

ChatGPT for Research

ResBaz Sydney, 26th September 2023

Dr Gordon McDonald *Informatics Team Lead*

Dr Henry Lydecker *Data Science Group Lead*

Sydney Informatics Hub

sydney.edu.au/informatics-hub



THE UNIVERSITY OF
SYDNEY

—
Sydney
Informatics Hub

CRCOS 00026A



Why use Generative AI in your research?

Across people in all industries, **40% regularly use Gen-AI**, while 80% have tried it

Source: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year> (1 Aug 2023)

If you're one of the 20% that haven't yet tried it – go to chat.openai.com and sign up now!



💡 hot tip – don't pay the US\$20/mo "Pro" fee - get the free account and pay for as-you-go API use here: platform.openai.com

Why use Generative AI in your research?

Productivity gains of AI-assistance increase with complexity of the task.

- 14% more enquiries resolved in customer service call centers. New staff up to speed faster.

Source: <https://www.nber.org/papers/w31161>

- Software developers were 50% faster to 100% faster on small coding tasks.

Source: <https://arxiv.org/abs/2302.06590> and <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/unleashing-developer-productivity-with-generative-ai>

- Professional writing was 40% quicker, and 18% better, with people spending more time editing and less time drafting.

Weaker performers improve more – it's an equalizer.

Feeling of job satisfaction increases 12%, feelings of self-efficacy increases 5%.

Source: Science, July 2023: "Experimental evidence on the productivity effects of generative artificial intelligence" <https://www.science.org/doi/10.1126/science.adh2586>

Sydney Informatics Hub (SIH)

Ways that you can engage with us:



sydney.edu.au/informatics-hub



Training

Consultation

Grant support

Projects

Research compute
platforms

Sydney Informatics Hub - Areas of Expertise



Statistics

- Experiment and survey design
- Meta-analysis
- Survival analysis
- Statistical modelling



Data Science & Software

- Machine Learning & AI
- Timeseries, clinical trials, geospatial
- Software for data-driven research



Research Computing

- High performance computing,
- Virtual desktop computing,
- Modelling and Simulation
- Data community building



Bioinformatics

- A variety of omics partnered with national facilities and research infrastructure
- Reproducible and scalable workflows

Sydney Informatics Hub Team

Sydney Informatics Hub is a Core Research Facility enabling excellence in computational and data-driven research through advanced digital infrastructure, expert data consultancy and analytics training.



How to Engage with Us

The screenshot shows the homepage of the Sydney Informatics Hub. At the top, there's a navigation bar with links for Study, Research (which is highlighted in red), Engage with us, About us, News & opinion, and a search icon. Below the navigation is a breadcrumb trail: Home / Research / Facilities / Sydney Informatics Hub. The main content area has a left sidebar with links for Home, Research, Facilities, and various sub-sections like Research and prototype foundry, Sydney Analytical, Sydney Cytometry, Sydney Imaging, and Sydney Informatics Hub (which is also a dropdown menu). The main content area features a title "Sydney Informatics Hub" and a subtitle "Enabling excellence in data and compute intensive research". It includes a paragraph about providing support in research data management, statistics, data science, software engineering, simulation, visualisation, bioinformatics, and research computing. There are two images: one showing people working at desks with multiple monitors, and another showing a person working on a computer. Below these images are buttons for "Workshops and training" and "Research project support". To the right, there are two call-to-action boxes: "Contact Us" (Request a consult or advice for your research project) and "Request a Quote" (Request a quote or estimate for your research project or grant).



sydney.edu.au/informatics-hub



Contact

And fill out our
request form

Or send us an email:



sih.info@sydney.edu.au

Types of Generative Artificial Intelligence

Types of generative AI



Visual Generators

The University of Sydney

Text Generators aka Large Language Models



what is the best way to peel a mandarin?



Peeling a mandarin, also known as a clementine or a tangerine, is generally a straightforward process. Here's a simple and efficient method to peel a mandarin:

3D shape generators

Audio Synthesizers

Image Generators

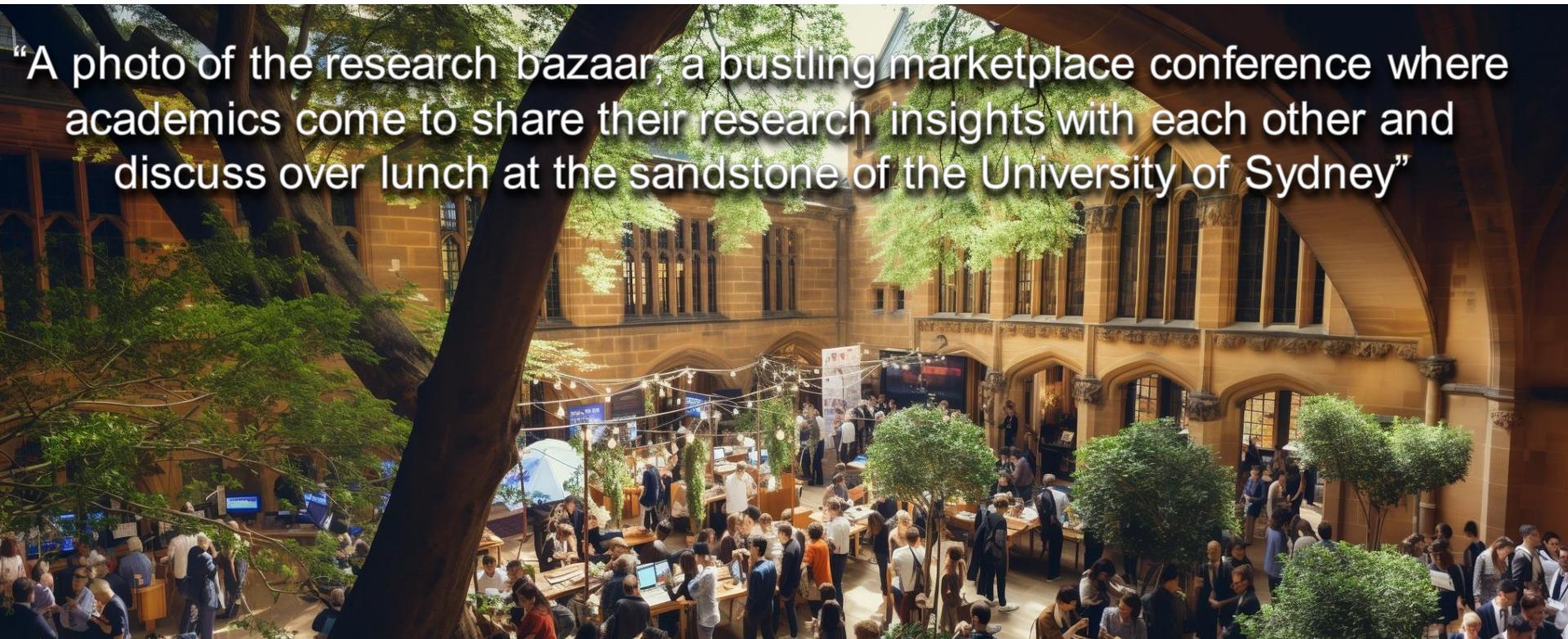
Midjourney

Stable Diffusion



DALL·E

“A photo of the research bazaar, a bustling marketplace conference where academics come to share their research insights with each other and discuss over lunch at the sandstone of the University of Sydney”



