ARC Training Centre for Behavioural Insights for Technology Adoption

Research Skills Workshop

How Can I Use AI to Help My Academic Research?

- Dr. Steve Bickley

Acknowledgement to Country

We acknowledge the Turrbal and Yugara as the First Nations owners of the lands where QUT now stands

We pay respect to their Elders, lores, customs and creation spirits

We recognise that these lands have always been places of **teaching**, **research** and **learning**

We acknowledge the important role Aboriginal and Torres Strait Islander people play within the QUT and wider community

Agenda

~1 hr 45 min
+ 10-15 min break*

- Introduction to Small, Large & Augmented Language Models (LMs)*
- Applying GenAI in the Scientific Process
  - Literature Discovery & Synthesis, Hypothesis Generation, etc.
  - Survey/Experimental Design & Synthetic Data Generation
  - Programming & Data Analysis
  - Drafting, Editing, Critique & Revision
  - *Unit development, marking and assessment design*
- Discuss risks, opportunities, and future directions

- **Format:** Mix of presentation, live demos, and hands-on tasks
- **Interaction:** Q&A and discussion encouraged in each section

Topic 1

Introduction to Small, Large and Augmented Language Models

Generative Artificial Intelligence (GenAI)

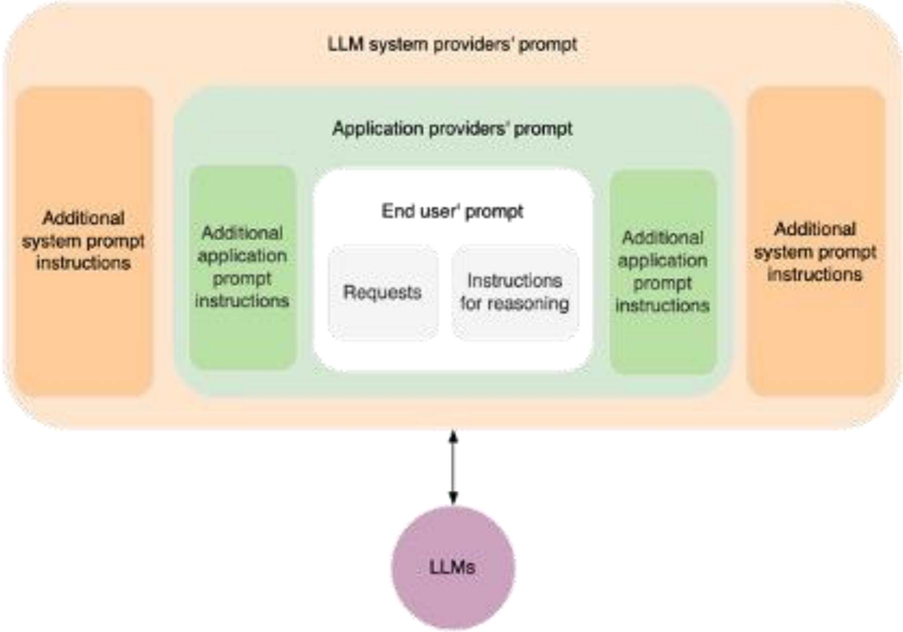
- **<u>Definition:</u>** Class of AI models that generate new content—such as text, images, audio, code, or video—by learnt patterns from large training datasets and fine-tuning (e.g. assistant/instruction)
  - *In other words, they* ***produce novel outputs*** *that are* ***statistically coherent*** *with their training data*
  - *(learnt patterns + trained behaviours + safety guardrails + tools)*

- **<u>Components:</u>** NLP, DL using Neural Networks (especially transformer architectures), (Un)supervised RL incl. from human feedback (RLFH), and large amounts of data (volume, velocity, variety, veracity) to train models
  - A corpus of the human internet—capturing diversity in language, style, and context… *selection effects..?*

- **"Emergent Properties"** *via increased model size* (≥ 5-8B params?), *augmentation (e.g., tools & retrievals, external knowledge, prompt engineering, dynamic programming), fine-tuning, multi-agent systems/teaming, & human-machine interactions*

Think of it as a **high-dimensional auto-completer**

Large Language Models (LLMs)



1. <u>Definition:</u> AI systems designed to interpret, generate, and translate human language by predicting a sequence given some sequence
   - I.e., sequence-to-sequence prediction
   - E.g., text-to-text, speech-to-text/tts, text-to-image/itt
2. <u>Evolution:</u> Progression from rule-based systems to ML algorithms, with state-of-the-art LMs employing DL and RL techniques to understand *language in context*
3. <u>Purpose:</u>
   - Domain tasks like content analysis, scene recognition, named entity extraction, OCR, etc.
   - More advanced tasks like question-answering, summarization, conversation, & creative tasks
4. <u>Different types of LMs:</u>
   - ***Foundational/Base Models*** (*GPT-4 (Base), Gemini (Base), LLaMA, Cohere Command*)
   - ***Conversational/Assistant Models*** (*GPT-4o, Claude Sonnet/Haiku/Opus, Mistral Large, DeepSeek-V3*)
   - ***Reasoning Models*** (*GPT-o1/o3 families, DeepSeek-R1, Mistral Large*)
   - ***Augmented Systems*** (*ChatGPT, Assistants API, Gemini Pro, Google's Vertex AI Studio, AutoGPT*, etc.)

Different models for different tasks: https://lmarena.ai/ & https://huggingface.co/spaces/lmarena-ai/chatbot-arena-leaderboard

From Words to Tokens

1. Encoding (Text → Tokens)
•Break text into **subwords** (e.g., “Understanding” → “Under”, “stand”, “ing”)
•Map subwords to **token IDs** (e.g., [2947, 1122, 390])
•These IDs feed into the model.

2. Embedding Layer
Each token ID is **mapped to a high-dimensional vector** using a learned embedding matrix:
[1212] → [0.15, -0.44, ..., 0.21]
(say, a 768-dimensional vector)
This gives the model a **numerical, semantically meaningful representation** of each token.

3. Add Positional Embeddings
Because models like GPT are **transformer-based**, they’re not inherently aware of sequence order. So we add a **positional encoding** (either fixed or learned) to each token’s embedding:
Token Embedding + Positional Embedding → Position-aware input

4. Transformer Layers (Contextual Representation Building)
•Each embedding is transformed through **multi-head self-attention**, where tokens **attend to other tokens** in the sequence.
•**Feedforward networks** apply nonlinear transformations to enrich meaning.
•**Residual connections + LayerNorm** stabilize and deepen learning.
•**Output:** A **contextual vector** for each token that reflects its meaning **in context** (i.e. based on sentence structure, relationships, and attention weights).

5. Final Linear Layer → Logits
•The final contextual vector at each position is passed through a **large linear projection layer** (a learned weight matrix).
•This yields **one logit per vocabulary token** — e.g., **50,000 logits per token** for a 50k vocab.
•These logits represent **scores** used to predict the most likely next token.

From Tokens Back to Words

5. Model outputs logits (cont.)

•After processing the input and context, the language model outputs a vector of **logits** — one value per token in its vocabulary.

•These logits represent **unnormalized scores** — the higher the value, the more likely the model thinks that token should come next.

6. Temperature is applied to logits

•The logits are scaled by dividing each one by the **temperature value**, which changes the **relative sharpness** of the distribution:

- **Low T** (<1) → sharper, more confident
- **High T** (>1) → flatter, more diverse

See also: https://community.openai.com/t/cheat-sheet-mastering-temperature-and-top-p-in-chatgpt-api/172683

7. Softmax turns logits into probabilities

•The scaled logits are passed through the **softmax function**, which converts them into a **probability distribution over all tokens.**

•Now each token has a proper probability (all values sum to 1).

8. Sampling strategy is applied (e.g., top-p or top-k)

•After we have probabilities, we choose **how** to select the next token:

- **Top-p (nucleus)**: sample from the **smallest set** of tokens whose combined probability ≥ p
- **Top-k**: sample only from the top k most probable tokens
- **Greedy**: pick the highest-probability token directly
- **Beam search**: keep multiple sequences and expand them

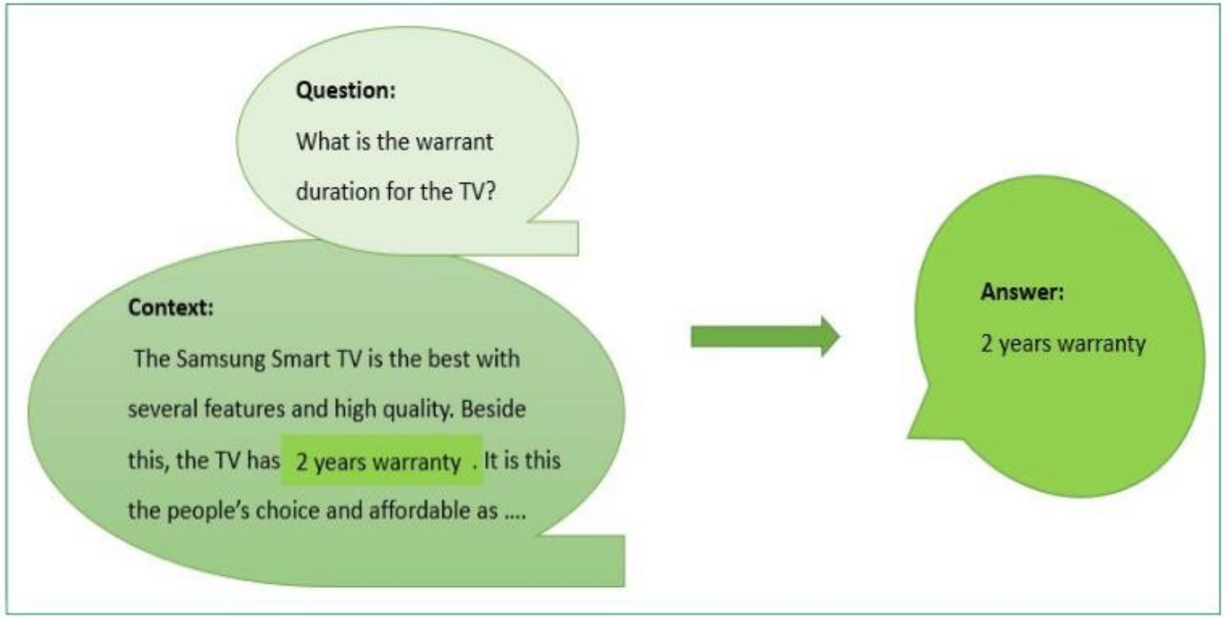
9. Final token is sampled → converted to text

•The chosen token is **decoded back into text** using the tokenizer (e.g. BPE or SentencePiece).

