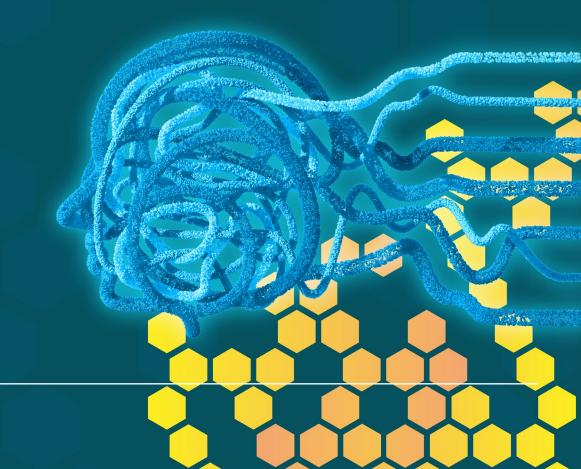


9 May 2024

The Al tools revolution:

Accelerating efficiency in data science

Nick Snellgrove Shanna Tervoort-Mcleod





Al Tools

A polarising subject, for some!

In many domains: chatbots, insurance, health, medical, travel, finance, translations, writing assistants....

Generative Al

"At a high level, generative AI refers to a category of AI models and tools designed to create new content, such as text, images, videos, music, or code. Generative AI uses a variety of techniques—including neural networks and deep learning algorithms—to identify patterns and generate new outcomes based on them."

Github Blog



- o. Dreaking boundaries: Future-Ready Data Science with Al 1001s
- 9. "Al-Driven Horizons: Advanced Data Science Navigation with Powerful Tools"
- 10. "Unveiling Excellence: Mastering Data Science with Cutting-Edge Al Tools"



You

Can the title focus more on Al tools making your work more efficient



ChatGPT

Certainly! Here are 10 titles that emphasize the efficiency boost from Al tools:

- 1. "Efficiency Unleashed: Empowering Work with Al Tools in Data Science"
- 2. "Streamlining Success: Boosting Efficiency with Al Tools in Data Workflows"
- 3. "Al Tools Mastery: Enhancing Efficiency in Data Science Endeavors"
- 4. "Efficient Data Science: Navigating Success with Cutting-Edge Al Tools"
- 5. "Unlocking Efficiency: Transformative Impact of Al Tools in Data Work"
- 6. "Al Tools Revolution: Accelerating Efficiency in Data Science"
- 7. "Efficiency Amplified: Harnessing Al Tools for Data Science Excellence"
- 8. "Precision in Practice: Al Tools Shaping Efficient Data Science Workflows"
- 9. "Efficient Insights: Mastering Data Science with Advanced Al Tools"
- 10. "Strategic Efficiency: Achieving Success through AI Tools in Data Science"



Message ChatGPT...

Tools for data science / analytics

Level of abstraction

• Customer ▶ Waiter ▶ Chef ▶ Prep-cook

For R programming workflows

- GitHub Copilot in RStudio and Posit Workbench (since late September 2023)
- Al powered code utilities
- Online generative AI tools

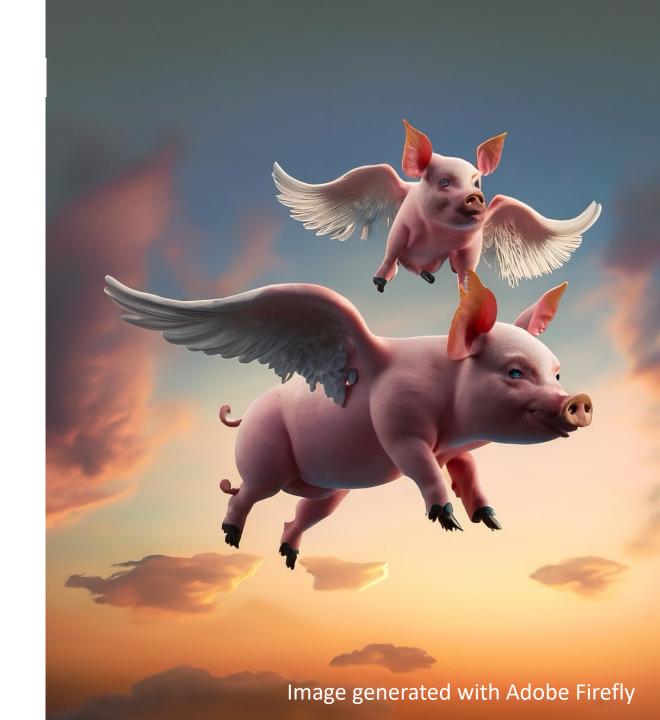




Al and design

What's possible with AI?

- Create impossible images quickly
- Powerful tool to help designers and nondesigners
- Lots of potential to assist in data visualisation

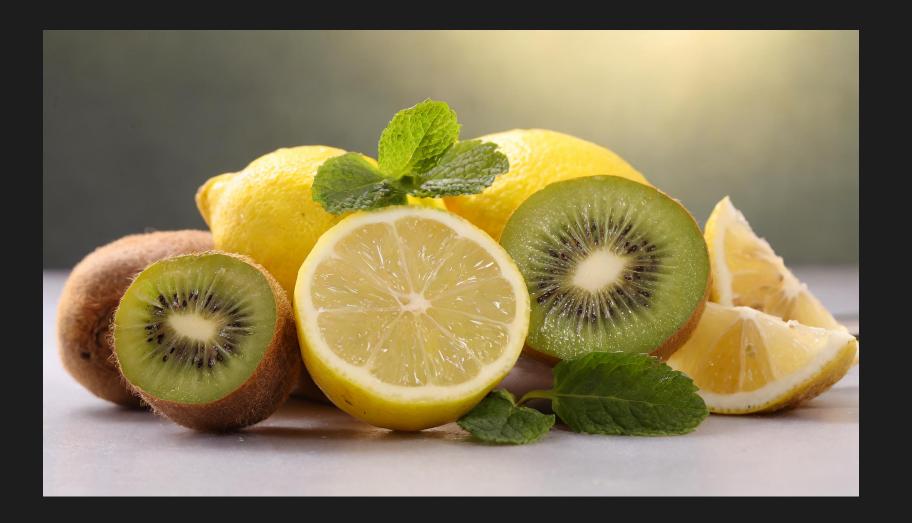


Limitations

- Unable to accurately produce data visualisation images at this stage
 - Can be useful for elements of web/dashboard design and infographics
- Al is not a designer, it is a design tool
 - Good for first draft design ideas
 - · Quickly ideate
 - Alter existing materials
- Easy to get cool result, but hard to get specific results
 - Use a lot of context in your prompts