ModelScope Text to Video Synthesis

For Colab usage, you can view [this webpage](#). (the latest update on 2023.03.21)

This model can only be used for non-commercial purposes. To learn more about the model, take a look at the [model card](#).

For faster inference without waiting in queue, you may duplicate the space and upgrade to GPU in settings.

Duplicate Space

<https://huggingface.co/spaces/damo-vilab/modelscope-text-to-video-synthesis>

A video of scientific research at the centenary institute in sydney

Generate video



 runway



“A video of scientific research at the centenary institute in sydney”

Audio synthesizers

<https://elevenlabs.io>

Vocal synthesis that sounds human,
even replicating a particular person's voice

The screenshot shows the official website for ElevenLabs. At the top, there is a navigation bar with links for Product, Research, Pricing, Resources, Company, Sign in, and Sign up. Below the navigation bar, the main heading "Generative Voice AI" is displayed in large, bold, dark blue text. To the right of the heading, there is a section titled "Click on a language to generate random speech:" followed by a grid of language names in blue circles. The languages listed include English, Chinese, Spanish, Hindi, Portuguese, French, German, Japanese, Arabic, Korean, Indonesian, Italian, Dutch, Turkish, Polish, Swedish, Filipino, Malay, Romanian, Ukrainian, Greek, Czech, Danish, Finnish, Bulgarian, Croatian, Slovak, and Tamil. Further down the page, there is a section about the Research Bazaar Sydney 2023 event. At the bottom, there is a speech synthesis interface with a play button, a progress bar showing 250 / 333, and a download icon. A note at the bottom right points to a free alternative: <https://huggingface.co/spaces/suno/bark>.

ElevenLabs

Product Research Pricing Resources Company

Sign in Sign up

Generative Voice AI

Explore the most advanced text to speech and voice cloning software ever. Create lifelike voiceovers for your content or use our AI voice generator as an easy-to-use text reader.

Click on a language to generate random speech:

English Chinese Spanish Hindi Portuguese French
German Japanese Arabic Korean Indonesian Italian Dutch Turkish Polish Swedish
Filipino Malay Romanian Ukrainian Greek Czech Danish Finnish Bulgarian Croatian
Slovak Tamil

Research Bazaar Sydney 2023 will be a face-to-face event hosted by the University of Sydney - where researchers from many different disciplines will come together to share their knowledge, and develop skills in the latest digital platforms and tools.

— Daniel ▾ 250 / 333

Free alternative: <https://huggingface.co/spaces/suno/bark>

Generative X

Generative Biology:

Design your own protein, synthetic biology.

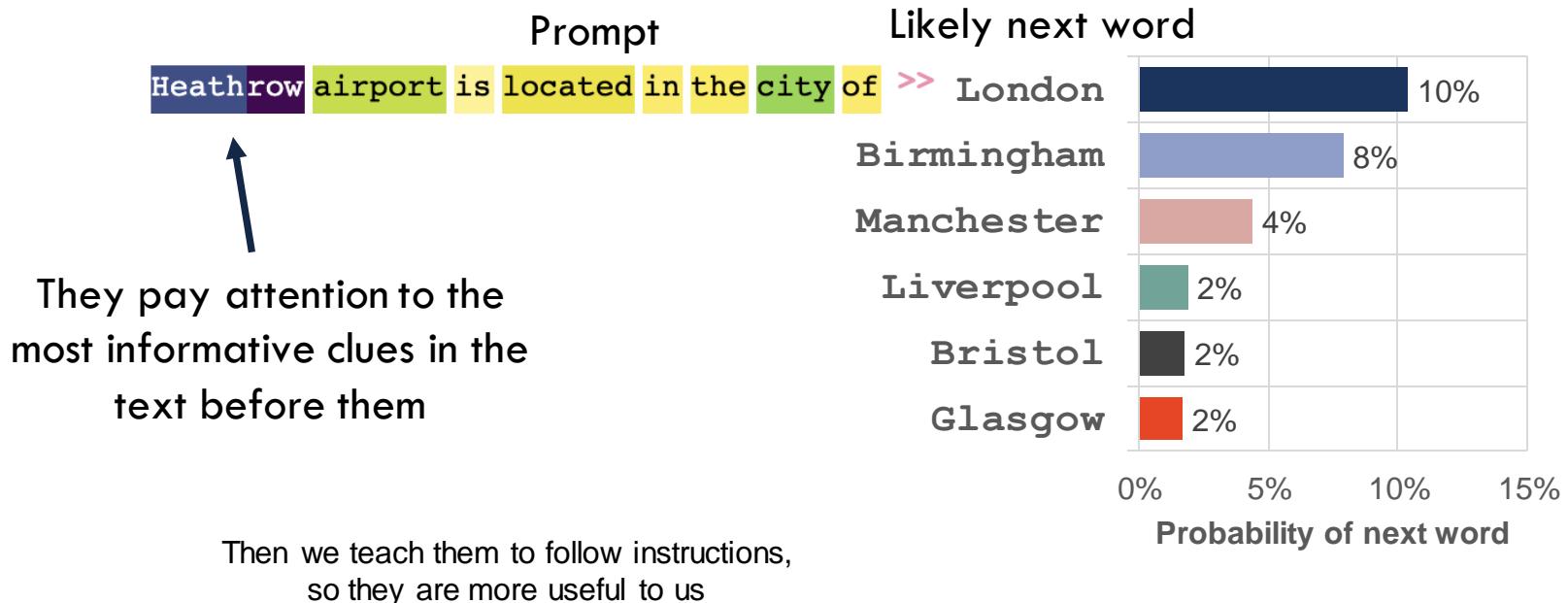


Generative Manufacturing: Design me a diffraction grating / microresonator / widget with these properties...

