

<https://bit.ly/birn-data>

Jonathan Soma
Columbia University
js4571@columbia.edu
[@dangerscarf](https://twitter.com/dangerscarf)

AI for Reporting

tricks for the newsroom

The Lede Program for Data Journalism

ledeprogram.com

Home Coursework Apply Tuition & Fees FAQs About

The Lede Program

An intensive, remote-friendly summer program from Columbia's Graduate School of Journalism designed to equip journalists and storytellers of all kinds with the skills needed to turn data into narrative

The Lede Program

Columbia Journalism School's Lede Program is a 10-week intensive on coding, data analysis, and visual storytelling. For 2024, the course will be once again operate with **in-person and remote options**.

No Prerequisites

We start from zero, **with no prior experience in data or coding necessary** – You'll learn it all along the way. If you can turn on your computer, you're ready to go!

10 weeks

M.S. Data Journalism | Columbia University

journalism.columbia.edu/ms-data-journalism

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

COLUMBIA JOURNALISM SCHOOL

News Giving Rentals Apply

Academics People Professional Learning Centers Community

About

Home Academics M.S. Data Journalism

Program Spotlight: M.S. in Data Journalism

COLUMBIA JOURNALISM SCHOOL

Share

PROGRAM SPOTLIGHT

M.S. in Data Journalism

Watch on YouTube

M.S. Data Journalism

Journalism in the 21st century involves finding, collecting and analyzing data for storytelling, presentation and investigation

12 months

<https://bit.ly/birn-data>

Jonathan Soma
Columbia University
js4571@columbia.edu
@dangerscarf

AI for Reporting

tricks for the newsroom

Scraping Python + Playwright

The image is a collage of three screenshots from a computer interface, likely a Jupyter Notebook or similar environment, demonstrating data analysis and web scraping.

- Top Left Screenshot:** A Jupyter Notebook cell showing Python code for language detection using the langdetect library. The code defines a function `detect_language` that takes a text column from a DataFrame and returns the detected language. It handles exceptions and adds a 'lang' column to the DataFrame. The cell output shows the progress of the `progress_apply` operation.
- Bottom Left Screenshot:** A screenshot of a web browser displaying a service provider's website for ground restoration and clearance services. The URL is <https://www.ludwigshafen24.com>. The page content includes a heading "Data Analysis Python + pandas" and some descriptive text about their services.
- Bottom Right Screenshot:** A Jupyter Notebook cell showing Python code using the Playwright library to extract data from a web page. The code uses asynchronous functions to wait for elements to appear and then extract data from a table using locators like ".trRecord". The extracted data is stored in a list and appended to a DataFrame.

A screenshot of a Google Colab notebook titled "02 - Using AI about AI (completed)". The code cell contains Python code for data analysis:

```
# Assuming df is your DataFrame
def detect_language(text):
    try:
        return detect(text)
    except LangDetectException:
        return 'Unknown' # In case language detection fails

# Add the 'lang' column to the DataFrame
df['lang'] = df['text'].progress_apply(detect_language)
```

Data Analysis
Python + pandas

AI wrote our code once we knew
how to run Python and knew the
word “pandas” and “dataframe”

A screenshot of a Jupyter Notebook titled "Untitled6.ipynb". The code cell contains Python code for web scraping using Playwright:

```
# Function to extract data from the current page
async def extract_data(page):
    # Create a DataFrame to store the data
    data = []

    # Extract the table rows (trRecord class)
    rows = page.locator(".trRecord")
    count = await rows.count()

    for i in range(count):
        # Extract the data from each row
        data.append(await rows.nth(i).inner_text())
```

Scraping
Python + Playwright

AI wrote our code once we knew
how to inspect website code and
that we use should Playwright

The background of the image features a dense, repeating pattern of stylized flowers and birds. The flowers are primarily yellow and pink, with intricate petal structures. Interspersed among the flowers are small, detailed depictions of birds, some perched on branches and others in flight. The overall effect is reminiscent of traditional Asian or Art Nouveau decorative arts.

A.I. is

magic patterns

Playground

Complete ▾

Your presets ▾

Save

View code

Share

...



Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished! 0
Would you care to join me for a meal? I know of
street. Their pastries are simply divine. What do
for kings and queens?

fam = 92.57%

r = 4.18%

starving = 1.60%

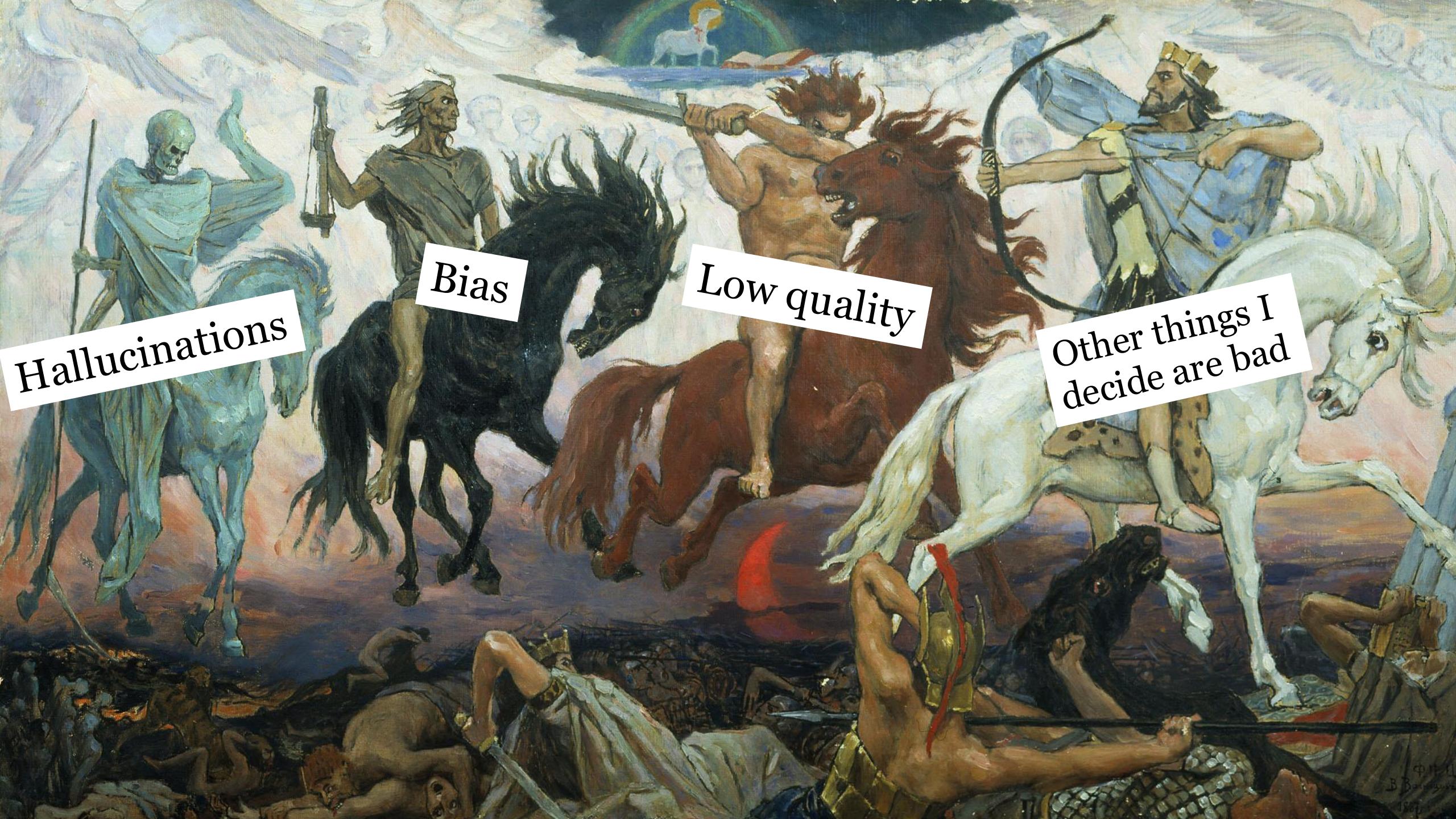
faint = 0.41%

wasting = 0.27%

star = 0.24%

parch = 0.17%

with = 0.09%



Hallucinations

Bias

Low quality

Other things I
decide are bad



A.I. should be used for
error-resilient tasks

“explain that???”