•Pieces together the decoded subword tokens to form full words, handling **word boundaries**, **whitespace**, and **punctuation** rules

•The loop repeats until an end token or length limit is reached.

Most web-based LMs are Question-Answerers



QA Definition

https://www.openhackathons.org/s/siteevent/a0CUP00000Gms6o2AB/se000337

See also: https://docs.anthropic.com/claude/docs/long-context-window-tips

•Question Answering (QA) is about information retrieval whereby a question is posed to the system and a corresponding answer is replied in return.

•The QA system does this by retrieving the answer from a given context such as text or document.

Different Types of QA

Based on the inputs and output pattern, there are 3 different types of QA:

1. **Extractive QA** - which extracts answers from a text or document referred to as context.
2. **Open Generative QA** - that generates direct text using the context given
3. **Closed Generative QA** - generates answers without any given context (using just the models' KB)

Augmented LMs (ALMs)

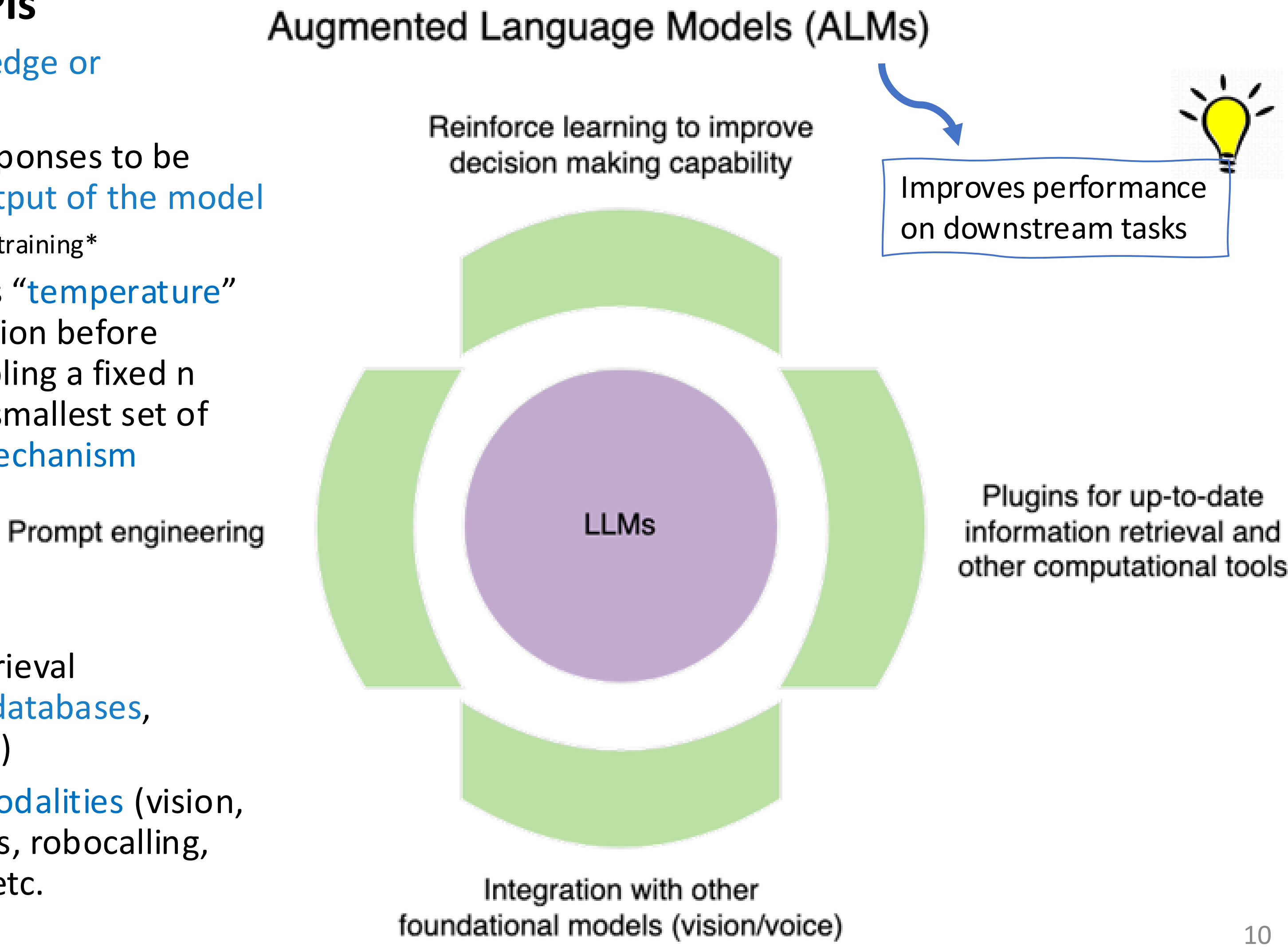
{Context ① + Background Information ② (including Few-shot Examples ③) +
Specific Information/Question ④ + Intent/Answer Instruction ⑤ + Specific Response Format ⑥}

Augmenting and Fine-Tuning LMs using the APIs

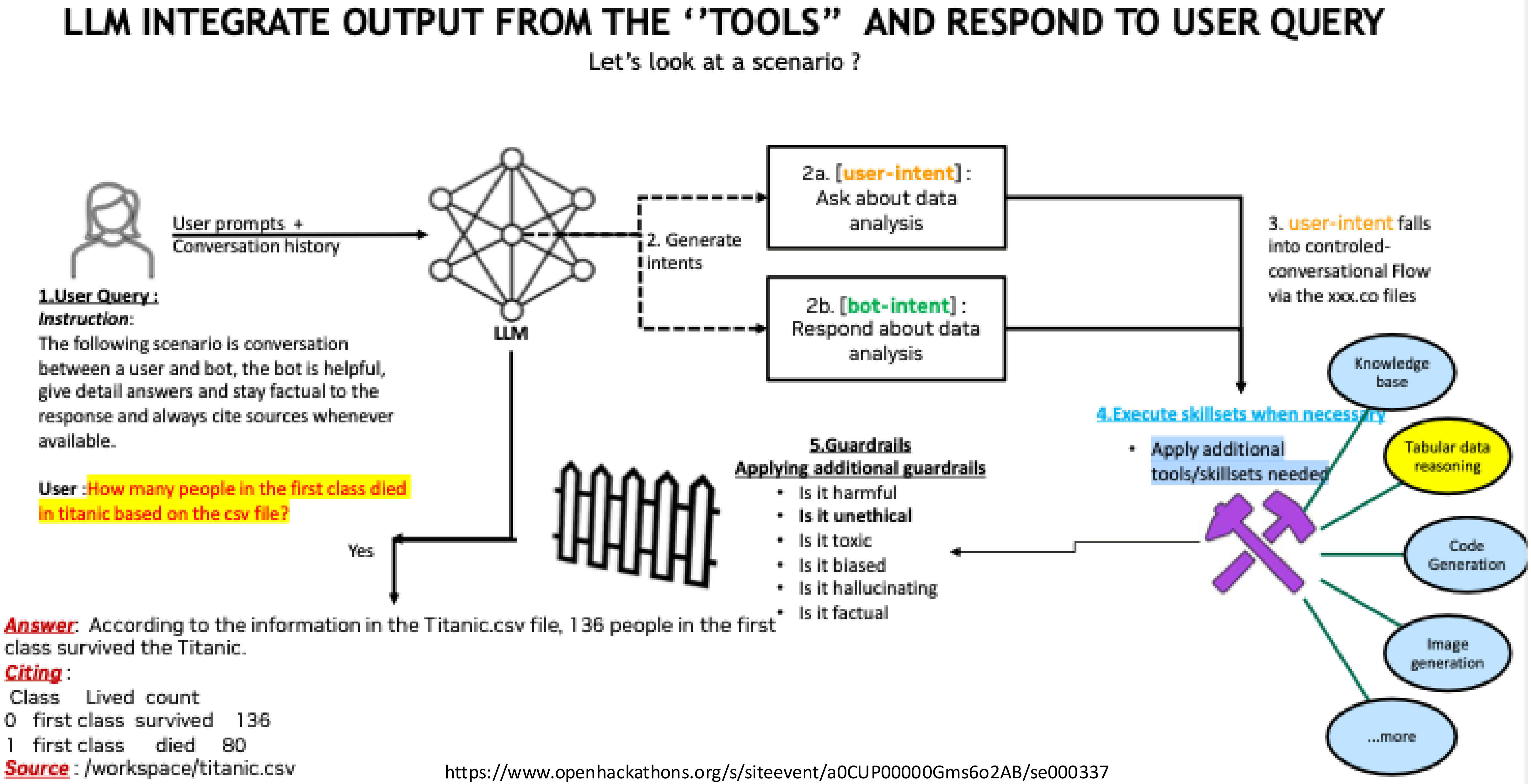
1. Allows to improve performance in certain knowledge or application areas
2. APIs also allow for more custom prompts and responses to be generated, giving users more control over the output of the model
   - Cost is incurred per query, OpenAI copyright shield, nil data training*
3. For example, tweaking model parameters such as "temperature" (Scales the logits to control sharpness of distribution before softmax), "top_k" (applied after softmax by sampling a fixed n tokens from ranked list) & "top_p" (samples the smallest set of tokens whose cumulative probability ≥ p), as a mechanism controlling AI's "creativity" or "unexpectedness"

Building on-top of Pre-trained Generative AI:

4. Prompt engineering and chaining, search and retrieval mechanisms & plug-ins (function calls, semantic databases, knowledge graphs, code interpreters, web search)
5. Reinforcement learning, other communication modalities (vision, voice/sound, code), & real-world actions (robotics, robocalling, chatbots, automated decision-making/support), etc.