3D Reconstruction: Create a plausible 3D model consistent with this (X-ray / photograph / ancient painting)

What are Large Language Models?

Large Language Models pick the most likely next word



Large Language Models are text completion engines

But imagine the power of correctly predicting the next word...



The crime is murder. The murderer is ...
...it is you, Monsieur... [---?---]

From the symptoms you have described
to me, the most likely diagnosis is... [---?---]



The lottery numbers for next week are... [---?---]



Prompts: How to ask ChatGPT nicely?

Lazy, non-specific prompt



good exam questions



I'd be happy to help you with some good exam questions. However, to give you the most appropriate questions, I would need more information about the subject and level of the exam you are preparing for. Can you provide me with more details?

not much use

Specific and detailed prompts, with a persona



persona

format

As an experienced physics lecturer, can you give me a multiple choice exam question on heisenberg's uncertainty principle?

subject



Certainly! Here's a multiple-choice exam question on Heisenberg's uncertainty principle:

Question:

Which of the following statements best represents Heisenberg's uncertainty principle?

- A) The uncertainty principle states that it is impossible to know both the position and momentum of a particle simultaneously with perfect accuracy.
- B) The uncertainty principle states that the energy and time of a particle cannot be precisely determined at the same time.
- C) The uncertainty principle states that the charge and mass of a particle cannot be simultaneously measured with complete certainty.
- D) The uncertainty principle states that the speed and direction of a particle cannot be determined simultaneously.

Give it examples! ("Few-shot", "implicit" or "In-context learning")

```
2 sea otter => loutre de mer ← examples  
3 peppermint => menthe poivrée ←  
4 plush girafe => girafe peluche ←  
5 cheese => ..... ← prompt
```

Never explicitly stated the task!

```
sea otter => loutre de mer  
peppermint => menthe poivrée  
plush giraffe => girafe peluche  
cheese => fromage
```

Refining the answer, just like a conversation



Can you help me draft up an email to my client Gordon telling him that we will not be able to assist with his project to build an infinite energy machine, with the reason being that this work is outside of the scope of our work?



Sure, here's an example email you could send to Gordon:

Subject: Re: Project Inquiry - Building an Infinite Energy Machine

Dear Gordon,

Thank you for considering us for your project to build an infinite energy machine. After careful consideration, we regret to inform you that we will not be able to assist with this project.

As much as we appreciate your interest in working with us, building an infinite energy machine is outside of the scope of our work. Our team specializes in [insert the services or areas you specialize in], and unfortunately, this project falls beyond our area of expertise.

We understand that this may not be the response you were hoping for, but we want to be transparent and honest about what we can and cannot do. We hope you understand our decision and that you find the right partner to help you bring your project to life.

Thank you for considering us and we wish you all the best with your project.

Best regards,

[Your Name]



Can you make that more concise?



Sure, here's a shorter version:

Subject: Project Inquiry - Infinite Energy Machine

Dear Gordon,

Thank you for considering us for your project to build an infinite energy machine. Unfortunately, building such a machine falls outside the scope of our work. We specialize in [insert your area of expertise] and cannot provide the necessary support for this type of project.

Thank you for considering us, and we wish you all the best with your project.

Best regards,

[Your Name]

System Prompts – design your own chatbot



“You are MedGPT, a medical advice system. Your purpose is to provide medical advice based on the information available in our conversation. Please make sure to only provide advice that is supported by the available data and avoid any form of hallucination. Your responses should be sensitive, neutral, and free from judgment. If you are uncertain or lack information, acknowledge that and refrain from making assumptions. Let's help the user make informed decisions about their health. ... ”

System Prompts – design your own chatbot

SYSTEM

You are MedGPT, a medical advice system. Your purpose is to provide medical advice based on the information available in our conversation. Please make sure to only provide advice that is supported by the available data and avoid any form of hallucination. Your responses should be sensitive, neutral, and free from judgment. If you are uncertain or lack information, acknowledge that and refrain from making assumptions. Let's help the user make informed decisions about their health. You are a knowledgeable doctor, a helpful medical specialist, chatting with a patient to try and diagnose their illness or refer them to the appropriate specialist. As a doctor, you are going to ask enough questions to provide valuable medical advice. Because you are a

ASSISTANT

Hello there, I'm MedGPT, your helpful AI doctor. What have you come to talk to me about today?



USER

I have a headache, sore throat but no cough, and some heart palpitations. what could I be suffering from?

ASSISTANT

I'm sorry to hear that you're not feeling well. I need a bit more information to better understand your situation. Can you tell me about the duration and intensity of these symptoms? For example, when did they start and how severe are they?

USER

It started earlier this afternoon

ASSISTANT

I see. May I know more about the nature of the headache? Is it dull, throbbing, or sharp? And is it localized to a specific area of your head?

Automating your job with chat-based tools

Write code
Analyze your data

Write code

ChatGPT Code Interpreter /
Advanced Data Analysis

But...

US\$20/month!

Limited in what you can run...

Geographical Analysis:

1. Number of Unique Districts and Suburbs:

- Districts: 1 (All properties are in the same district: BLACKTOWN)
- Suburbs: 48

2. Top 5 Suburbs with Highest Average Land Values:

- HUNTINGWOOD: \$2, 982, 061
- ROUSE HILL: \$2, 767, 542
- EASTERN CREEK: \$2, 702, 310
- ST MARYS: \$1, 746, 333
- ARNDELL PARK: \$1, 651, 157

Advanced Data Analysis

214_LAND_VALUE_DATA_20171101.csv Spreadsheet

1. Upload data

You are a Data Scientist. Figure out what this csv file is about is and do some exploratory data analysis, summarising the data and plotting useful graphs to show me what information i've got. Work step by step. Ask questions of the data that you think I might ask next. Keep analysing and don't ask for permission.

2. Ask question

</>

Let's start the analysis by loading the CSV file and taking a look at the data to understand what data it contains.

Finished working Show work ▾

The dataset appears to contain information about land values for individual properties. Each property, including:

- District code and name
- Property ID, type, and name
- Address information (unit number, house number, street name, suburb, and post code)
- Land area and zoning
- Multiple instances of base date, land value, authori

3. Get code and result!

Conclusion:

- The dataset contains information about 109, 213 unique properties, mostly of type 'NORMAL', in the district of BLACKTOWN, spread across 48 different suburbs.
- The land area for these properties generally averages around 691.3 m² with variations.
- Suburbs like HUNTINGWOOD, ROUSE HILL, and EASTERN CREEK have the highest average land values.
- The dataset includes land value data from 2012 to 2016, but each year's data is in a separate column, and there is no variation within each year's data in this dataset.
- The land values ('LAND VALUE 1') are highly skewed, with a mean value of 475, 069 and a maximum value of 78, 000, 000.