Prompt

kiwi



Prompt

kiwi bird



Prompt kiwi, New Zealand flightless bird



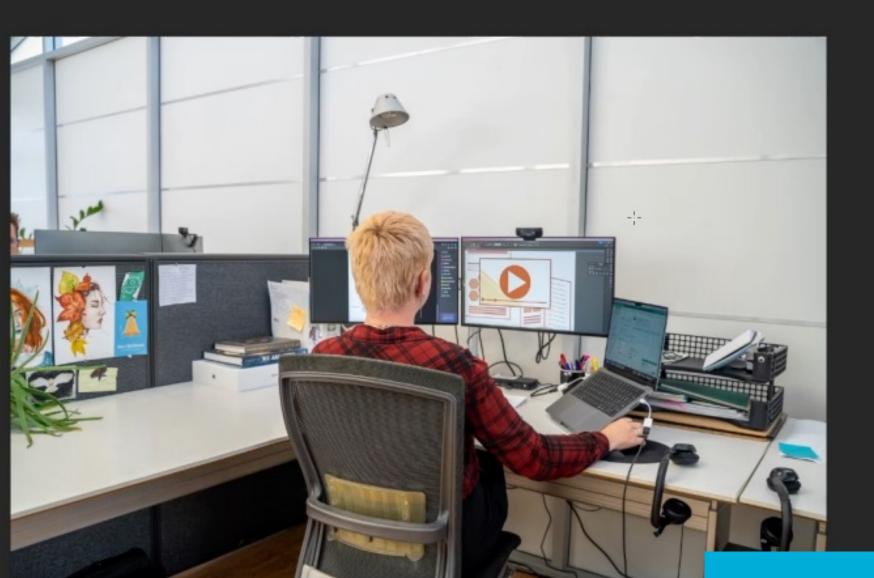
Prompt kiwi, New Zealand flightless bird, long beak



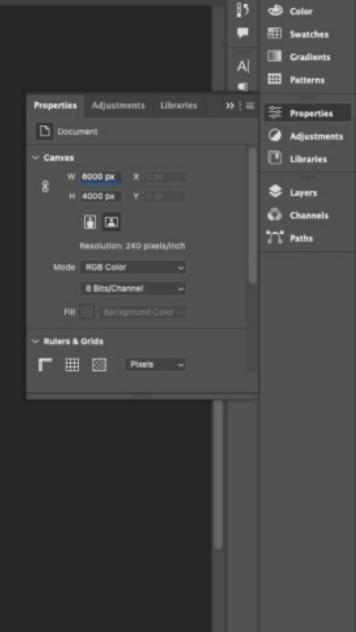
Use AI as a tool

- Creating images
- Custom icons
- Font and colour choice
- Creating videos
- UI/UX design