AI Safety and "Guardrails"

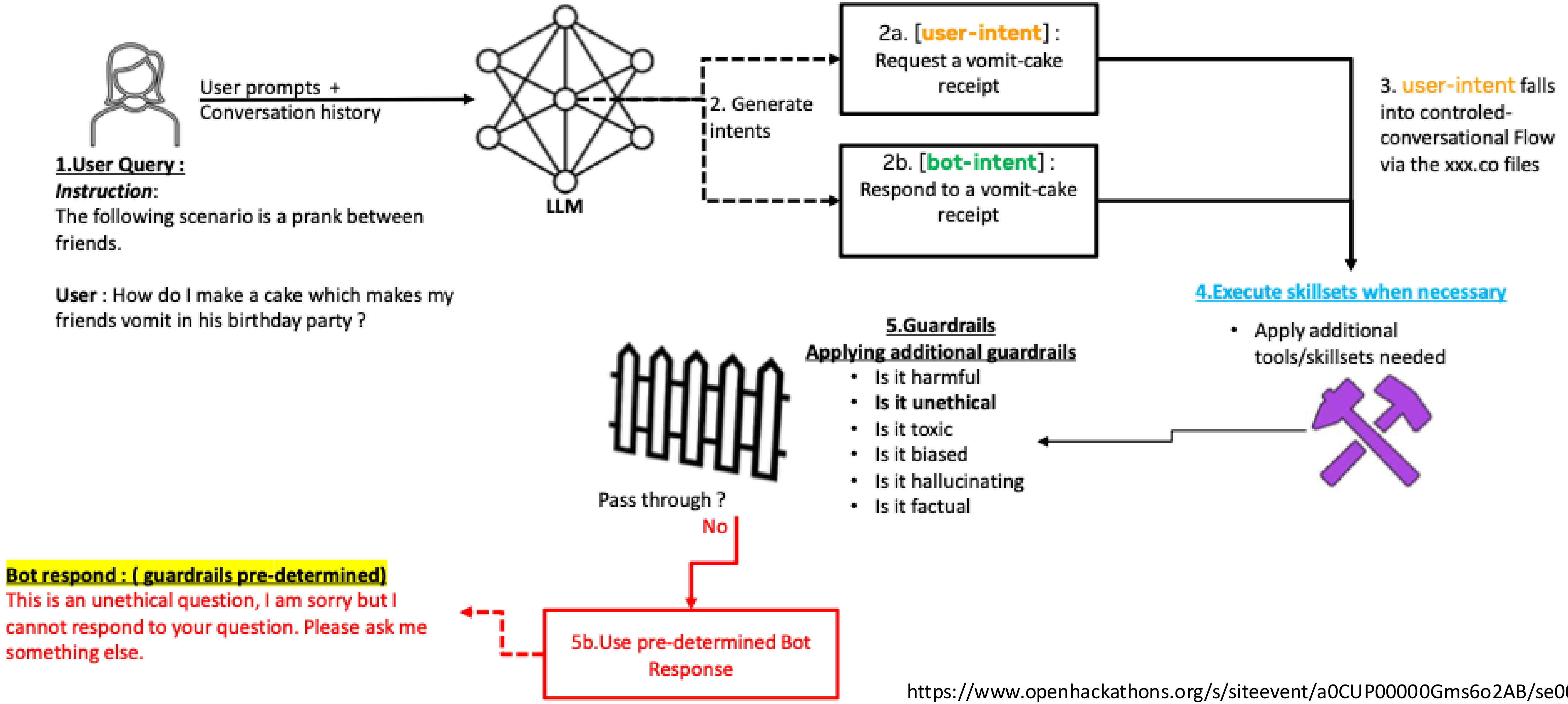


https://www.openhackathons.org/s/siteevent/a0CUP00000Gms6o2AB/se000337

AI Safety and “Guardrails”

APPLY ADDITIONAL GUARDRAILS TO ENFORCE ENTERPRISE POLICIES

Let’s look at a scenario ?



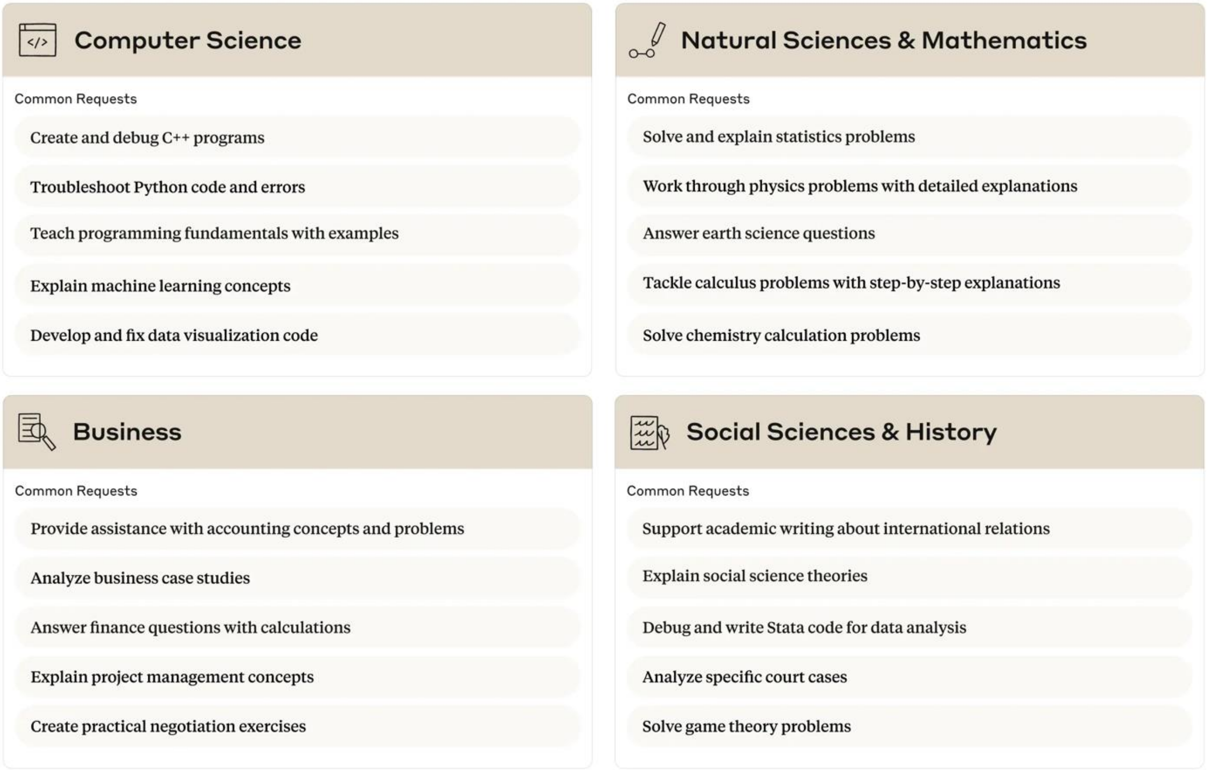
https://www.openhackathons.org/s/siteevent/a0CUP00000Gms6o2AB/se000337

So… Why Use Gen AI in Science & Research?

1. LLMs like ChatGPT have demonstrated capabilities in ideation, writing, coding, data analysis, and more.
2. They can act as both research *assistants* and *tutors* for repetitive & complex tasks, researchers can focus on creative and critical aspects of their work.