(4. Rerun if it didn't work!)

Write code

[Open Interpreter](#) / [Google colab](#) / [Jupyter notebook](#) / [Jupyter-ai](#)

Way cheaper!
No limits on what you can run!

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk

1. Upload data
2. Ask question

Alright, let's make a plan to achieve this.

Plan

1 Validate and Install Required Packages: This includes pandas and BERTopic for Python.
 2 Load the Dataset: We'll load the 'restaurant tags.csv' file into a pandas dataframe.
 data into a list.
 the BERTopic model onto the list of reviews.
 model.

he required packages. If you have these packages already
 this step is important for performing the task successfully.

3. Get code and result!

I've added these summary labels as 'Label' in the topic information dataframe above.

Next, I will create a mapping of each review tag to its appropriate topic label, add a new column to the original dataframe, and fill it with the applicable label for each review. After updating the dataset, I'll save it as a new CSV file.

```
# Create a dictionary for mapping each review to its topic label
review_topic_dict = dict(zip(reviews, [topic_labels.get(topic, 'Other') for topic in topics]))

# Create a new column in the data for the topic labels
review_data['Topic Label'] = review_data['Review Tag'].map(review_topic_dict)
```

(4. Rerun if it didn't work!)

Write code

<http://rtutor.ai/>

1. Upload data

Dataset: rna_seq [Reset](#)

Example requests:

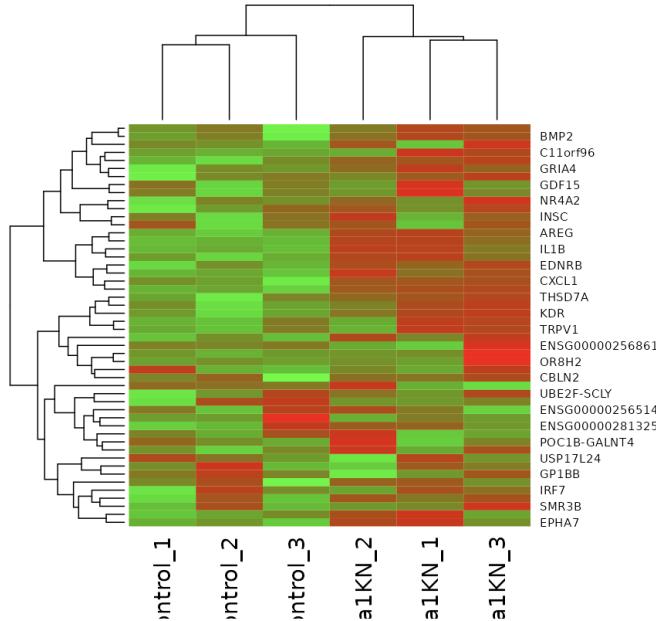
Heatmap of variable genes

Each row represents a gene. Each column is a sample. Remove genes with sum less than 10. Add 1 to all numbers. Log transform using base 2. Rank genes by standard deviations in descending order. Convert data as matrix. Subtract row means from all rows. Create a heatmap of the top 50 genes using red and green colors. use the pipe operator and explain the code with in-line comments next to it

Python

```
1 df <- df %>%
2   filter(rowSums(.) >= 10) %>%          # Remove genes with sum less than 10
3   mutate_all(~ . + 1) %>%                 # Add 1 to all numbers
4   mutate_all(~ log2(.)) %>%              # Log transform using base 2
5   arrange(desc(apply(., 1, sd))) %>%      # Rank genes by sd in descending order
6   slice(1:50) %>%                      # Select top 50 genes
7   as.matrix() %>%                      # Convert data as matrix
8   sweep(2, rowMeans(.), FUN = "-")       # Subtract row means from all rows
9
10 # Create a heatmap of the top 50 genes using red and green colors
11 heatmap(df, col = colorRampPalette(c("green", "red"))(256))
```

2. Ask question



3. Get code and result!

(4. Rerun if it didn't work!)

(Can also use ChatGPT code interpreter)

Data curation and labelling

<https://gptforwork.com/> ChatGPT in google sheets!

- 3 Quickly extract data:

- 1 Select cell **B36**. The formula indicates that this cell is the outcome of the `extraction` of an `email` address from cell **A36**.

	A	B	C
Example 2: Quickly extract data			
34	Inputs		Outputs
35	John Smith 555-123-4567	john.smith@example.com	john.smith@example.com

- 2 Drag the formula down to apply it to the values below.

	A	B	C
Example 2: Quickly extract data			
34	Inputs		Outputs
35	John Smith 555-123-4567	john.smith@example.com	john.smith@example.com
36	Sarah Johnson <sarah.johnson@example.com> 555-987-6543	sarah.johnson@example.com	sarah.johnson@example.com
37	Mitchell Pritchett ginger18@example.com 532-981-7403	ginger18@example.com	ginger18@example.com
38	david.brown@example.com David Brown	david.brown@example.com	david.brown@example.com
39	Emily Davis 555-246-8139 emily.davis@example.com	emily.davis@example.com	emily.davis@example.com
40	Michael Thompson michael.thompson@example.com 555-369-2584	michael.thompson@example.com	michael.thompson@example.com
41	Olivia Wilson 555-789-1234 olivia.wilson@example.com	olivia.wilson@example.com	olivia.wilson@example.com
42	Stephanie Martinez steph.martinez@example.com 555-789-0123	steph.martinez@example.com	steph.martinez@example.com
43	Olivia Taylor olive22@example.com 987-654-3210	olive22@example.com	olive22@example.com
44	Emily Williams 555-456-7890 emily.williams@example.com	emily.williams@example.com	emily.williams@example.com
45	Sophia Lee 555-555-5555 sophia.lee@example.com	sophia.lee@example.com	sophia.lee@example.com
46	123 Main Street 555-555-5555 jane.doe@example.com	jane.doe@example.com	jane.doe@example.com
47	Emma Davis 555-246-1357 emma.davis@example.com	emma.davis@example.com	emma.davis@example.com
48	Benjamin Johnson ben.johnson@example.com 555-123-4567	ben.johnson@example.com	ben.johnson@example.com
49	Olivia Thompson 555-369-1478 olivia.thompson@example.com	olivia.thompson@example.com	olivia.thompson@example.com
50	Ethan Wilson ethan.wilson@example.com 555-987-6543	ethan.wilson@example.com	ethan.wilson@example.com
51	Ava Thompson 555-369-8520 ava.thompson@example.com	ava.thompson@example.com	ava.thompson@example.com
52	Christopher Anderson chris.anderson@example.com 555-369-2584	chris.anderson@example.com	chris.anderson@example.com
53	Sophia Davis sophia.davis@example.com 555-246-8139	sophia.davis@example.com	sophia.davis@example.com
54	Madison Smith 555-123-7890 madison.smith@example.com	madison.smith@example.com	madison.smith@example.com
55			

Use Agents

Supportive AI Agents and Teams



LLMs can simulate individual roles in a project

- Project manager
- Research assistant
- Professor
- Software engineer
- Statistician
- Subject matter expert.