JS

You

Write a sonnet about my fat, mean cat Mulberry.

ChatGPT

Beneath the moon, Mulberry takes her throne,
A cat of girth, with fur like midnight's shade.
Her eyes, twin emeralds, cold as stone,
Within their depths, her fierce commands are made.

Her claws, like daggers, glint in soft moonlight,
A queen who rules with strict and stern decree.
Her growl can turn the bravest heart to fright.
Her hiss a symphony of

low-quality output is ignorable

when shadows dance and silent

She curls beside me, softening I

Revealing tender moments, warr

For though she's fat and often se

In dreams, sweet Mulberry's a gei



New York Sessions

Year Session

2023-
2024

General Assembly

20091

3380

[Browse](#)

tight feedback loops

```
# Python function to calculate fibonacci sequence
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

errors are expected

[Introduced](#)

[Completed](#)

[Legislation](#)

(a.k.a. "intern work")



```
# Python function to calculate fibonacci sequence
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

tight feedback loops

tight feedback loops

Website Scraping Guide

chatgpt.com/c/91f73e39-05c9-4487-b969-d90ef5fef264

ChatGPT 4o

Handling Pagination

Since the results span multiple pages, we'll need to handle pagination.

```
python
# All of our data will go here
all_data = pd.DataFrame()

while True:
    # Scrape the data from the current page
    df = await scrape_current_page(page)
    all_data = pd.concat([all_data, df])

    try:
        # Click the "Next Page" button
        await page.locator("a:has-text('Next Page')").click()
    except:
        # If the "Next Page" button is not present, break
        break

# Display the first few rows of the data
all_data.head()
```

Message ChatGPT

ChatGPT can make mistakes. Check important info.

Untitled6.ipynb (2) - JupyterLab

File Edit View Run Kernel Tabs Settings Help

Untitled6.ipynb

Notebook Python 3 (ipykernel)

```
# Display the first few rows of the data
all_data.head()
```

NameError
Cell In[4], line 42
38 all_data = pd.DataFrame()
40 while True:
41 # Extract data
---> 42 data = await extract_data(page)
43 all_data = pd.concat([all_data, data])
45 # Try clicking the next page button

Cell In[4], line 32, in extract_data
23 license_status = page.locator("a:has-text('Last Name')").click()
24 last_name_goto = page.locator("a:has-text('Last Name')").click()
25 data.append([
26 "License Status",
27 "Company Name",
28 "License Number",
29 "Board Name",
30 "License Type",
31 "Last Name",
32 "First Name",
33 })
35 return pd.DataFrame(data)

NameError: name 'last_name_goto' is not defined

[8]: !pwd
/Users/soma

Simple 0 \$ 3 Python

This guide should cover the entire process for scraping the Appraisal Management Companies licenses from the website. If you encounter any issues or need further adjustments, feel free to reach out!

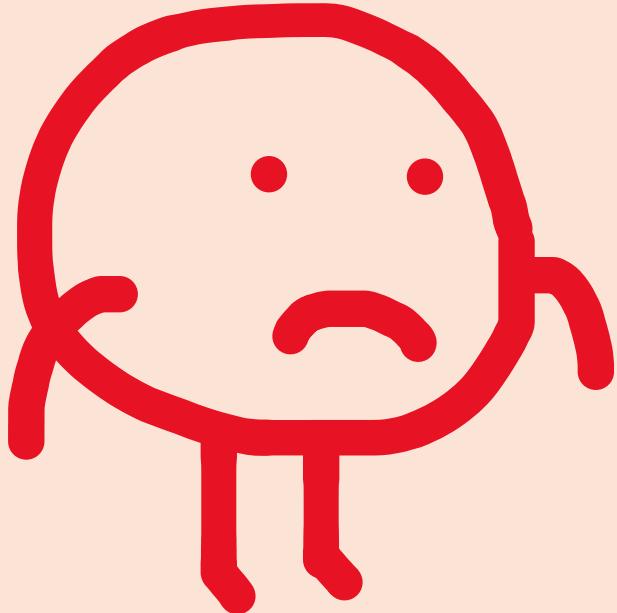
Error: Locator.click: Error: strict mode violation:
locator("input[type='submit'][value='Search']") resolved to 2 elements:
1) <input type="submit" value="Search" class="btn btn-primary" />
aka locator("input[name='j_id0||j_id1||j_id14||j_id73']")
2) <input type="submit" value="Search" class="btn btn-primary" />
aka locator("input[name='j_id0||j_id1||j_id14||j_id79']")

