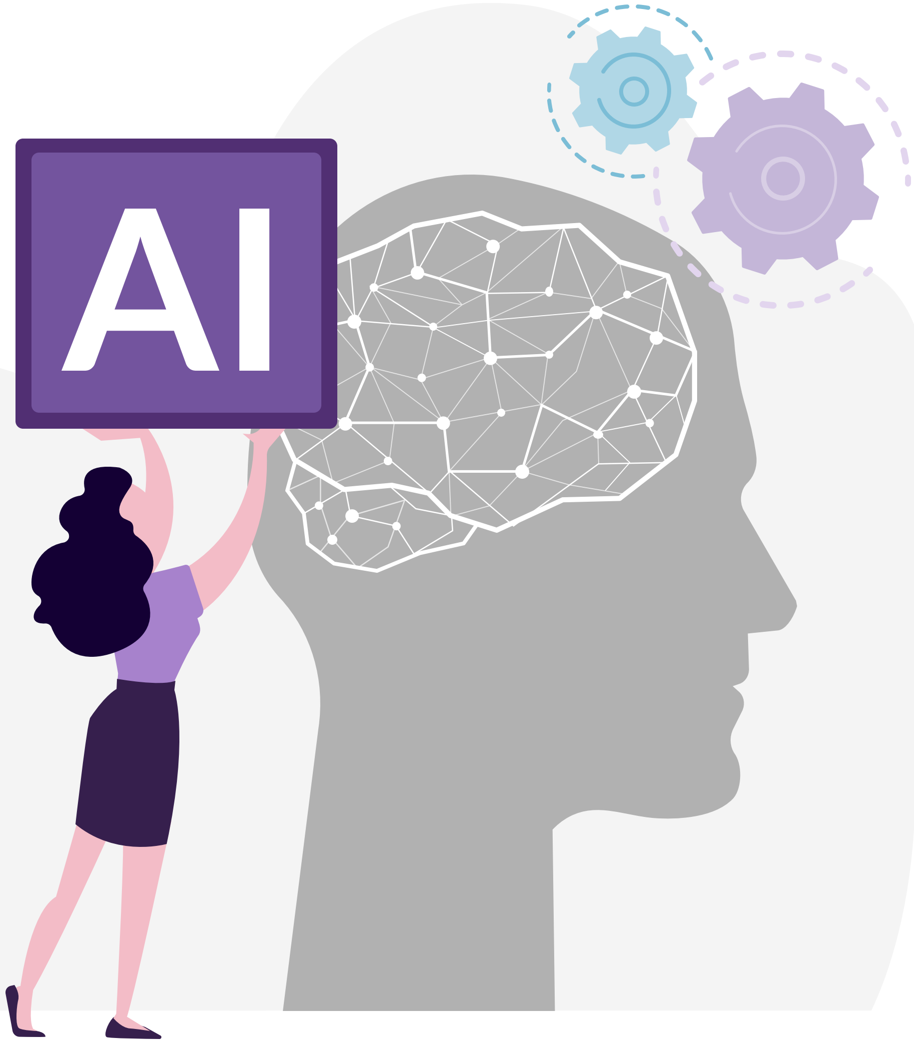
Let’s play – do you think like AI?

BING lexicon (positive or negative)	
Brilliant	Legend
Dodgy	Hopeless
Rubbish	Lamington