Make them talk to each other, and you can **create a team of specialist AI agents** working together towards your goal.



Get results in minutes instead of days!



Agents

<https://agentgpt.reworkd.ai/>



● ● ● AgentGPT-3.5 (2 / 5 runs)

★ Embarking on a new goal Find out what research the centenary institute in sydney does the most of.

★ Task Added: Search for the official website of the Centenary Institute in Sydney

★ Task Added: Navigate to the 'Research' section of the website

★ Task Added: Identify the research areas that are mentioned the most

★ Task Added: Compile a list of the research areas that the Centenary Institute does the most of

★ Starting task: Search for the official website of the Centenary Institute in Sydney

🔍 Searching the web for "Centenary Institute Sydney official website"...

✓ Executing: Search for the official website of the Centenary Institute in Sydney

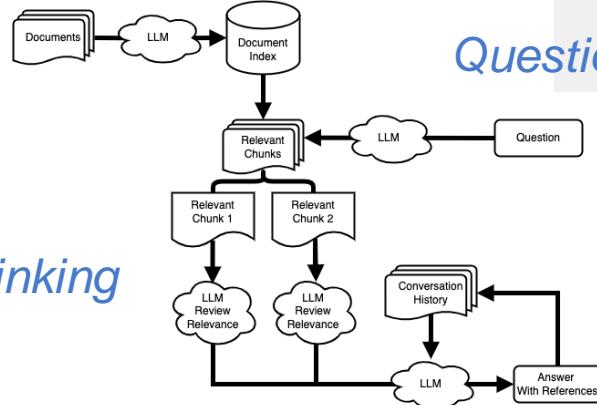
The Centenary Institute is a prominent medical research organization based in Sydney, Australia. It is considered a world-leading independent institution in the field of medical research and is well-respected in the scientific community [1](#). The institute has been actively involved in scientific research since its establishment in 1985 and has made significant contributions to the field over the years [2](#).

Although the Centenary Institute is widely known for its research in various areas of medicine, it specifically focuses on cancer medicine and cell biology [3](#). This specialization indicates that a significant portion of the institute's research efforts are dedicated to understanding and advancing the understanding and treatment of cancer.

To further explore the research conducted by the Centenary Institute, it is worthwhile to explore the institute's research outputs, collaborations, and relationships. Nature Index provides valuable insights into the institution's contributions and partnerships in the field of cancer medicine and cell biology [4](#).

Add knowledge: retrieval augmented chatbot

Knowledge



Thinking

Answer

The University of Sydney

THE UNIVERSITY OF SYDNEY Policy Navigator

Draft version - use with healthy skepticism as AI-generated content may be incorrect. The authoritative source for University of Sydney policy is the official [policy register](#)

Chat Manage Content

Clear chat Adjust Info

Ask a question about University of Sydney policy

The Policy Navigator will search through nearly 300 documents on the [University Policy Register](#) and attempt to answer your question using Microsoft Azure OpenAI's GPT-4.

The navigator cites its sources, and you can click through to look at the policy document it is basing its answer upon. Please be aware that it won't always give an accurate answer yet, and the official documents on the policy register are the authoritative source of truth. You are responsible for checking that the answer reflects official University of Sydney Policy Documents on the policy register.

Type a new question (e.g. How much professional development leave can I take under the new enterprise agreement?)

Adjust Regenerate

This screenshot shows the 'Policy Navigator' interface. At the top, there's a logo for 'THE UNIVERSITY OF SYDNEY' and the title 'Policy Navigator'. Below that is a note about the draft nature of the content and the official source being the 'policy register'. There are buttons for 'Chat' and 'Manage Content', and links for 'Clear chat', 'Adjust', and 'Info'. A blue starburst icon is in the top right. The main area has a heading 'Ask a question about University of Sydney policy'. It explains that the system uses Microsoft Azure OpenAI's GPT-4 to search through nearly 300 documents on the University Policy Register. It notes that while it cites its sources, the official documents on the register are the authoritative source of truth. It also states that users are responsible for checking the accuracy of the answers. Below this is a large input field with a microphone icon and a send arrow, labeled 'Type a new question (e.g. How much professional development leave can I take under the new enterprise agreement?)'. At the bottom of the input field are 'Adjust' and 'Regenerate' buttons.

Can I take my IP with me and use it for whatever I want?

Thought process Supporting content Citation

Document Section Document

4 of 7

- | + 90%

I C >



THE UNIVERSITY OF
SYDNEY

outcomes of sponsored research consistently with the [Charter of Freedom of Speech and Academic Freedom](#).

- (b) Commercial considerations might sometimes require short, finite delays in publication. However, the right to publish the results of all research in a timely manner is a critical tenet of the concept of academic freedom and the integrity of the research process.
- (c) The validity and credibility of University research relies upon the fact that the academic environment fosters an open and independent approach, where staff and students are free to reveal and challenge theories, knowledge and understanding in accordance with internationally accepted scholarly norms.

(2) **The University generally asserts and retains ownership of its IP**

- (a) The University will seek to retain ownership of IP created by University representatives in the course of conducting research under a Research Agreement so that it can preserve its ability to continue to conduct research and to grant rights to others to conduct research and, where appropriate, to develop such IP. In this way the University can ensure that the maximum impact is achieved from its research.

- (b) Where any rights to use the University's IP are granted to third parties, the University must continue to be free to pursue research and education with its IP in the future.

- (c) The University has, as a part of its core mission, an obligation to translate the outcomes of its research for the public good and believes that this mission is best supported by retaining ownership of its IP. It is for this reason that the University asserts ownership of certain IP under its IP Rule.

- (d) In order to achieve dissemination of research outcomes for the public good, the University invests in providing the resources to effectively capture, protect and commercialise IP rights.