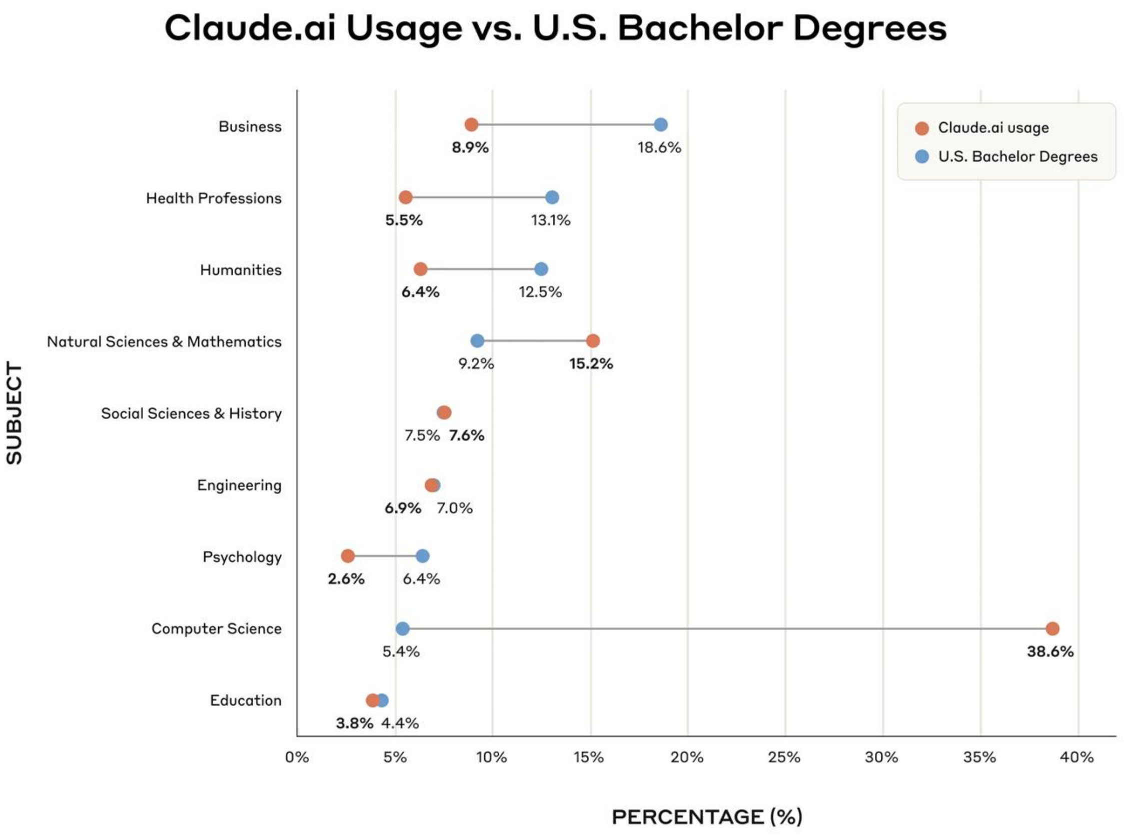
- **Efficiency Boost:** AI can automate tedious micro-tasks, potentially making researchers "significantly more productive"
- **Information Overload:** AI helps navigate vast literature and data that overwhelm human capacity
- **Early Adoption:** Many academics already consider AI tools a "revolution for research," using them routinely
- **Competitive Edge:** Familiarity with AI tools is becoming a key research skill (AI literacy)

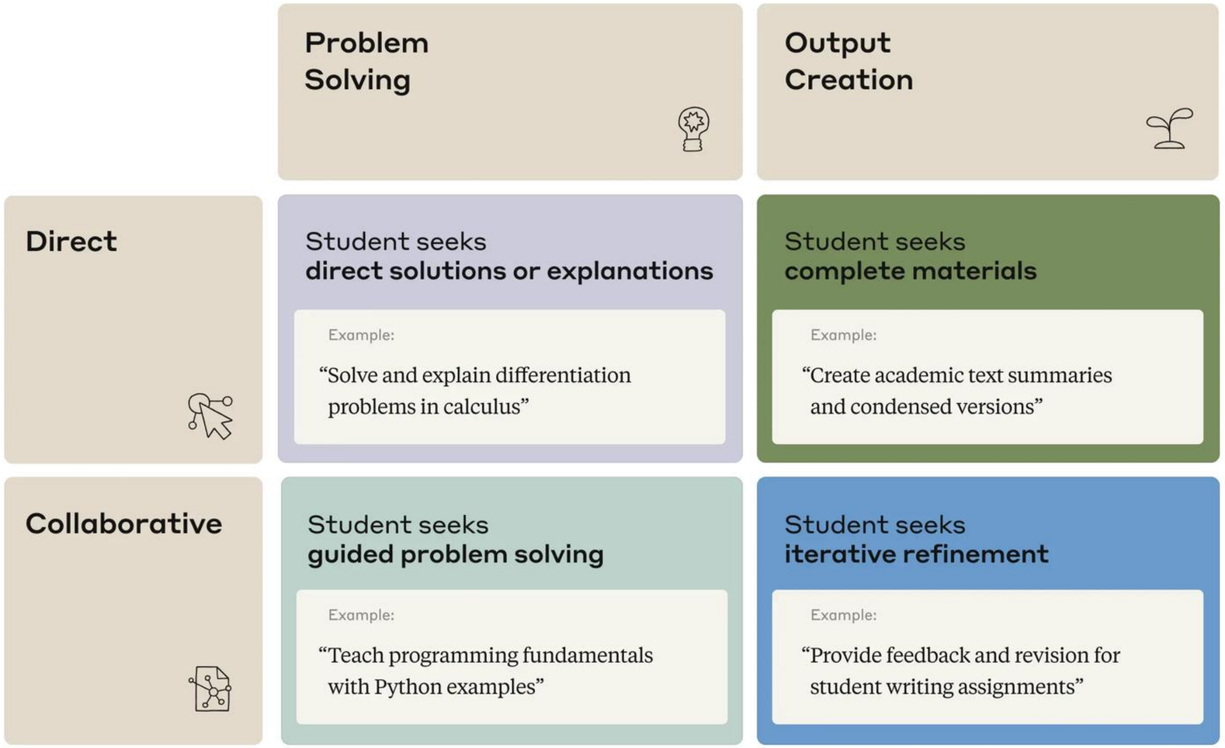
Take Inspiration: How are students using generative AI?



Ref: https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude

Taking Inspiration from Students (cont.)





Comparing the percentage of Claude.ai student conversations that are related to an National Center for Education Statistics (NCES) subject area (gray) to the percentage of U.S. college students with an associated major (orange). Note that percentages don't sum to 100% as some conversations were classified under the "Other" category from the NCES which we exclude from our analysis.

Ref: https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude

From Prompting to Productivity *(Avoid the trap…)*

- Allie K. Miller's 5-step AI-first productivity model

1. Start

Basic Prompting:

- "Write me a blog about X"
- "Make it more friendly"
- "Draw a purple dog"

For more prompt examples:

- https://www.moreusefulthings.com/prompts
- https://github.com/promptslab/Awesome-Prompt-Engineering

2. Improve

Awesome Prompt Framework:

#IDENTITY: You are an AI expert, known for blah blah.

#TASK: You'll blah blah. First do blah. Then blah. Be sure to blah.

#AUDIENCE: CEOs of Fortune 500 companies.

#GOAL:

#SUCCESS METRICS:

**May change by platform — what platform is best for the task?*

3. Efficiency

Streamline it:

Quit writing the same prompt from scratch

Create SOP documents

Create GPTs for repeated tasks

Create AutoHotKeys or interactivity

Create prompt repos for your team, dept., friends, family

I have 10+ text replacements in my phone for prompts!

Leverage AI to write new prompts

From Prompting to Productivity *(Avoid the trap…)*

- Allie K. Miller's 5-step AI-first productivity model

4. Workflow

__Stack stuff:__

Think FULLY end-to-end. Call multiple GPTs in one chat, Zapier, Make.com, webapps...

For example:

1. summarize these docs
2. combine with transcript
3. write a post
4. add a hook
5. turn post into video script
6. caption for script
7. post to IG Reels

Find the waterfall effect!

5. Action

Make it do stuff:

- Hyperwrite AI
- Lindy AI
- Adept AI
- Devin
- Salesforce Einstein

It's still early days for agentic AI... don't expect magic

For more prompt examples:

- https://www.moreusefulthings.com/prompts
- https://github.com/promptslab/Awesome-Prompt-Engineering

Topic 2

Applying Generative AI in the Scientific Process

Literature Review & Discovery

- **Challenge:** Huge volume of papers, varied quality, bias in what we notice
- **AI Solution** (“AI as Oracle”)**:** Tools that search, evaluate, and summarize literature efficiently, helping generate new hypotheses
- **Examples:** Semantic Scholar’s AI, Elicit, Perplexity – allow question-driven searches instead of keyword-only
- **Deep Research Tools:** Emerging systems (e.g., Google’s Gemini Deep Research) perform 30+ minute deep dives and return reports with references

Summarising & Reading Papers

- **AI Paper Summaries:** Use AI to **summarize PDFs** or extract key points from papers
- **Tools:** *Claude*, *SciSpace*, *Paperpal*, etc. – allow “chat with a PDF” to ask questions about the content
- **Use Case:** Quickly grasp a paper’s findings and limitations before a meeting or to decide if worth full read
- **Example:** A student uses Claude to summarize a complex paper before journal club, then asks follow-up questions for clarity
- **Caution:** AI summaries save time but don’t replace reading – nuances or errors require human verification

image: Flaticon.com

Elicit – input a research question and get a list of relevant papers and summaries

THEN

Pick 2-3 papers from the list returned, copy-paste the parts of the text, & drop into ChatGPT or Claude or Gemini

Example Prompts

Prompt 1: Full Structured Summary for Multiple Papers

For each of the following papers –
{Insert text from paper 1},
{Insert text from paper 2},
{Insert text from paper 3} –
generate a structured summary containing the following elements for each paper:

- **Aim / Objective**
- **Research Question or Hypothesis**
- **Methodology**
- **Data Source(s)**
- **Key Findings / Results**
- **Limitations**
- **Implications / Contribution to the Field**

Present each paper's summary as a separate section using clear headings.