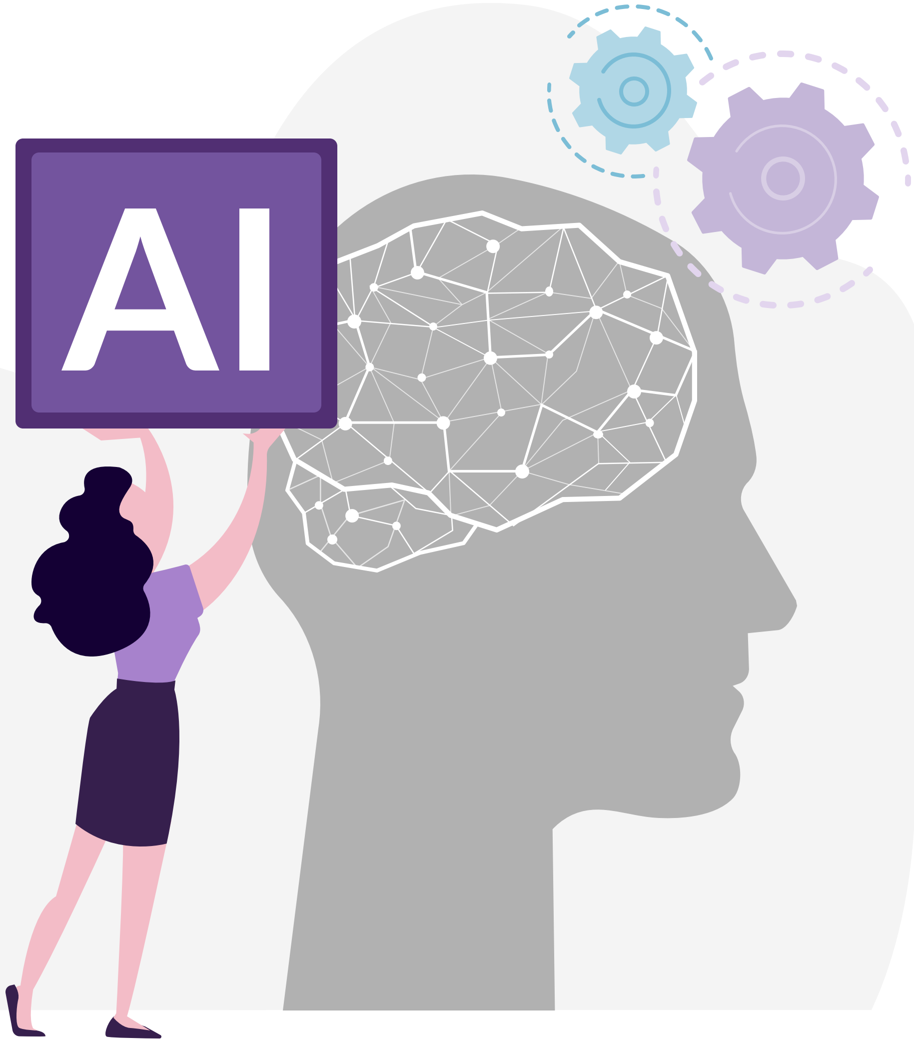
Let’s play – do you think like AI?

AFINN lexicon (-5 to +5)	
Brilliant	Legend
Dodgy	Hopeless
Rubbish	Lamington



Let’s play – do you think like AI?

NRC lexicon	
Maps words to multiple emotions: (“negative”, “positive”, “anger”, “anticipation”, “disgust”, “fear”, “joy”, “sadness”, “surprise”, or “trust”)	
Brilliant	Legend
Dodgy	Hopeless
Rubbish	Lamington



Limitations – Sentiment Analysis

Loss of context

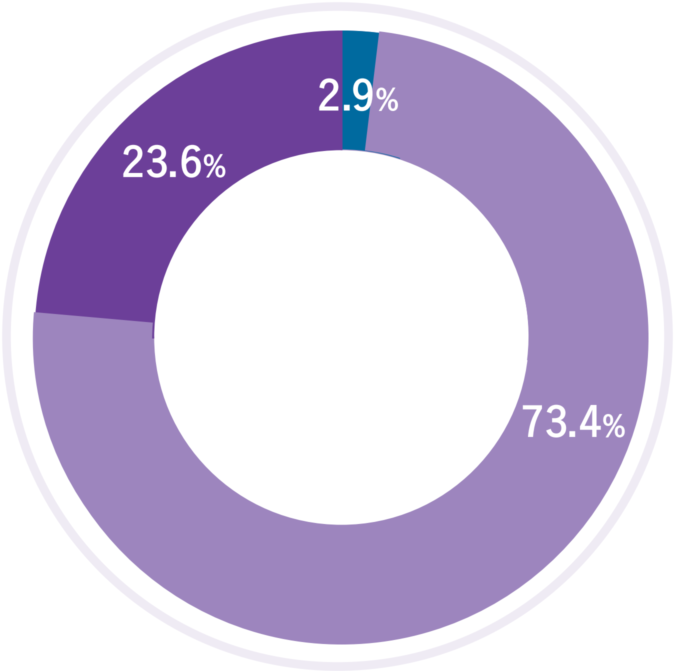
Slang, misspellings and out-of-vocabulary