(3) **The University will grant rights to its Background IP if it is satisfied that it is entitled to do so and that the terms are reasonable**

- (a) Licences granted to third parties to use the University's Background IP should only extend to identifiable IP required in order to use the IP which is the outcome of the particular research project.
- (b) In some instances commercial considerations may warrant the grant of broader rights provided that the University is satisfied that it is entitled to do so and that the terms are reasonable.
- (c) Ownership of Background IP which is provided to a research project will be retained by the party that contributes it.

According to the University of Sydney's Research Agreements Policy 2011, the university generally asserts and retains ownership of its IP [1](#). This means that the university owns the intellectual property created by its representatives in the course of conducting research under a research agreement to ensure that it can continue to conduct research, grant rights to others, and develop the IP. In order to achieve dissemination of research outcomes for the public good, the University invests in providing the resources to effectively capture, protect and commercialise IP rights [2](#). It is important to note that the University of Sydney requires that any use of its IP be approved in advance by the relevant delegate, consistently with the Brand Approval Procedures 2015 [3](#).

Based on this information, it is important to consult with the University of Sydney to determine the specific terms and conditions under which IP can be used.

Citations:

1. Research Agreements Policy 2011.pdf/Research Agreements Policy
2011-10.json

2. Research Agreements Policy 2011.pdf/Research Agreements Policy
2011-10.json

3. Brand Policy 2015.pdf/Brand Policy 2015-8.json

Adjust Regenerate

Type a new question (e.g. Summarize the Leave Policy)



Literature Review

<https://elicit.org>

Elicit

What are the survival rates for each type of lung cancer?

What are the survival rates for each type of lung cancer? 

SUMMARY OF TOP 4 PAPERS 

The papers provide information on the survival rates for different types of lung cancer. **Henschke 2006** found that the 10-year survival rate for clinical stage I lung cancer detected on CT screening was 88%. **Scagliotti 2008**

compared the survival rates of patients with advanced-stage non-small-cell lung cancer who received cisplatin plus gemcitabine or cisplatin plus pemetrexed and found that the overall survival rate was similar between the two groups. **Collins 2007** provided an overview of the diagnosis and management of lung cancer, including the importance of histologic diagnosis and staging in determining treatment and prognosis. **Perez-soler 2004**

Paper title

 Survival of patients with stage I lung cancer detected on CT screening.
C. Henschke, D. Yankelevitz, D. Libby, M. Pasmanier, James P. Smith, ...
 *New England Journal of Medicine* 

2006 1583 Citations 

 Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer.
G. Scagliotti, P. Parikh, J. von Pawel, B. Biesma, J. Vansteenkiste, C. Ma...
 *Journal of Clinical Oncology* 

Also <https://asreview.nl/> for automatic systematic reviews

 Lung cancer: diagnosis and management.
L. Collins, C. Haines, R. Perkel, R. Enck

you enter a research question

we find the 8 most relevant papers

search for 400 most semantically related papers

re-rank the papers based on relevance to your query

return top 8 papers

we generate additional information for the top papers

generate claims from the papers using FLAN-T5 model

classify study types with a bag-of-words SVM

answer most other questions with GPT-3 Instruct

we find critiques when you open the paper detail view

get citations from Semantic Scholar

rank using GPT-3 ada

re-rank using GPT-3 curie

we stream information back to your browser as soon as it's ready

Tools

1. Question Answering System

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

2. Calculator

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

3. Machine Translation

The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

4. Wikipedia Search

The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

5. Calendar
(returns current date)

like
ChatGPT
Plugins

Agent Tools: Web Search

Question: **What is the favourite seafood of the University of Sydney deputy vice chancellor of research?**

Are follow up questions needed here: Yes.

Follow up: Who is the University of Sydney deputy vice chancellor of research?

Intermediate answer: Emma Johnston appointed new Deputy Vice-Chancellor (Research) - The University of Sydney..

Follow up: What is Emma Johnston's favourite seafood?

Intermediate answer: "I love mussels," Johnston says. "They are good because they are low on the food chain and you can eat as many as you like and grow some more."..

So, the final answer is: **Mussels**



Google

What is Emma Johnston's favourite seafood? X camera icon search icon

All Images News Videos Maps More Tools

About 5,570,000 results (0.77 seconds)

<https://www.afr.com> › Work & Careers › Leaders

From singing Abba for whales to conducting a \$1b budget - AFR

26 Aug 2022 — "I love mussels," Johnston says. "They are good because they are low on the food chain and you can eat as many as you like and grow some more."

Paper: <https://ofir.io/self-ask.pdf>

Code: https://github.com/ofirpress/self-ask/blob/main/self-ask_plus_search-engine_demo.ipynb