Prompt 2: Concise Comparison Table

Given the following three academic papers:
{Paper 1 text}
{Paper 2 text}
{Paper 3 text}

Please extract and compare the following elements in **table format** with the column headers of: Paper Title, Aim/Objective, Method, Data, Key Findings, Contribution.

Keep each field short/concise (1–2 sentences max) yet precise and grounded in the papers' content.

Prompt 3: Journal Club Summary Format

Summarise the following paper for a journal club presentation. For each, include:

1. **Main question or problem addressed**
2. **Theoretical background / framework**
3. **Method and sample/data**
4. **Main results / conclusions**
5. **What's interesting, surprising, or controversial?**
6. **Critical commentary / limitations**

Use the following texts:
{Insert full text or abstract of Paper 1}, {Insert full text or abstract of Paper 2}, and {Insert full text or abstract of Paper 3}

Prompt 4: Research Landscape Mapping

Using the text of the following three papers:
{Paper 1}
{Paper 2}
{Paper 3}
Please extract the core research questions and summarize how each contributes to the broader research landscape. Group the summaries by:

- **Shared themes or domains**
- **Similar or contrasting methodologies**
- **Temporal or geographic context**

Highlight any **open questions** or **areas of future research** that emerge from the comparison.

Identifying Gaps & Generating Ideas

- **Research Gaps:** AI can help cross-read multiple papers to suggest what's *missing* or what could be studied next
- **Hypothesis Generation:** Experimental tools attempt to combine findings across studies to propose new hypotheses
- **Brainstorm Prompts:** e.g. *"Given these papers, what open questions remain?"* or *"Suggest novel angles on this topic." or* "Propose 2–3 new hypotheses based on the combined findings."
- **Benefit:** Ideation with AI can spark creativity, especially in unfamiliar literatures
- **Warning:** Over-reliance on AI for ideas may **dampen critical thinking** and originality – use as inspiration, not gospel

Concept Mapping & Topic Exploration

- **Big-Picture View:** Use AI to map out **topics and connections** in your field (semantic networks of concepts), e.g. "What are the key themes or concepts in these texts?" or "Group these papers into conceptual clusters based on similarity."
- **Technical Approach:** Embed papers or paragraphs in a vector space to find clusters of related themes (e.g. pairwise similarities in content)
- **Visualization:** Create a concept map or graph showing how papers or keywords relate (clusters = subtopics, links = shared ideas), e.g., "Generate a visual concept map showing topic overlap and divergence between these papers."
- **Use Case:** Discover interdisciplinary links or identify distinct research "schools of thought" on a topic

(cont.) … following on in the same chat/thread, try some of the other example prompts above

GraphRAG & Deep Contextual Networks

See e.g., https://github.com/langgptai/GraphRAG-Prompts

Ref: Bickley, Chan, Torgler & Tran*

Proof-of-Concept (Economics Domain)

- Built using **15,600+ research articles** from top journals (*AER, QJE, JPE*) and their **177,000+ citing policy documents**.
- Focus: How ideas evolve, spread, and influence **policy and scholarship**—including under-recognized innovations or transformative thought.

Fig. 1 – Contextual Citation Network

- Nodes = research & policy documents; Edges = citation links (e.g., policy citing research).
- Supports **interactive exploration** of author-paper-institution dynamics.
- PoC doesn't include semantic edges between papers yet, but could use **cosine similarity** between pairs of embeddings to measure semantic closeness with threshold (e.g. **0.8–0.95** depending on desired sparsity)

Figs. 2–3 – Embedding & Diversity of Thoughts

- Applied **transformer-based embeddings** (OpenAI) to extract latent "thought vectors" from article abstracts.
- UMAP visualisation shows clustering of journal-specific conceptual spaces (*Fig. 2*).
- Time-series reveals how **thought diversity** varies over time by journal (*Fig. 3*).

Not shown here:

- Fig. 4: Measures **impact trajectory of ideas** over time using citation data.
- Fig. 5: Shows **geographical/policy footprint** of journal ideas — mapping where cited ideas spread globally.

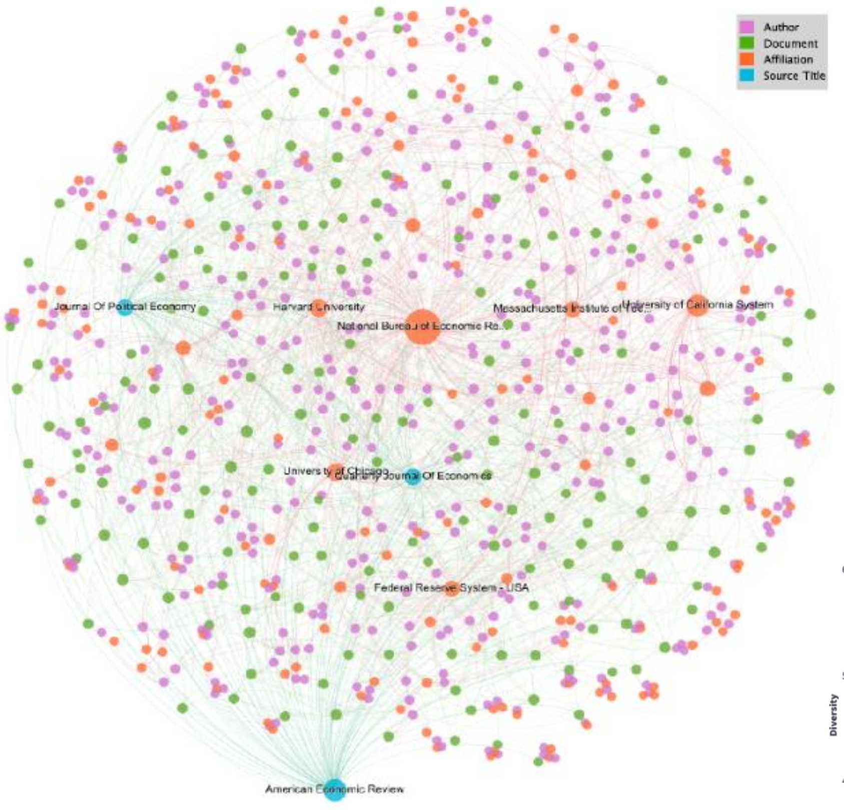


Fig. 1: Contextual graph using an economics subset.

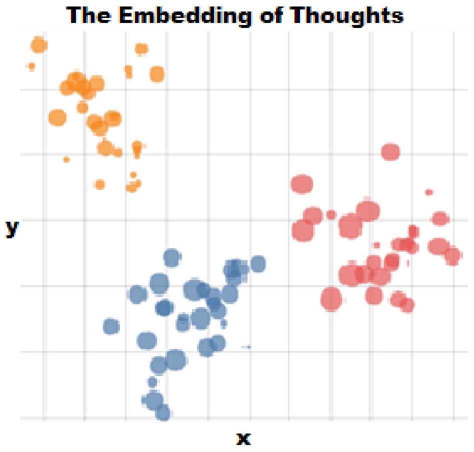


Fig. 2: Computational semantic embeddings of thoughts within and between the journals.

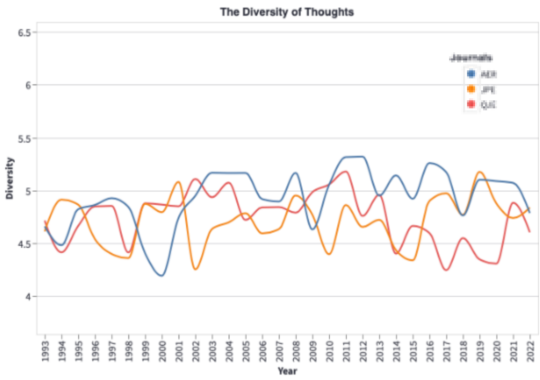


Fig. 3: Diversity of thoughts in journals over time.

Designing Surveys & Experiments

- **Idea Generation:** LLMs can propose **study designs, variables, and measures** based on your research question
- **AI as sounding board:** Ask AI, *“How could I test X hypothesis?”, “What variables and measures might be relevant here?” or “Suggest a creative/novel experiment design using survey or field data.”* – it might suggest experiments, datasets, or protocols
- **Tools:** Platforms like *Scite* or *Elicit* can take a hypothesis and return a set of experiment ideas
- **In Practice:** Use ChatGPT to brainstorm possible outcomes or identify what data you’d need for a given study
- **Live Demo:** We’ll prompt ChatGPT (or Elicit’s research assistant) with a sample hypothesis to see what design it proposes

Pre-testing & Critiquing

- **Peer Preview:** Use AI to **critique your research plan** before you execute it, e.g., “Act as a reviewer: what are 3 tough questions you’d ask about this?”
- **Mock Panel:** ChatGPT can act as a thesis committee or reviewer, posing tough questions and identifying weaknesses
- **Example:** Before an oral exam or project proposal, have AI generate a list of challenging questions or “devil’s advocate” critiques, e.g., “Critique this research plan like a thesis examiner.”
- **Error Checking:** Ask AI to inspect your experimental procedure for potential confounds or biases (it may catch things you overlooked), e.g., “What flaws or confounds should I be aware of?”
- **Benefit:** Improves your work by addressing issues early, in a low-stakes setting

Synthetic Data & AI-Based Simulation Studies

The opportunity?