Handling complex sentiment

Real-life example: Sentiment Analysis

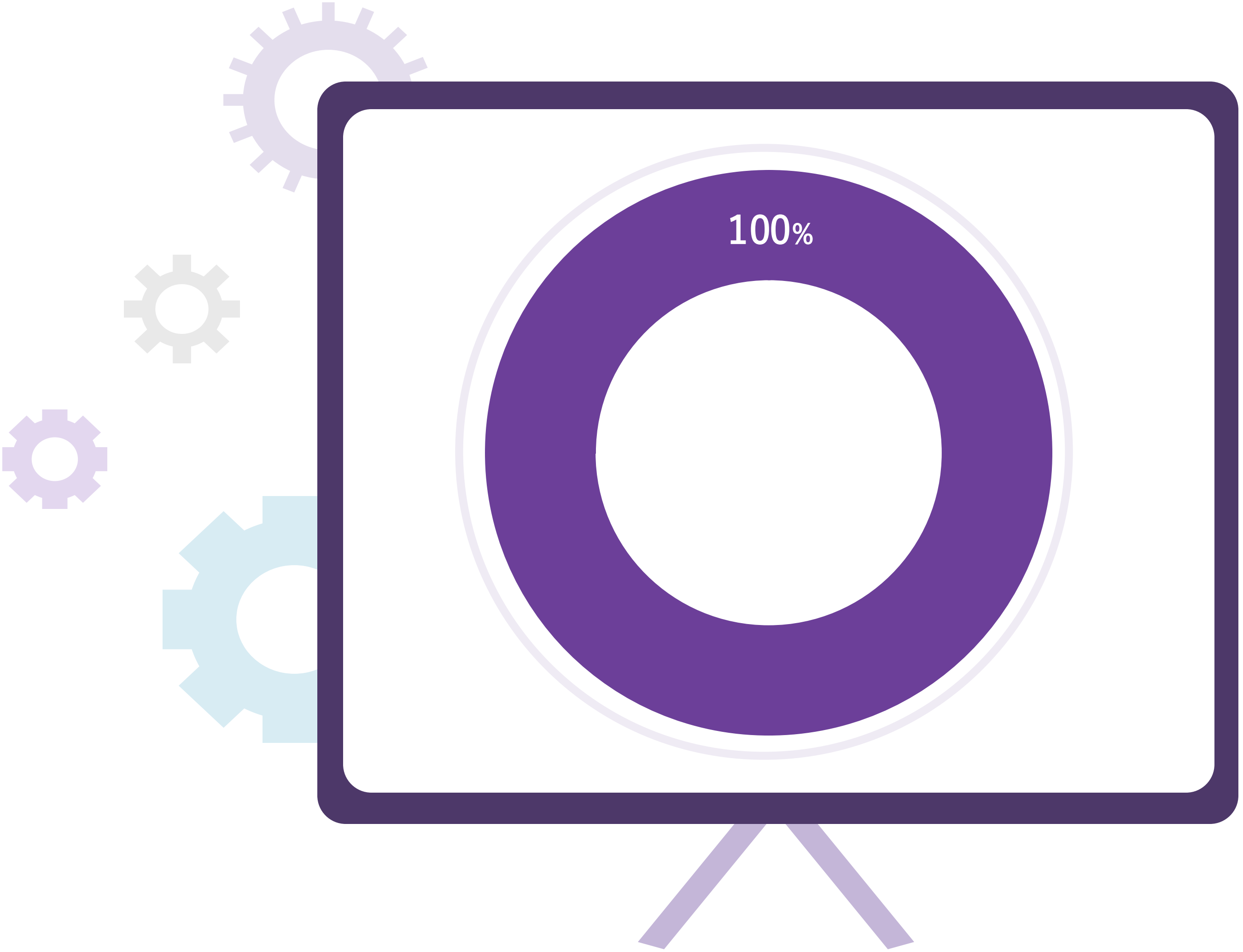
Even with its imperfections, it still tells a story!

Sentiment Analysis – BING + Meltwater – “police” across NSW



●	Positive	2.9%
●	Not rated	0%
●	Neutral	73.4%
●	Negative	23.6%

Sentiment Analysis – BING + Meltwater – “police” in Mt Druitt



Positive	0%
Not rated	0%
Neutral	0%
Negative	100%

Nitty-gritty of HOW

R, Python and the like

Courses to learn to code

- DataCamp
- Coursera
- Edx
- Many more!