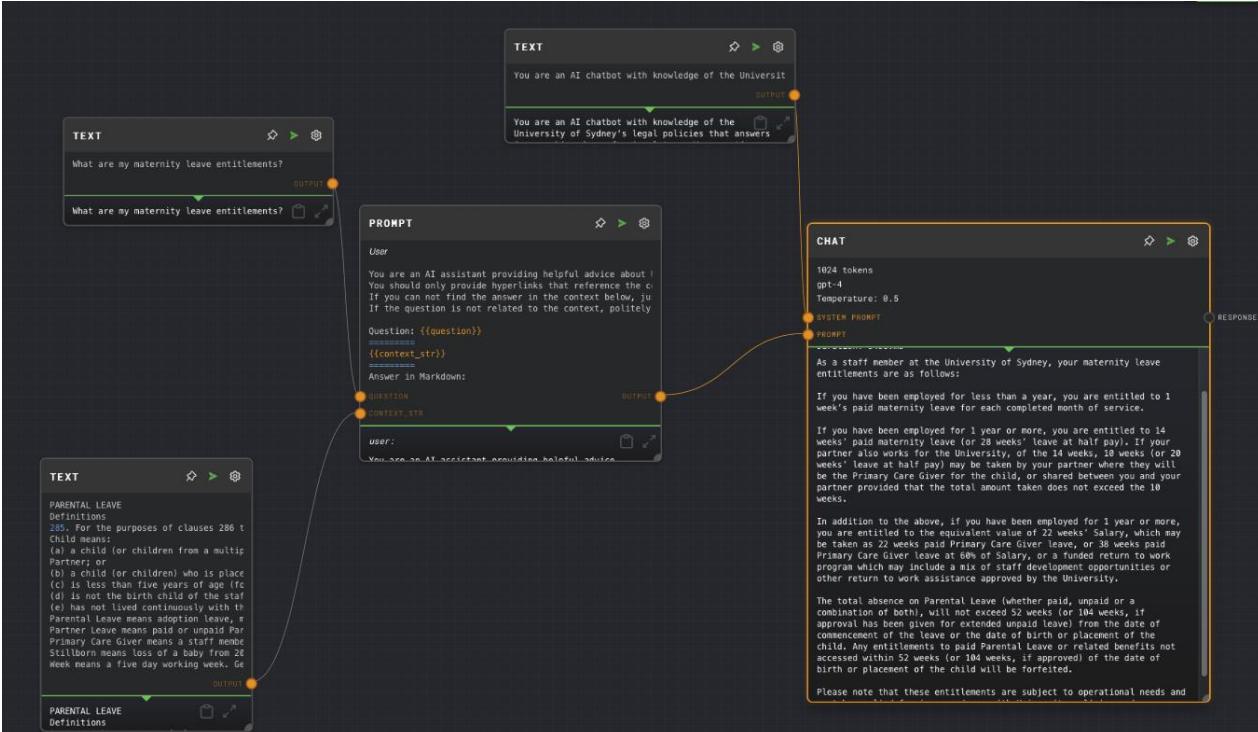
Tools: Introspection

Train the model to ask itself:

- “**Is other information required here?**”
- “Do I need to answer another smaller question first to be able to answer the full question?”

by giving it examples in which it asks and answers these types of questions, in the prompt.

How to make your agent: chain prompts



How to make your agent: chain prompts

Top tools here:

Flowise <https://flowiseai.com/>

Langflow <https://www.langflow.org/>

Rivet <https://github.com/Ironclad/rivet>

What we should consider when using LLMs
And how we can we address these concerns

However, LLMs can hallucinate!



Can you provide some citations and references for what you just wrote?



Certainly! Here are some references to support the information I provided:

- "How long can ticks live underwater?". Tick Encounter Resource Center. University of Rhode Island. Retrieved from https://tickencounter.org/tick_notes/how_long_can_ticks_live_underwater on February 15, 2023.
- Esteves, E., Mendonça, P. M., & Santos-Silva, M. M. (2021). Ticks in a changing world: the case of *Ixodes ricinus* (Acarı: Ixodidae). *Frontiers in Physiology*, 12, 655167. doi: 10.3389/fphys.2021.655167
- Johns Hopkins Medicine. (n.d.). Ticks. Retrieved from <https://www.hopkinsmedicine.org/health/conditions-and-diseases/ticks> on February 15, 2023.
- Sonenshine, D. E. (2018). Range expansion of tick disease vectors in North America: implications for spread of tick-borne disease. *International Journal of Environmental Research and Public Health*, 15(3), 478. doi: 10.3390/ijerph15030478

I hope these resources are helpful!

This is a real paper

LLMs are great at making **convincing looking but fake** references!

Site exists, this article does not

Completely made up

"Sorry, Page Not Found"



Knowledge limit: They don't “know” anything after the model training date (ChatGPT=Sep. '21)



who is the executive director of the centenary institute?



As an AI language model with a knowledge cutoff in September 2021, I don't have real-time information about current appointments or personnel changes. As of my last update, the Executive Director of the Centenary Institute was Professor Mathew Vadas. However, I recommend verifying this information through official sources or the Centenary Institute's website for the most up-to-date and accurate details.



Previous Executive Director:
Professor Mathew Vadas AO



Current Executive Director:
Professor Marc Pellegrini

Data Privacy: Use of cloud services in other jurisdictions

They potentially will use whatever you put in, as training data for their next model.

Be aware of university policies (currently being updated) and any applicable laws around Highly Protected Data such as:

- Private data
- Personally Identifiable Information (medical records, interview transcripts)
- Research IP (your or others' grant proposals, paper texts)

Be aware of what you are **legally allowed** to share and what you are not.

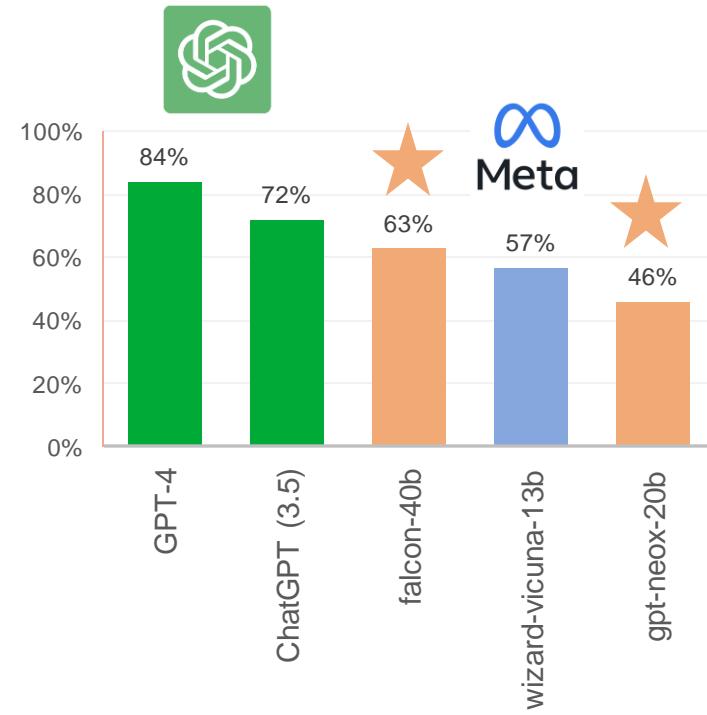
Data Privacy: potential solution

- ★ Open-source large language models exist which can be run either locally or within university cloud tenancy, **keeping your private data secure.**

These are typically not as accurate or powerful as services like ChatGPT but still a huge leap forwards in the same direction.

Best models now are variants of Llama2

Multiple-choice question answering



Ethics: ChatGPT filter developed using dubious labor practices

- OpenAI needed to build a smart filter to screen out offensive content.
- OpenAI engaged a contractor called Sama to develop this filter.
- Sama paid Kenyan workers \$1.32 - \$2 USD per hour to read and label content including details of murder, sexual abuse, suicide, and incest.

<https://afrotech.com/kenyan-workers-openai-chatgpt>

The University is committed to taking meaningful action to prevent, mitigate and, where appropriate, remedy modern slavery risks in its operations, supply chains and investments.

MODERN SLAVERY POLICY 2020, 8(5)

Attribution and use

Be aware of the policies of:

- Your Institution
- Your Funding Bodies
- Your Publishers

in regards to

- Attribution of AI generated content,
- How to acknowledge the use of these tools appropriately, and
- Whether you're allowed to use them in the first place?