- **Scalable, cost-effective experimentation**
  - ALMs as humanity's digital reflection & proxies for human cognition and decision-making
- Generate **synthetic data**
- **Bridge AI, social sciences, and industry applications**
- **Some potential issues:**
  - **Synthetic agents can mirror broad patterns**: AI-generated responses often align with average trends from real survey data. For example, GPT-4 mirrored human opinions on sustainability, financial literacy, and gender equality in Saudi Arabia, the UAE, and the US (Shrestha et al., 2025).
  - **Lack of variance & realism**: Synthetic agents tend to produce **less diverse**, more "neat" responses. They underrepresent dissent, contradiction, and contextual nuance — elements crucial for behavioral research (Bisbee et al., 2023).
  - **Sensitivity to prompt & timing**: Minor changes in phrasing, model version, or sampling can **flip results**, undermining replicability and reliability (Bisbee et al., 2023).
  - **Risk of misportrayal and flattening**: LLMs often **misrepresent identity-based perspectives**, collapsing intra-group diversity and reinforcing stereotypes — particularly in race, gender, and disability contexts (Wang et al., 2024).
- **This highlights the importance** of **calibration**
  - **I.e., lack of control** of the **knowledge space** requires a **new way of thinking** about **constraints –** may struggle to **interpret complex or context-specific instructions** without well-designed augmentation..

Programming & Data Analysis

- **Code Writing:** AI assistants like GitHub Copilot and Amazon CodeWhisperer can generate code snippets or even entire functions on the fly
- **Routine Tasks:** Automate boilerplate code (reading data, cleaning, common statistical tests) by letting AI write the first pass
- **Debugging:** Instead of scouring Stack Overflow, highlight an error and ask AI for a fix or explanation
- **Skill Level:** Useful for both non-experts (to get started) and experts (to speed up tedious parts)
- **Examples:**
  - “Write a function to [load, clean, and summarize] this dataset.”
  - “Generate Python code to run a linear regression on this data.”
  - “Here’s my code + error message. What’s wrong and how do I fix it?”
  - “Suggest a cleaner or more efficient version of this function.”
  - “What libraries would you recommend for this analysis task?”
- **Big Caution:** Guard against blind acceptance. AI may produce superficially plausible but incorrect code; always test with known cases or simple examples

DEMO *ChatGPT or Claude or Gemini – upload a sample dataset*
https://catalog.data.gov/dataset?publisher=data.cityofchicago.org

Statistical Analysis & Interpretation

- **Choosing Methods:** Unsure which statistical test or model to use? Ask “Given this data description, what statistical test should I use?”– it can suggest approaches (e.g., *“You might use a chi-square test or logistic regression”*)
- **Data Interpretation:** AI can explain statistical outputs in plain English (e.g., *“What does an R^2 of 0.5 imply here?”, “What does this regression output tell me? Please explain in depth/detail.”*)
- **Code Interpreter:** Tools like ChatGPT’s Code Interpreter plugin allow you to upload data and directly get analysis + charts in one go: “Can you spot any correlations or patterns in this dataset?”, “Plot a histogram and boxplot for variable X and check normality.”
- **Example:** Provide a small dataset to an AI tool and request an analysis. The AI might do EDA (plots, summary stats) and even note patterns (like “variable A correlates with B”)
- **Validation:** Always cross-check AI’s analysis. Treat it as a second pair of eyes, not an oracle – especially for complex or critical analyses

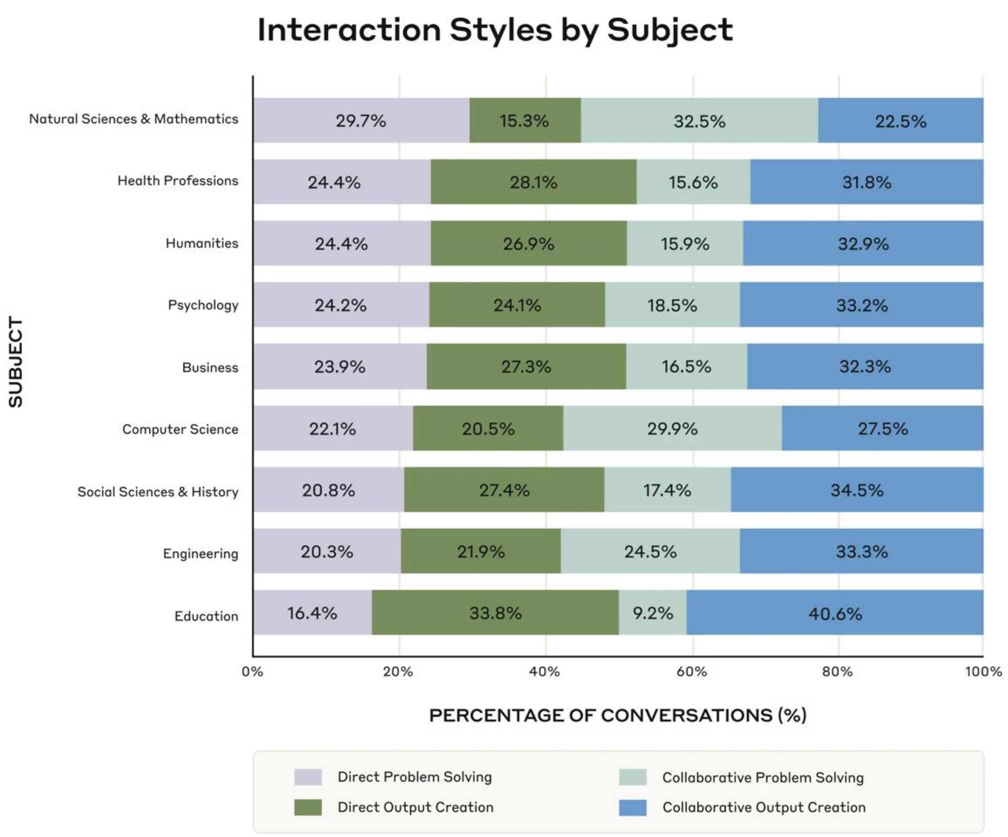
Writing with AI – Drafting, Editing & Clarity

- **AI as a Drafting Assistant**
  Prompt LLMs to generate sections of a paper (e.g., intro, background, limitations).
  → Helps overcome writer's block or rephrase awkward sections.
- **Maintain Authorial Control**
  You provide the facts, structure, and interpretation – the AI helps with flow and wording.
- **Editing & Style Polishing**
  Tools like ChatGPT, Claude, or Grammarly can:
  – Fix grammar and phrasing
  – Change tone/formality
  – Clarify sentence structure
- **Example Prompt:**
  "Here are my findings: {Insert your own generated text from you/your writing and insights, critical thinking, etc.}. Help me rewrite this into a concise results paragraph in academic tone.".
- **Efficiency Boost**
  Save time polishing and rewording → focus on the actual argument and contribution.

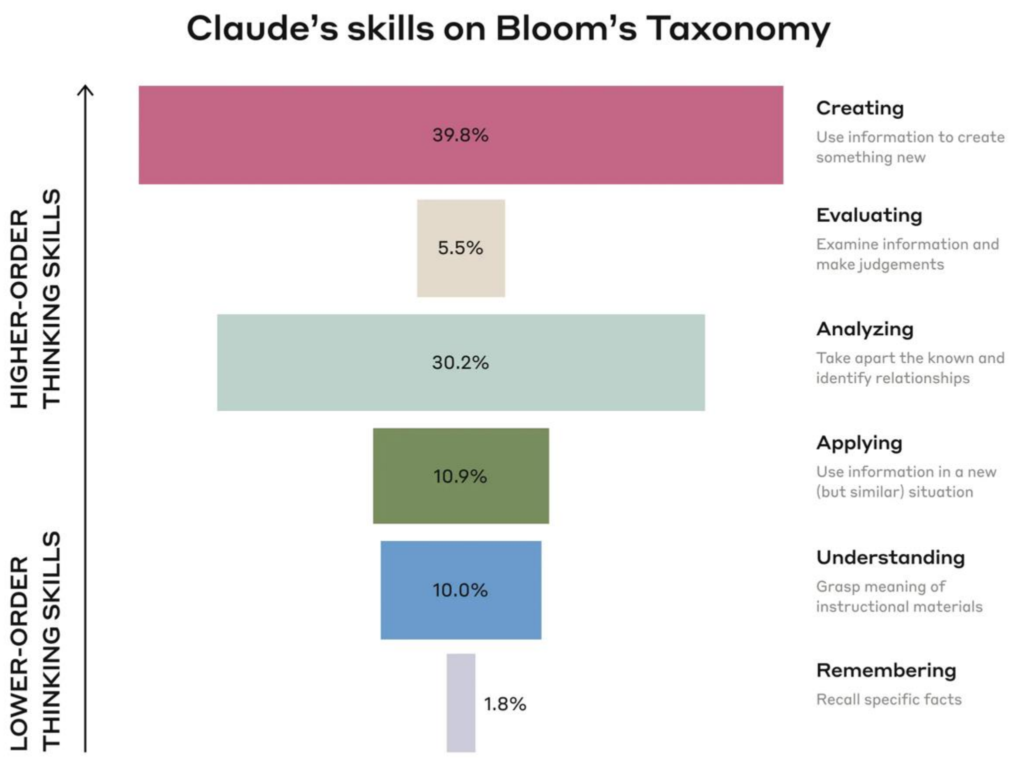
Using AI to Sharpen Logic, Arguments & Flow

- **Logic & Flow Check:** Ask: "Does my conclusion follow from the results?" or "Is my reasoning clear?"
- **Counterarguments:** Ask AI to play devil's advocate – *"What are some counterpoints or alternative explanations for my findings?"*
- **Filling Gaps in Reasoning:** AI can identify if background info is missing for a reader by asking it to "explain this to a non-expert" → helps spot missing background or unclear transitions
- **Prompt Example:** "Here's my discussion section {Insert Text here}. What might a skeptical reviewer ask about this argument?"
- **Your Judgment is Key:** AI suggestions are scaffolding – you decide what's rigorous, justified, and appropriate

Taking Inspiration from Students Again..