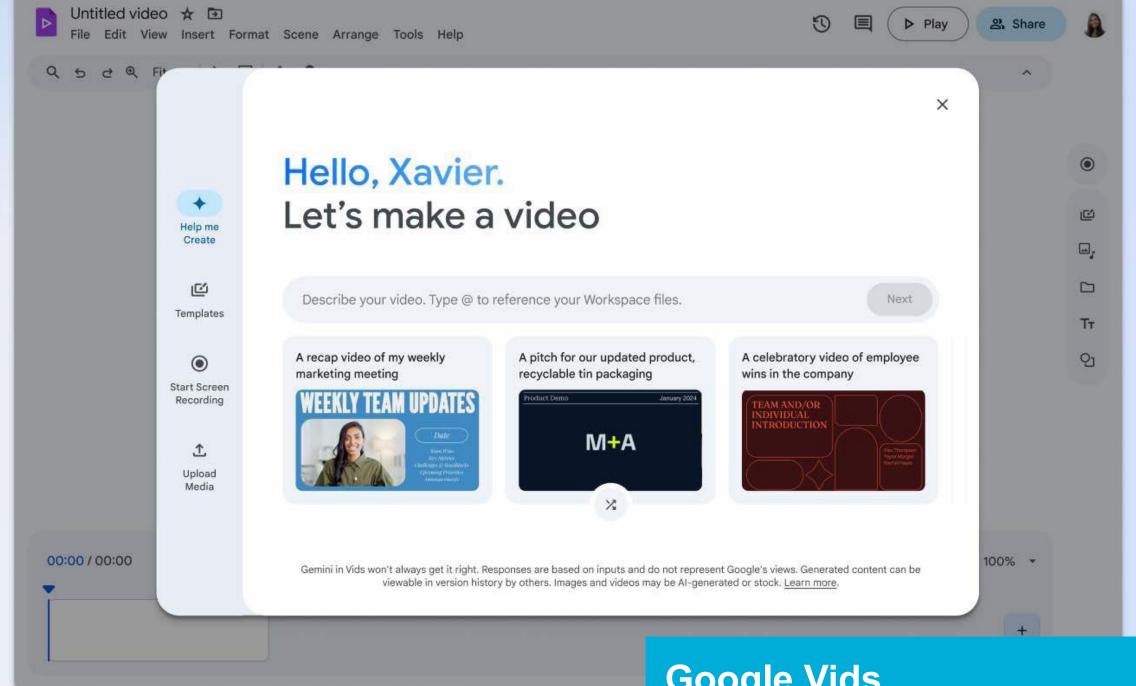




Remove background 🗅 🕜 …

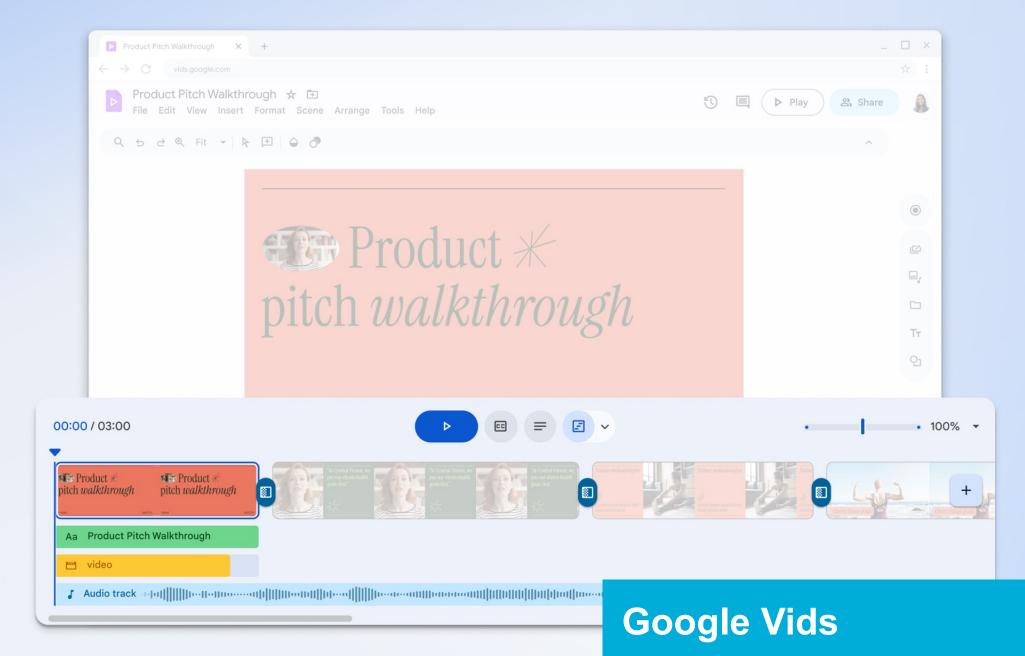


Adobe Generative Fill



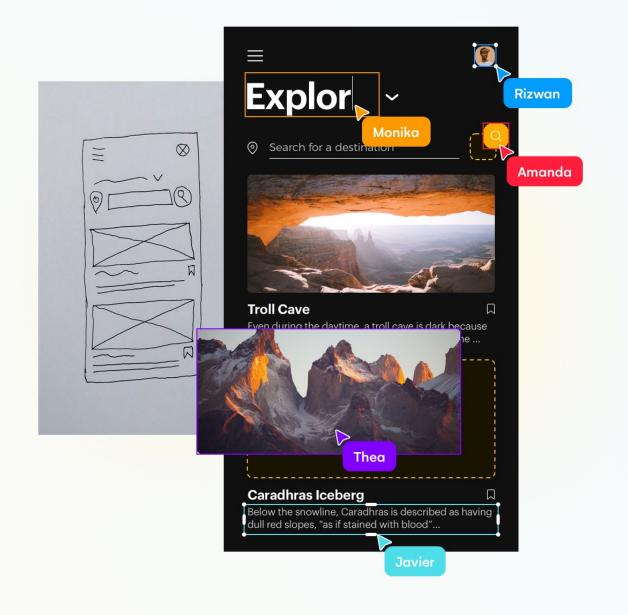
Google Vids

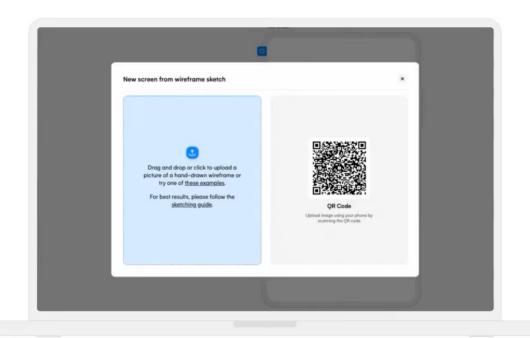
Google Vids - new Workspace app coming soon

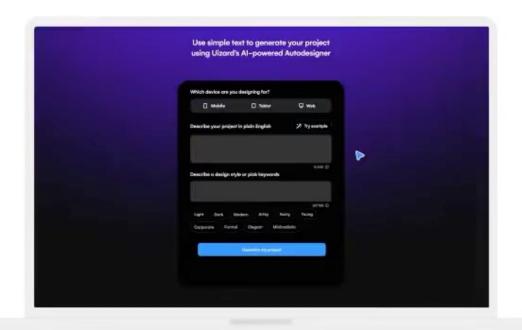


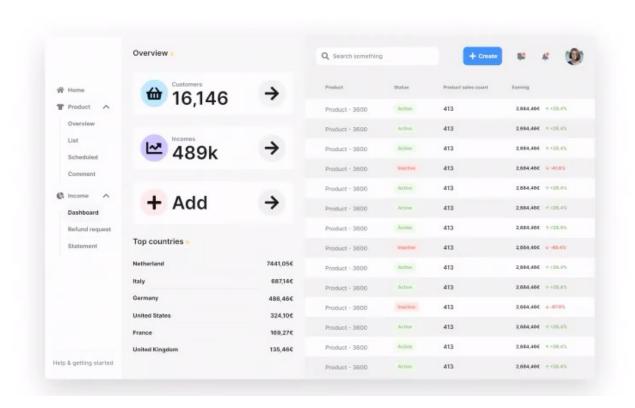


The easiest and fastest tool to design your ideas









Uizard

Ethics

- Al generated art for commercial use is a potential breach of copywrite to artists.
 - Adobe Firefly is guaranteed to be ethical as is based on work that has been paid for and given with permission.
- Always check the rules of the tool you are using.
- Although it may not be stated, it is good practice to cite the AI tool used to create your image.
- Tip to avoid copyright issues is to reverse google image search https://lens.google.com/