Stack Overflow

- The Google for coders

LLMs

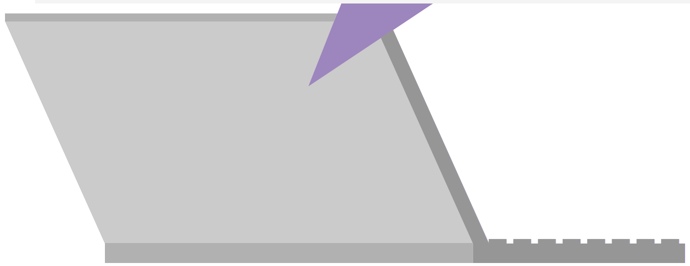
- ChatGPT
  - Paid: Specific GPTs for writing code
- Bing Copilot
- Bard

LLM Prompts

- Code to merge many excel files together in R
- Code to tokenise words in text in python
- Code to load a csv file into python
- Please, could you take a look at this dataa and run any interesting analysis you would like. Tell me the most interesting things you find. Thanks!

R code – ChatGPT – I’m happy to email

```
#############################################################
# Topic Modeling & Sentiment Analysis Workflow in R
# This script shows how to:
# - Clean text data
# - Remove stop words
# - Standardise words
# - Build a document-term matrix
# - Run Latent Dirichlet Allocation (LDA) for topics
# - Perform sentiment analysis using lexicons
#############################################################
```



```
# 1. Install and load required packages -------------------

install.packages(c("tidyverse", "tidytext", "topicmodels"))

library(tidyverse)   # data wrangling + visualisation
library(tidytext)    # text mining (lexicons + tokenisation)
library(topicmodels) # topic modeling (LDA)

# 2. Load your text data ---------------------------------

# Example: load a CSV file with at least two columns:
# "article" = document ID, "text" = raw text
df <- read_csv("text.csv")

# 3. Tokenise text into words and remove stop words --------

df_token <- df %>%
  unnest_tokens(word, text) %>%    # split text into words
  anti_join(stop_words)            # remove common words like "the", "and"

# 4. Standardise word forms (basic stem)
```

Emerging trends – Credit Chattie!

01

Transition from Self-Developed LLMs to Agentic Ecosystems

- **AI agents (“autopilot” mode)** empowered by LLMs are increasingly performing tasks autonomously – shifting from co-pilot to full automation in workflows and decision-making.
- **Multimodal LLMs** capable of processing text, images, and audio in real time are now mainstream, enabling more adaptive and dynamic AI assistants.
- Example: India’s Kruti AI showcases agentic architecture – breaking requests into sub-tasks, interfacing with APIs/databases, and orchestrating actions with >90% accuracy.

02

Accessibility: From Awareness to Compliance

- While accessibility remains a key concern, research shows that over **80% of deployed chatbots still have critical accessibility issues**, highlighting widespread gaps in inclusive design.
- AI advancements in assistive technologies and inclusive design are improving – but implementation lags behind capability.

Emerging trends

03

Specialised Assistants Within Multi-Agent Frameworks

- AI ecosystems are evolving such that **specialised assistants now operate under orchestrator layers**, coordinating across domains like health, finance, personal productivity, and more.
- Agents can now handle complex, multi-step tasks autonomously – a leap from today's single-purpose bots toward collaborative AI teams

04

Two Dominant UX Approaches Emerge

- **Zero-knowledge ("protected") chatbots:** Simplified, user-friendly interfaces requiring minimal expertise – favoured in consumer and low-barrier contexts.
- **Walled LLMs with controlled access:** Tailored, gated systems for enterprises and specialised tasks – offering performance and oversight, but requiring more expertise to use and manage

Emerging trends

05

Mixed Modeling: Blending Quantitative, Qualitative, and Agentic Signals

- Hybrid approaches continue ruling the day – **combining qualitative feedback, quantitative metrics, and agent-generated structure** to refine AI output and understanding.
- Increased adoption of agentic pipelines that include **multi-modal data, internal evaluation, and cross-agent agreement checks**, enhancing reliability and explainability.

Happy to chat more!

Dr Squirrel Main
0414 842 859
s.main@acilallen.com.au

acilallen.com.au

ACIL ALLEN