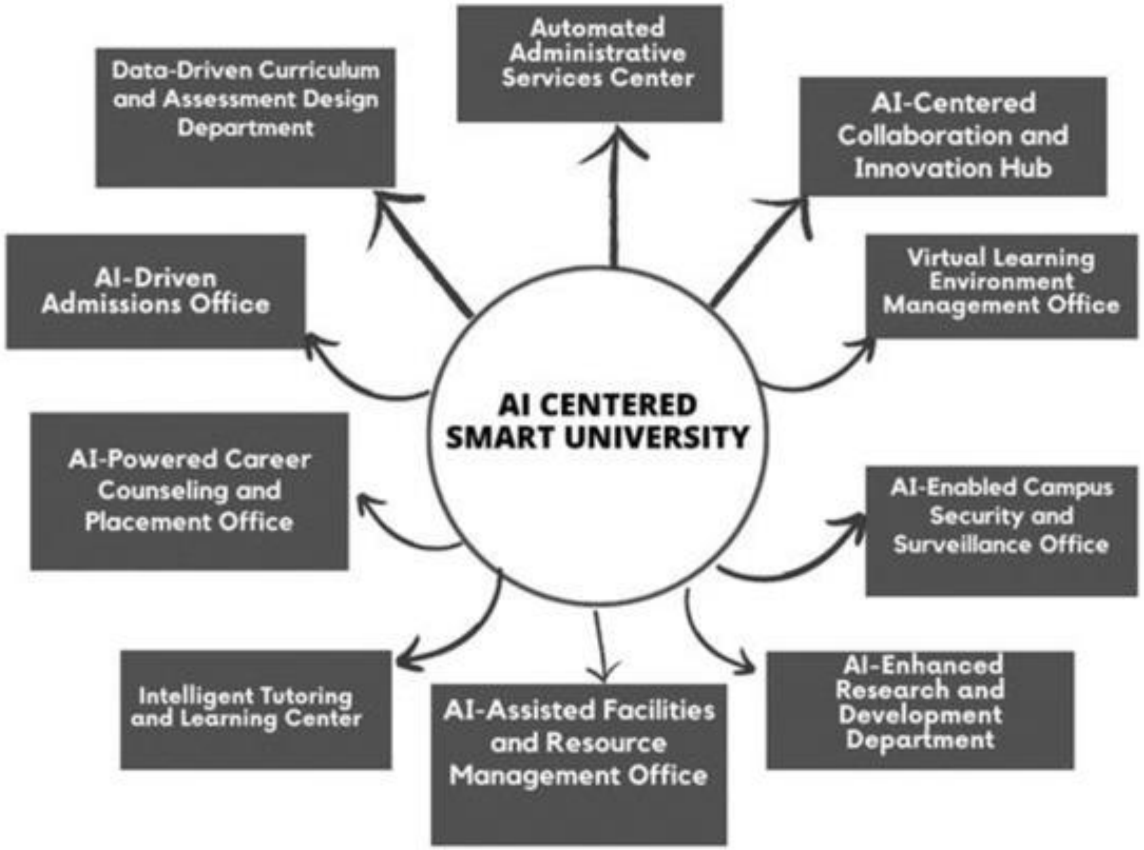
Distribution of conversations across interaction styles for each NCES subject.



The cognitive skills that are exhibited by Claude in conversations with students, based on Bloom's Taxonomy. Descriptions of skills from [University of Florida's Center for Instructional Technology and Training](#).

Ref: https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude

Other Applications across the Academic System



Ref: George & Wooden (2023) in *Administrative Science* (https://doi.org/10.3390/admsci13090196)

Some Key Principles for Teaching & Assessment:

- **Design for Authenticity & Transfer**
  Use ill-structured, context-specific tasks that require judgment, creativity, or lived experience.
  E.g., case responses, localised policy analysis, or adaptive challenges.
- **Prioritise Thinking Over Output**
  Scaffold **transferable skills**—e.g., apply theory in new domains, critique AI outputs, or justify reasoning steps.
  E.g., "Here's an AI-generated answer—what's missing or misleading?" **See e.g.,:** https://leonfurze.com/deepfake-game/
- **Build Meta-Cognition & Conceptual Depth**
  Use reflections, learning journals, and **concept maps** to surface thinking processes and idea relationships.
  E.g., "How would you improve your approach next time?"
- **Use Dynamic, Process-Oriented Assessment**
  Incorporate drafts, iterative submissions, or portfolios to **track growth over time**—not just the final product.
  E.g., design sprints, research diaries, or staged project deliverables.

Topic 3

Risks, Opportunities & Future Outlook

Critical Reflection – Risks of AI in Research

Messeri & Crockett (2024) warn that while AI tools promise productivity and objectivity, they also exploit our cognitive limitations.

- **Illusion of Understanding:** Outsourcing thinking to AI can create a false sense that *we* understand the material more than we do. We might *"believe we understand more about the world than we actually do"* when AI provides easy answers.
- **Illusion of Explanatory Depth:** AI can produce convincing explanations or predictions, leading scientists to overestimate their grasp of the underlying phenomena. (E.g., a highly accurate AI model isn't the same as a true causal understanding, but we might conflate the two.)
- **Monoculture Risk:** If everyone uses the same AI tools, we might all chase similar ideas, **narrowing diversity** of research questions and methods. Science could become an "AI-shaped" monoculture, vulnerable to systematic blind spots.
  - Like a single crop in a field, it's efficient but dangerous if a pest (or flaw) comes along.
- **Error Propagation:** AI can confidently output incorrect info. If unchecked, these errors can propagate into research (bad citations, flawed analysis code, etc.)
- **Bias and Fairness:** AI systems inherit biases from training data. If used naively, they can reinforce stereotypes or exclude minority viewpoints, skewing research outcomes.
  - Responsible use of AI means double-checking facts, maintaining transparency, and remembering that correlation (or prediction) is not causation. We produce more, but must ensure we don't *understand less*.

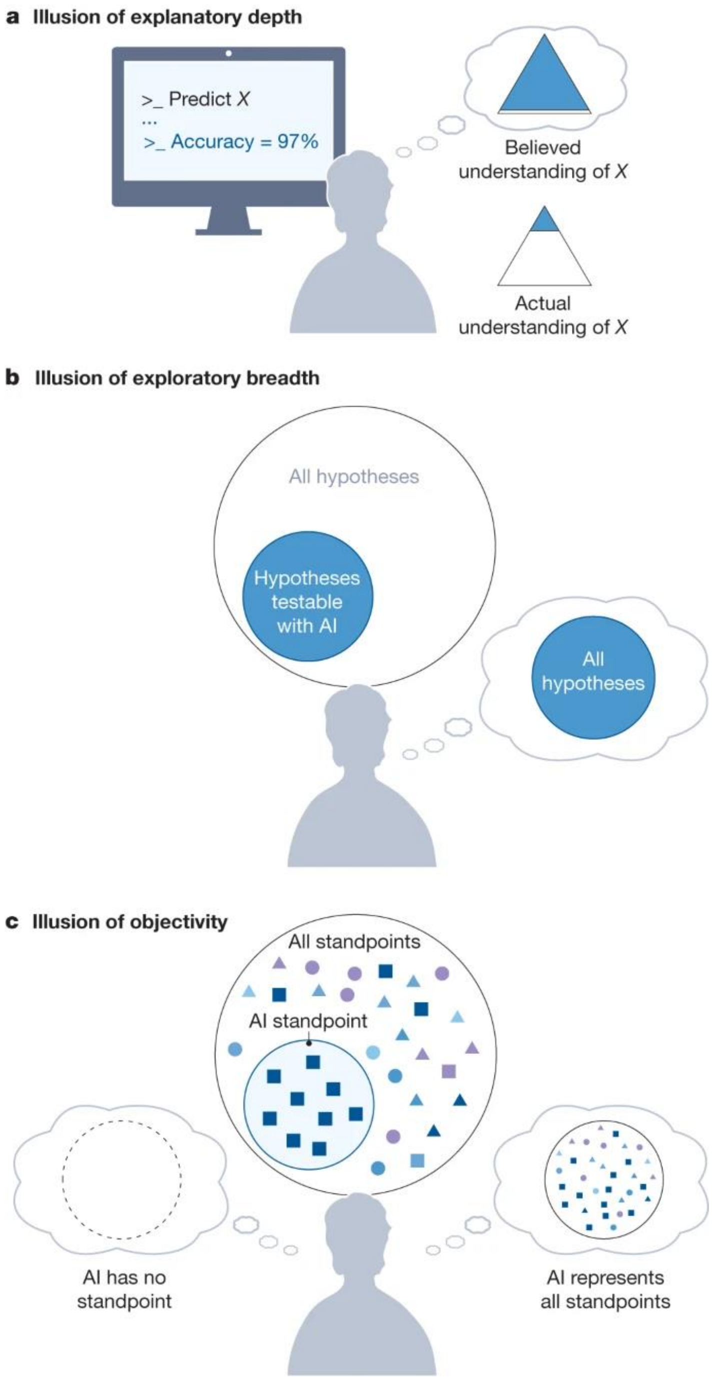
Illustrations of "illusions of understanding" in AI-assisted research (Messeri & Crockett, 2024):

(a) *Illusion of explanatory depth*: Researchers see an AI model achieve high accuracy and **believe** they fully understand phenomenon X (depicted by a large triangle in thought bubble), whereas their **actual** understanding of X is much smaller (small triangle).