Useful developer tools

Useful developer tools





ChatGPT – chat.openai.com

- Open ended conversational AI
- Set the context of a conversation with instructions, then converse to get results
- Make further comments or questions to clarify results

Copilot – github.com/features/copilot

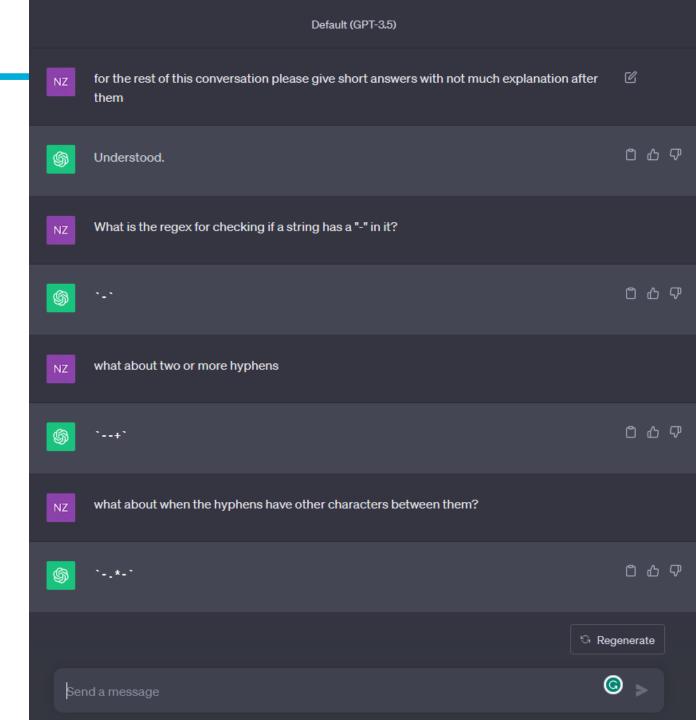
- Autocompletes code suggestions
- Integrates with RStudio and VS Code
- Improves coding practice, rapid iteration of ideas
- Open ended suggestion and prompts

Other useful tools

Phind – https://www.phind.com/
 Al Powered search engine and pair programming tool
 Can customise available sources for Al search

ChatGPT

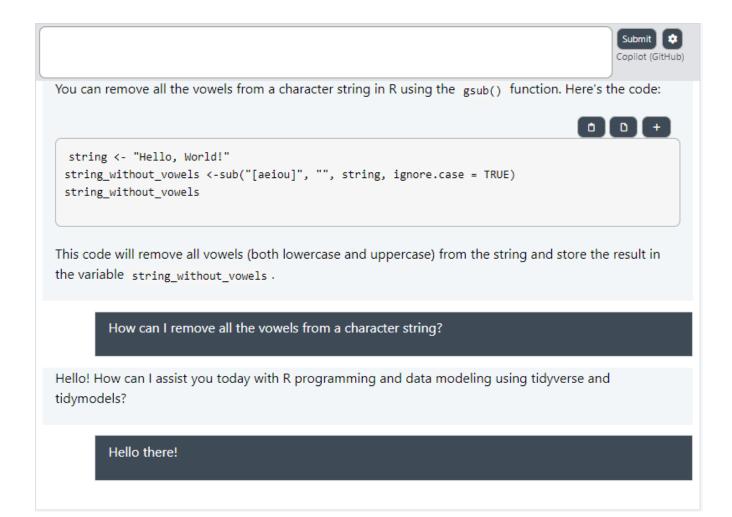
- Fluid conversation between ChatGPT and the developer
- Allows clarification to get better results (without retyping the whole prompt)
- Useful to find out how to do something specific with technology that is well established (e.g. Regex)