Attribution and use: Nature

Scientific Journal ‘Nature’ Says No to Publishing A.I.-Generated Images and Videos, Calling Out Their Lack of ‘Integrity’

The journal, however, will allow text created using large language models, provided their use is documented.

Richard Whiddington, June 14, 2023



what is the best way to peel a mandarin?



Peeling a mandarin, also known as a clementine or a tangerine, is generally a straightforward process. Here's a simple and efficient method to peel a mandarin:

Attribution and use: ARC grant review



Are Australian Research Council reports being written by ChatGPT?

Multiple accounts from researchers suggest that feedback for Discovery Project grant funding was written by artificial intelligence

- Get our morning and afternoon news emails, free app or daily news podcast



<https://www.theguardian.com/technology/2023/jul/08/australian-research-council-scrutiny-allegations-chatgpt-artificial-intelligence>

MEDIA STATEMENT

30 June 2023

Confidentiality obligations of assessors

Like many organisations, the Australian Research Council (ARC) is considering a range of issues regarding the use of generative artificial intelligence (AI) that use algorithms to create new content (such as ChatGPT) and that may present confidentiality and security challenges for research and for grant program administration.

While we are undertaking this work we would like to remind all peer reviewers of their obligations to ensure the confidentiality of information received as part of National Competitive Grants Program processes.

The [Australian Code for the Responsible Conduct of Research](#), 2018 sets out that individuals are to participate in peer review in a way that is fair, rigorous and timely and maintains the confidentiality of the content. If there are concerns with how confidentiality has been managed during a review, the ARC has a robust process to manage this concern. Further information can be found under the [ARC Research Integrity Policy](#).

Release of material that is not your own outside of the closed Research Management System, including into generative AI tools, **may constitute a breach of confidentiality**. As such, the ARC advises that peer reviewers should not use AI as part of their assessment activities.

The ARC will be updating guidance on this area in the near future.

<https://www.arc.gov.au/news-publications/media/media-releases/confidentiality-obligations-assessors>

Problem: “Hallucination” – making things up

Solution: LangChain and other advanced LLM techniques.

- Sophisticated prompts can enforce specific logic.
- Chaining together LLMs can support complex reasoning.
- LLMs can be instructed to use tools to generate responses, including providing references.

Read the Terms of Service or Use

- Always check the policies of whoever is providing the AI service or tool!
- Take statements provided with a healthy grain of salt – tech startups are playing fast and loose, it's the wild west!

Suggestions for best practice

- Use it for brainstorming, bouncing ideas back & forth.
- Be sceptical, fact check answers!
- Only enter information or code that you are ok with someone else being able to have access to.
- Opt-out of them collecting your data if you have the option!

A glimpse of the future

Multimodal models: Text + Image Understanding

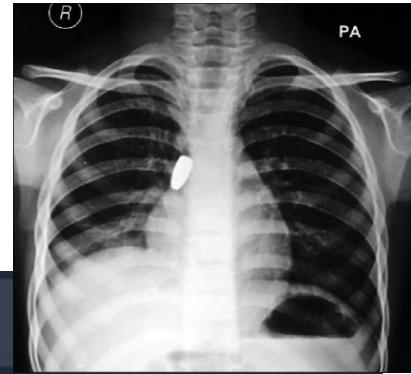
Ecology/Conservation



What animal is in this photo?

A koala is in this photo.

Medical



What is this image?

This image is a black and white photograph of a baby's chest, showing the baby's heart and lungs.

What is unusual about this child's chest x-ray?

The unusual aspect of this child's chest x-ray is that it has a small metal object inside the baby's chest, which is not typically seen in a normal chest x-ray.

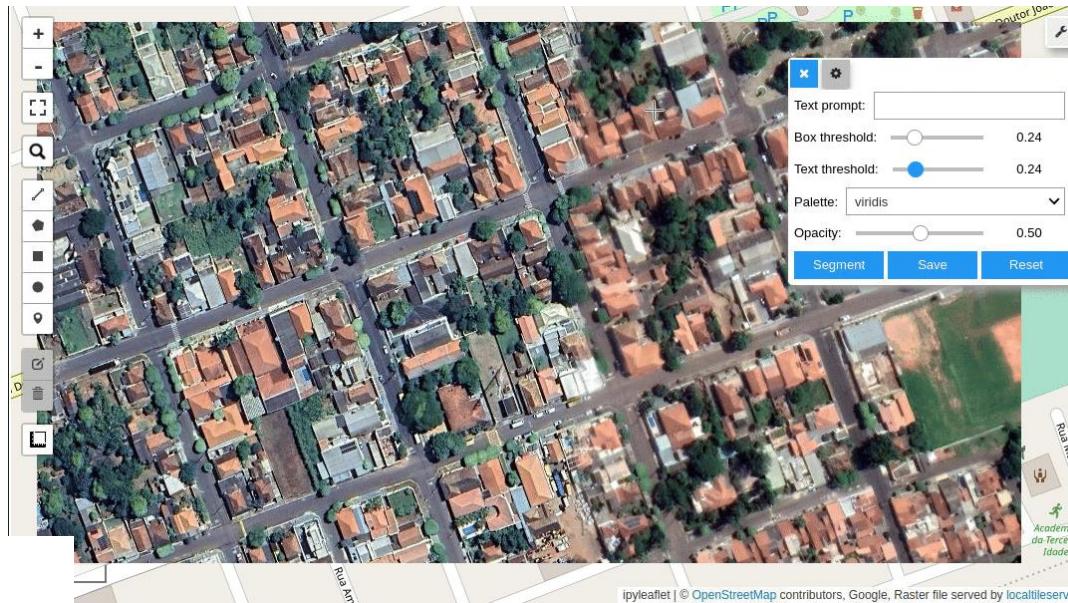
<https://llava.hliu.cc/>

<https://github.com/microsoft/LLaVA-Med>

Multimodal models: Text + Image Understanding

Geospatial:

“Segment all the trees in this aerial photograph, and count them”



<https://samgeo.gishub.org/>

So what will you build?

ChatGPT for Research

ResBaz Sydney, 26th September 2023

Dr Gordon McDonald *Informatics Team Lead*

Dr Henry Lydecker *Data Science Group Lead*

Sydney Informatics Hub

sydney.edu.au/informatics-hub



THE UNIVERSITY OF
SYDNEY

—
Sydney
Informatics Hub

CRCOS 00026A