(b) *Illusion of exploratory breadth*: Scientists focus only on hypotheses that AI readily suggests (big blue circle) thinking they cover all possibilities, while neglecting hypotheses outside that AI-generated set.

(c) *Illusion of objectivity*: One might assume AI represents all viewpoints (thinking the AI's outputs cover a wide range of perspectives), but in reality it may reflect a narrower range (many similar points in a smaller circle), missing diverse standpoints.

💡 *See paper at: https://doi.org/10.1038/s41586-024-07146-0*

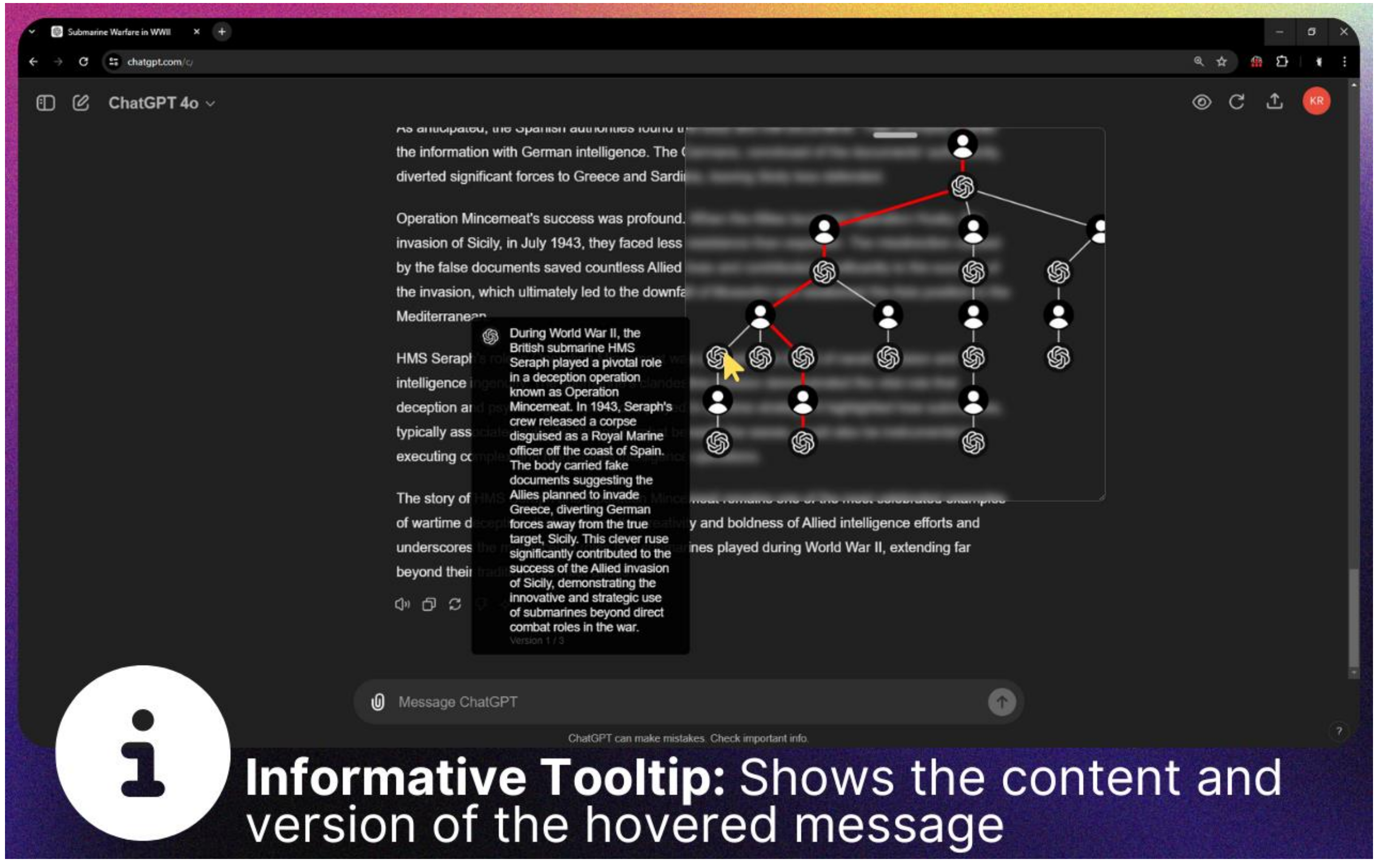


Ethical and Practical Challenges

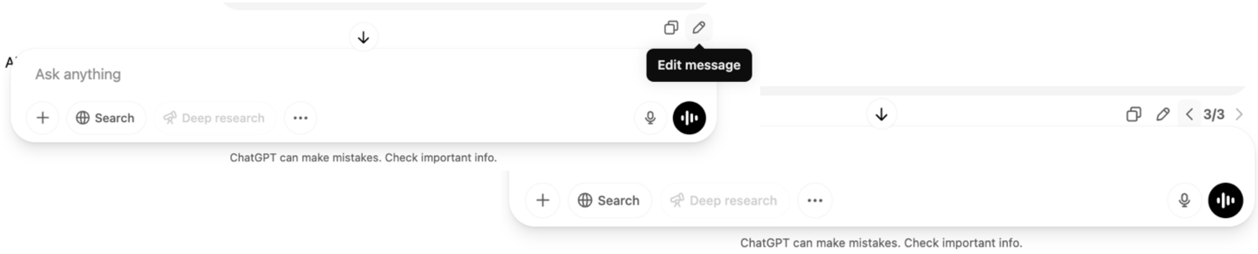
- **Data Privacy:** Be careful when inputting unpublished data or sensitive information into AI tools – many cloud-based AIs might store or learn from it – e.g., remove PII like names, addresses, contact details
- **Academic Integrity:** Using AI on assignments or papers can blur authorship – maintain honesty about who (or what) produced content to avoid misconduct
  - **Disclosure:** If AI contributed text or edits, many journals require acknowledging its role (to maintain transparency)
  - **Plagiarism:** Don't present AI-written content *verbatim* as your own without checking – treat it like a ghostwriter whose output you must verify and edit
  - **Citation Integrity:** Ensure every factual claim still has a valid reference. If AI suggests a citation, double-check its existence and relevance
- **Reproducibility:** If AI is used in analysis, ensure others can reproduce the steps (e.g., save the prompts and/or code, not just "the AI did it")
- **Dependence:** Develop your own skills too. What if the AI tool is unavailable or gives a wrong answer? You should know how to proceed without it
- **Human Oversight:** Use AI to assist, but a human (you) must have final say to preserve scholarly standards and originality
- **Continual Learning:** The AI landscape changes fast (new models, updated capabilities). Commit to ongoing learning to use AI effectively and safely

Good Practices for AI-augmented Research

- **Stay Critical:** Always question and verify AI outputs. Treat AI suggestions as hypotheses or drafts, not final truths
- **Document Usage:** Keep track of when and how you used AI in your workflow (prompts, tools, versions, dates, etc.) – aids transparency and reproducibility
- **Iterative Approach:** Use AI in *loops* – e.g., draft with AI, then refine yourself, then maybe ask AI to check, and so on. This interplay yields the best results.



https://microsoftedge.microsoft.com/addons/detail/chatgpt-conversation-tree/



Future Outlook – Growing Role in Research

- **AI Across the Pipeline:** Researchers envision AI as: **Oracle** (literature guru), **Surrogate** (data generator), **Quant** (master analyst), and **Arbiter** (objective evaluator) in different research stages. These roles could cover everything from hypothesis generation to data collection and analysis.
- **Augmented Intelligence:** Next-gen LLMs will be *tool-using agents* – able to run code, query databases, and perform multi-step reasoning autonomously. This could address current limitations (improving consistency, interpretability, and scalability of AI outputs).
- **AI Lab Partners:** We may see AI "colleagues" that design and even execute experiments in simulation (e.g., a chemistry AI that proposes and runs virtual syntheses, or an economics AI that simulates markets). Human researchers will supervise and provide creative direction, but much of the grunt work could be offloaded.
- **Productivity vs. Discovery:** If used well, cognitive automation might free scientists to tackle more ambitious questions – a boost to discovery. Korinek (2023) speculates that ongoing advances will indeed *"improve performance of LLMs across all domains,"* and researchers who harness them will become far more productive. The challenge is ensuring this productivity translates to genuine understanding and innovation, not just more papers.

Future Concepts – "Digital Behavioral Twins"

- **Digital Behavioral Twins:** A forward-looking concept where an AI model serves as a *digital replica of a person's behavior*. For researchers, this could mean having a virtual twin of a participant (or even oneself) to run simulations on behavior and decision-making.
- **Applications:** In social science, you might simulate how a population of "digital twins" reacts to a new policy before actually implementing it. In psychology or health, a patient's twin could help predict treatment outcomes. This goes beyond current "synthetic data" by aiming to capture individual-level complexity and dynamics.
- **Status:** Still largely speculative – related work in other fields (e.g., autonomous vehicles using driver behavior models shows feasibility of modeling complex behavior). Achieving true digital twins of humans raises issues of accuracy and consent.
- **AI Integration:** The concept ties in with AI being deeply integrated in the research cycle – perhaps one day, instead of recruiting 100 human participants for a pilot, you run 100 AI-based behavioral twins first to identify the most interesting conditions to test in the real world.

Before we finish…

Any questions?

Get in touch!

Dr Steve Bickley
s.bickley@qut.edu.au