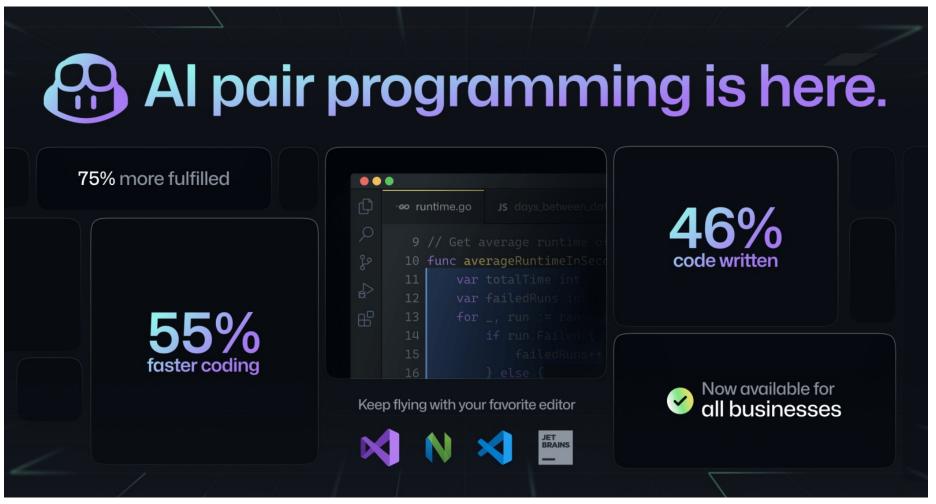
Al Powered R Packages

Extending The RStudio / R workflow with AI

- OpenAl
- GPTStudio
- GPTTools
- Chattr

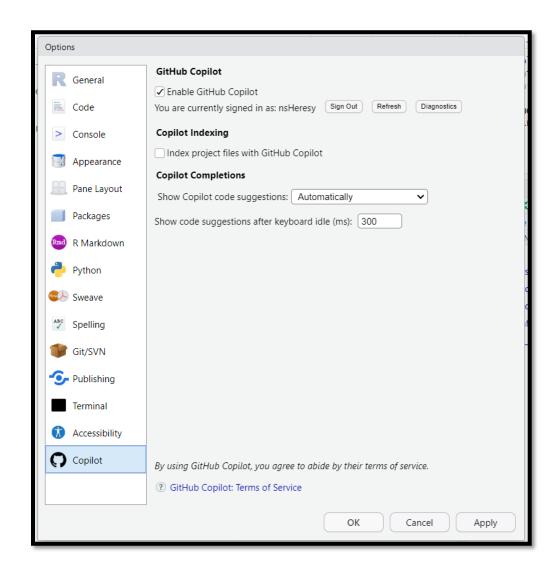


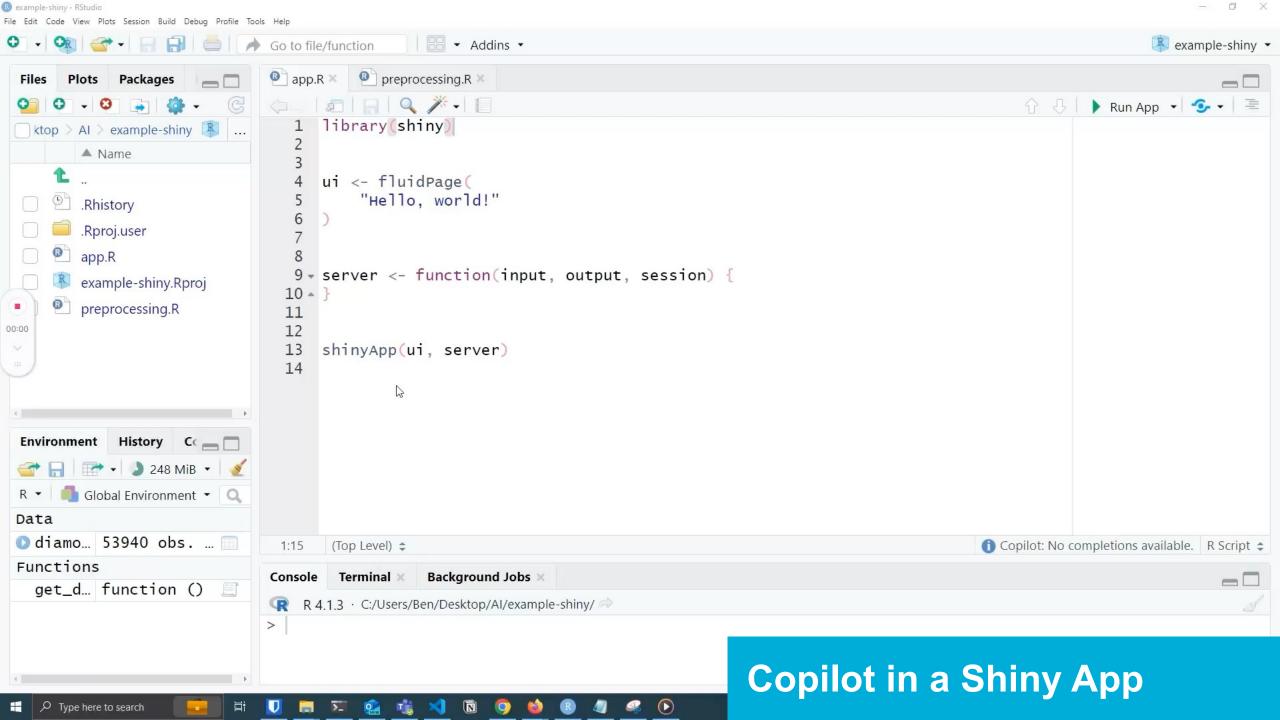
GitHub Copilot

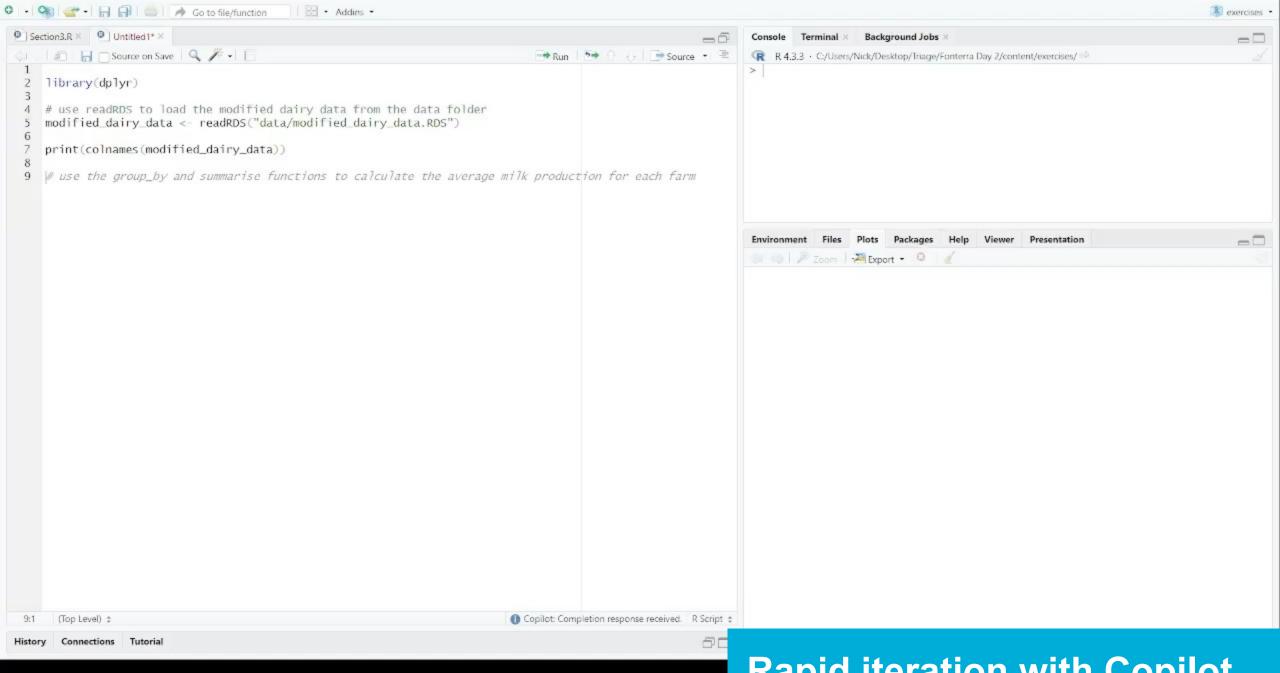


Copilot in RStudio

- Available from RStudio and Posit Workbench 2023.09.0.
- Requires a GitHub account with an active Copilot subscription.
- Needs to be enabled in the project or global settings.
- Consider privacy Indexing project files with Copilot







Rapid iteration with Copilot

Practical uses for large language models

Practical uses for large language models

- Generative AI models (machine learning)
- Identifies patterns and structures in text to generate new content
- Trained on massive amounts of text data
- Understands context and relationships between terms
- Useful in applications for discrete, specific tasks

https://platform.openai.com/examples



Example Shiny App: cleaning and extracting HTML

Forecast for the next 48 hours

| | Monday | | | Tuesda | ıy | | |
|---------------------------------|-------------------|----------------|---------|----------|----------|----------------|-----------------------|
| | Morning Afternoon | | Evening | Night | Morning | Afternoon | Evening |
| Forecast | | | * | * | | | |
| Temperature | 9 °C | 14 °C | 12 °C | 10 °C | 13 °C | 16 °C | 10 °C |
| | Mostly cloudy. | Mostly cloudy. | Clear. | Clear. | Cloudy. | Mostly cloudy. | Drizzle. Overcast. |
| Feels Like | 7 °C | 12 °C | 11 °C | 9 °C | 12 °C | 16 °C | 6 °C |
| Wind Speed | 16 km/h | 36 km/h | 13 km/h | 7 km/h | 8 km/h | 29 km/h | 24 km/h |
| Wind Direction | N ↓ | NNW \ | N ↓ | NNE ✓ | NNE ↓ | SSE | SSE 1 |
| Humidity | 75% | 63% | 69% | 76% | 66% | 60% | 80% |
| Dew Point | 5 °C | 7 °C | 7°C | 6 °C | 7 °C | 8 °C | 6 °C |
| Visibility | 11 km | 10 km | 9 km | 10 km | 11 km | 11 km | 7 km |
| Probability of Precipitation | 0% | 0% | 0% | 0% | 0% | 0% | 5% |
| Amount of Rain | - | - | - | - | - | - | 0.7 mm |

^{*} Updated Monday, 9 October 2023 8:27:47 a.m. Wellington time - Weather by CustomWeather, © 2023

OpenAl API

Overview

The OpenAl API is powered by a diverse set of models with different capabilities and price points. You can also make customizations to our models for your specific use case with fine-tuning.

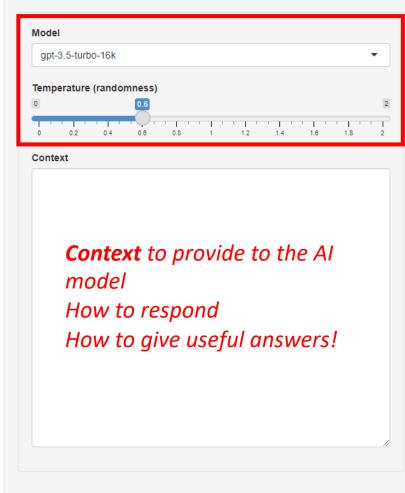
| MODEL | DESCRIPTION |
|-----------------------|----------------------------------------------------------------------------------------------------------------|
| GPT-4 Turbo and GPT-4 | A set of models that improve on GPT-3.5 and can understand as well as generate natural language or code |
| GPT-3.5 Turbo | A set of models that improve on GPT-3.5 and can understand as well as generate natural language or code |
| DALL·E | A model that can generate and edit images given a natural language prompt |
| TTS | A set of models that can convert text into natural sounding spoken audio |
| Whisper | A model that can convert audio into text |
| Embeddings | A set of models that can convert text into a numerical form |
| Moderation | A fine-tuned model that can detect whether text may be sensitive or unsafe |
| GPT base | A set of models without instruction following that can understand as well as generate natural language or code |
| Deprecated | A full list of models that have been deprecated along with the suggested replacement |

OpenAl API parameters

- Model
- Temperature
 - Between 0 and 2 how random will the generated response be?
- Max tokens
 - Maximum number of tokens to be used in the completion
- Messages
 - System: Instructions to the AI, context on how it should behave
 - User: Instructions given by the user
 - Assistant: A response which the AI has provided

Setting up OpenAl

Model **Settings**



Content for the AI model to interpret and provide an answer to.

(Note: This is distinctly connected to the context provided)

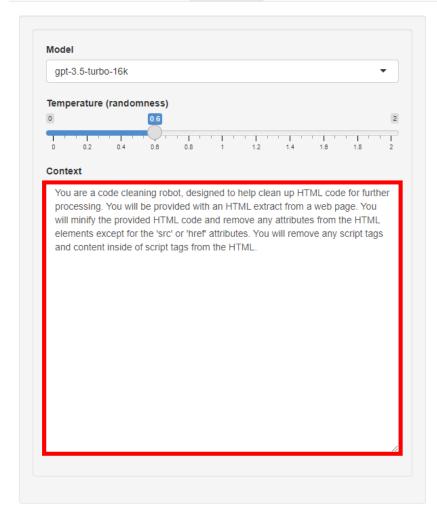
Settings + Context + Content = Result

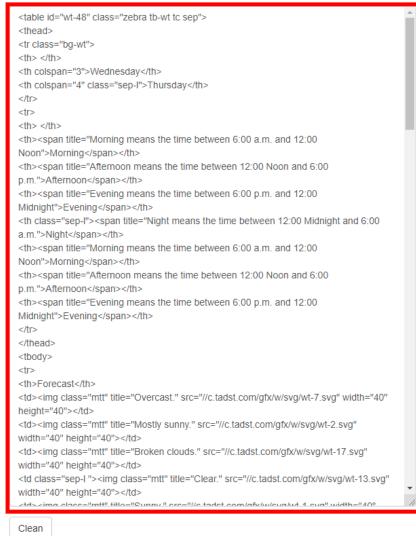
Setting up OpenAl

```
mod_hal_server <- function(id, msgContent, params) {
  moduleServer( id, function(input, output, session){
    ns <- session$ns
    aiOutput <- reactive({
      api_url <- Sys.getenv("OPENAI_ADDR")</pre>
      api_key <- Sys.getenv("OPENAI_KEY")</pre>
      data_to_send <- toJSON(list()</pre>
        temperature = params()$temperature,
        model = params()$model, #added
        messages = list(
                                                               Creating the API Query
          list(role = "system", content = params()$context),
          list(role = "user", content = msgContent())
       , auto_unbox = TRUE)
      response <- POST(
        url = api_url,
        add_headers(
          `Content-Type` = "application/ison",
                                                               Querying the API
          Authorization = paste("Bearer", api_key)
                                                               for a response
        body = data_to_send,
        encode = "json"
      validate(need(response\status_code == 200,
                    message = sprintf("API call failed. Status Code: %s",
                                       response$status_code)))
      content <- fromJSON(content(response, as = "text", encoding = "UTF-8"))</pre>
      return(content)
    })
    return(aiOutput)
```

Clean HTML

Extract Value



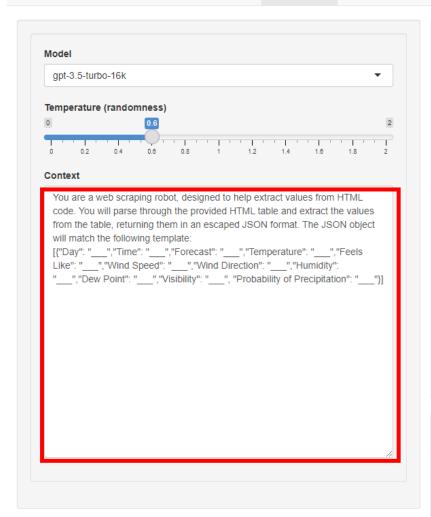


· Prompt token usage: 1578 tokens · Completion token usage: 1035 tokens · Total token usage: 2613 tokens

Finish reason: stop

<thead> colspan="3">WednesdayThursday MorningAfternoon< EveningNight<</pre> MorningAfternoon< Evening</thead>Forecast src="//c.tadst.com/gfx/w/svq/wt-7.svq"><imq src="//c.tadst.com/gfx/w/svq/wt-7.svq"> 2.svg"><imq src="//c.tadst.com/gfx/w/svg/wt-17.svg">class="sep-I"> 1.svg"> Temperature10 °C14 °C12 °C10 class="smaller">Overcast.Mostly sunny.<td class="smaller">Broken clouds.Clear.<td class="smaller">Sunny.Sunny.Scatterred clouds.Feels Like8 °C12 °C9 °C8 °C10 °C17 °C11 °C Wind Speed9 km/h20 km/h31 km/htd> class="sep-I">11 km/h27 km/h26 km/h17 km/h17 km/h18 km/h Wind DirectionESE
<</td>NW
</ NNW
<</td>class="sep-l">NNE
 NNE
^N
^NNE
 ^Humidity81%61%60% Dew Point7 °C6 °C5 °C3 °C5 °C7 °C8 °CVisibility8 °C8 °C</ km11 km10 km11 km11 km11 km 0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0%0% </tdody><tfoot>* Updated Wednesday, 11 October 2023 8:28:00 a.m. Wellington time - Weather by CustomWeather, © 2023ot>

Cleaning the input HTML



Prompt token usage: 1165 tokensCompletion token usage: 559 tokensTotal token usage: 1724 tokens

Finish reason: stop

Show 10 ∨ entries

Showing 1 to 7 of 7 entries

10 km11 km12 km9

| | Day 💠 | Time \$ | Forecast $\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$ | Temperature 🖣 | Feels Like | Wind Speed | Wind Direction | Humidity \$ | Dew Point | Visibility $\mbox{$\frac{1}{7}$}$ | Probability of Precipitation | A |
|---|-----------|-----------|--------------------------------------------------|---------------|---------------|---------------|-------------------|-------------|--------------|-----------------------------------|---------------------------------|----------|
| 1 | Wednesday | Morning | Overcast. | 10 °C | 8 °C | 9 km/h | ESE | 81% | 7 °C | 8 km | 2% | |
| 2 | Wednesday | Afternoon | Mostly sunny. | 14 °C | 12 °C | 20 km/h | NW | 61% | 6 °C | 11 km | 0% | |
| 3 | Wednesday | Evening | Broken clouds. | 12 °C | 9 °C | 31 km/h | NNW | 60% | 5 °C | 10 km | 0% | |
| 4 | Wednesday | Night | Clear. | 10 °C | 8 °C | 11 km/h | NNE | 66% | 3 °C | 10 km | 0% | |
| 5 | Thursday | Morning | Sunny. | 13 °C | 10 °C | 27 km/h | NNE | 60% | 5 °C | 11 km | 0% | |
| 6 | Thursday | Afternoon | Sunny. | 17 °C | 17 °C | 26 km/h | N | 51% | 7 °C | 12 km | 0% | |
| 7 | Thursday | Evening | Scattered clouds. | 13 °C | 11 °C | 17 km/h | NNE | 75% | 8 °C | 9 km | 0% | |

Extracting data from HTML

Be specific!

You are a web scraping robot, designed to help extract values from HTML code. You will parse through the provided HTML table and extract the values from the table, returning them in an escaped JSON format.

| • | Wednesday.Morning.Forecast $^{\hat{	au}}$ | Wednesday.Morning.Temperature | Wednesday.Morning.Feels.Like | $\textbf{Wednesday.Morning.Wind.Speed} ^{\hat{\neg}}$ | $\textbf{Wednesday.Morning.Wind.Direction} \ ^{\hat{\forall}}$ |
|---|-------------------------------------------|-------------------------------|------------------------------|--------------------------------------------------------|------------------------------------------------------------------|
| 1 | Overcast. | 10 °C | 8 °C | 9 km/h | ESE ↑ |

You are a web scraping robot, designed to help extract values from HTML code. You will parse through the provided HTML table and extract the values from the table, returning them in an escaped JSON format. The JSON object will match the following template:

```
[{"Day": "___","Time": "___","Forecast": "___","Temperature": "___","Feels Like":
"___","Wind Speed": "___","Wind Direction": "___","Humidity": "___","Dew Point":
"__","Visibility": "__", "Probability of Precipitation": "__"}]
```

| • | Day [‡] | Time [‡] | Forecast | Temperature [‡] | Feels Like | Wind Speed [‡] | Wind Direction | Humidity [‡] | Dew Point | Visibility [‡] | Probability of Precipitation $\hat{\ }$ |
|---|------------------|-------------------|-------------------|--------------------------|------------|-------------------------|----------------|-----------------------|-----------|-------------------------|-----------------------------------------|
| 1 | Wednesday | Morning | Overcast. | 10 °C | 8 °C | 9 km/h | ESE | 81% | 7 °C | 8 km | 2% |
| 2 | Wednesday | Afternoon | Mostly sunny. | 14 °C | 12 °C | 20 km/h | NW | 61% | 6 °C | 11 km | 0% |
| 3 | Wednesday | Evening | Broken clouds. | 12 °C | 9 °C | 31 km/h | NNW | 60% | 5 °C | 10 km | 0% |
| 4 | Thursday | Night | Clear. | 10 °C | 8 °C | 11 km/h | NNE | 66% | 3 °C | 10 km | 0% |
| 5 | Thursday | Morning | Sunny. | 13 °C | 10 °C | 27 km/h | NNE | 60% | 5 °C | 11 km | 0% |
| 6 | Thursday | Afternoon | Sunny. | 17 °C | 17 °C | 26 km/h | N | 51% | 7 °C | 12 km | 0% |
| 7 | Thursday | Evening | Scattered clouds. | 13 °C | 11 °C | 17 km/h | NNE | 75% | 8 °C | 9 km | 0% |

Validation

| | Wednesda | ay | | Thursday | | | | | |
|------------------------------|---------------------------|---------------|----------------|----------|-----------|---------|--------------------|--|--|
| | Morning Afternoon Evening | | Night | Morning | Afternoon | Evening | | | |
| Forecast | • | 6 | & | * | | | 4 | | |
| Temperature | 10 °C | 14 °C | 12 °C | 10 °C | 13 °C | 17 °C | 13 °C | | |
| | Overcast. | Mostly sunny. | Broken clouds. | Clear. | Sunny. | Sunny. | Scatterred clouds. | | |
| Feels Like | 8 °C | 12 °C | 9 °C | 8 °C | 10 °C | 17 °C | 11 °C | | |
| Wind Speed | 9 km/h | 20 km/h | 31 km/h | 11 km/h | 27 km/h | 26 km/h | 17 km/h | | |
| Wind Direction | ESE | NW | NNW | NNE ↓ | NNE ↓ | N ↓ | NNE ↓ | | |
| Humidity | 81% | 61% | 60% | 66% | 60% | 51% | 75% | | |
| Dew Point | 7 °C | 6 °C | 5 °C | 3 °C | 5°C | 7 °C | 8 °C | | |
| Visibility | 8 km | 11 km | 10 km | 10 km | 11 km | 12 km | 9 km | | |
| Probability of Precipitation | 2% | 0% | 0% | 0% | 0% | 0% | 0% | | |

^{*} Updated Wednesday, 11 October 2023 8:28:00 a.m. Wellington time - Weather by CustomWeather, © 2023

- Does the OpenAl output match the original content?
- More specific prompts leads to more predictable outcome
- Reduce temperature parameter value

| Sho | w 10 v ent | 10 v entries | | | | | | | | | | |
|------|-----------------------------|--------------|-----------------------|---------------|---------------|---------------|-------------------|---------------------------------|--------------|-----------------------------------|------------------------------|--|
| | Day \$ | Time \$ | Forecast \$\display\$ | Temperature 🖣 | Feels Like | Wind Speed | Wind Direction | Humidity $\mbox{$\frac{1}{7}$}$ | Dew Point | Visibility $\mbox{$\frac{1}{2}$}$ | Probability of Precipitation | |
| 1 | Wednesday | Morning | Overcast. | 10 °C | 8 °C | 9 km/h | ESE | 81% | 7 °C | 8 km | 2% | |
| 2 | Wednesday | Afternoon | Mostly sunny. | 14 °C | 12 °C | 20 km/h | NW | 61% | 6 °C | 11 km | 0% | |
| 3 | Wednesday | Evening | Broken clouds. | 12 °C | 9 °C | 31 km/h | NNW | 60% | 5 °C | 10 km | 0% | |
| 4 | Thursday | Night | Clear. | 10 °C | 8 °C | 11 km/h | NNE | 66% | 3 °C | 10 km | 0% | |
| 5 | Thursday | Morning | Sunny. | 13 °C | 10 °C | 27 km/h | NNE | 60% | 5 °C | 11 km | 0% | |
| 6 | Thursday | Afternoon | Sunny. | 17 °C | 17 °C | 26 km/h | N | 51% | 7 °C | 12 km | 0% | |
| 7 | Thursday | Evening | Scattered clouds. | 13 °C | 11 °C | 17 km/h | NNE | 75% | 8 °C | 9 km | 0% | |
| Show | Showing 1 to 7 of 7 entries | | | | | | | | | | Previous 1 Next | |

Tips

- Be specific!
- Be wary of 'hallucinations' and 'confidently incorrect' answers.
- Watch out for missing or incorrect values.
- Check expected prompt token size.
- Pick the right model for your use case & token usage!
- Be mindful of API usage and rate limits.

Food for thought

- Al and Al tools are a constantly evolving field
- New capabilities are constantly emerging
- Still requires careful validation and consideration
- Potential to greatly enhance the work we do

"Artificial intelligence is not a substitute for human intelligence; it is a tool to amplify human creativity and ingenuity."

Fei-Fei Li, Co-Director of the Stanford Institute for Human-Centered Artificial Intelligence and IT Professor at the Graduate School of Business