Call log:
waiting for locator("input[type='submit'][value='Search']")

The error indicates that the selector is matching two elements, which violates the "strict mode" in Playwright. This happens because both buttons have the same attributes. We need to be more specific in selecting the correct "Search" button.

Message ChatGPT

ChatGPT can make mistakes. Check important info.



Ex-USA Today writer fabricates parts of major stories

deseret.com/2004/3/21/19818659/ex-usa-today-writer-fabricated-parts-of-major-stories/

Deseret News | Deseret Magazine | Church News Print Subscriptions

DeseretNews

Newsletter Sign Up → Latest The West Utah Sports Opinion Magazine More X f | Search

U.S. & WORLD

Ex-USA Today writer fabricated parts of major stories

Published: March 21, 2004, 12:00 a.m. MST

By Deseret News, Blake Morrison, USA Today

Seven weeks into an examination of former USA Today reporter Jack Kelley's work, a team of journalists has found strong evidence that Kelley fabricated substantial portions of at least eight major stories, lifted nearly two dozen quotes or other material from competing publications, lied in speeches he gave for the newspaper and conspired to mislead those investigating his work.

Perhaps Kelley's most egregious misdeed occurred in 2000, when he used a snapshot he took of a Cuban hotel worker to authenticate a story he made up about a woman who died fleeing Cuba by boat. The woman in the photo neither fled by boat nor died, and a USA Today reporter located her this month. If Cuban authorities had learned she was the woman in the picture, she says, she could have lost her job and her chance to emigrate.

Kelley, 43, resigned from the newspaper in January after he admitted conspiring with a translator to mislead editors overseeing an inquiry into his work. At the time, newspaper

VIEW COMMENTS **SHARE**

Most Popular



2 Obamas, 1 Latter-day Saint: Highlights from Tuesday night at the Democratic convention



JS

You

Write a sonnet about my fat, mean cat Mulberry.



ChatGPT

Beneath the moon, Mulberry takes her throne,
A cat of girth, with fur like midnight's shade.
Her eyes, twin emeralds, cold as stone,
Within their depths, her fierce commands are made.

Her claws, like daggers, glint in soft moonlight,
A queen who rules with strict and stern decree.
Her growl can turn the bravest heart to fright.
Her hiss a symphony of

low-quality output is ignorable

when shadows dance and silence wraps the air,
She curls beside me, softening her might,
Revealing tender moments, warm and rare.

For though she's fat and often seems so mean,
In dreams, sweet Mulberry's a gentle queen.



Grammarly



Rooms that are tiny can be tricky to decorate but they can also be a lot of fun. So when a client challenged us to give her pocket size space a summer makeover for under \$500 dollars, we just couldn't say no. Transforming a very small space doesn't have to blow your budget. Small things like finding a vintage piece of furniture from a relative or adding a fresh coat of paint to your own dated items can add a stylish splash to any abode.

Correctness

2 alerts



Clarity

A bit unclear



Engagement

A bit bland



Delivery

Slightly off



Sauna

lauramiina.github.io/suomi-sauna/



Traditional wood-burning smoke sauna in Lehmonkärki in southern Finland. Smoke saunas take hours to heat up. Photo: Mila Kukkonen

FINNS WANT TO SWEAT THEIR TROUBLES AWAY

Finns love sauna, and build saunas everywhere they go. Still, finding a perfect Finnish sauna experience abroad is tough. It can be equally challenging to get foreign friends to fully grasp the importance of sauna culture in Finns' lives.

By [Laura Kukkonen](#)

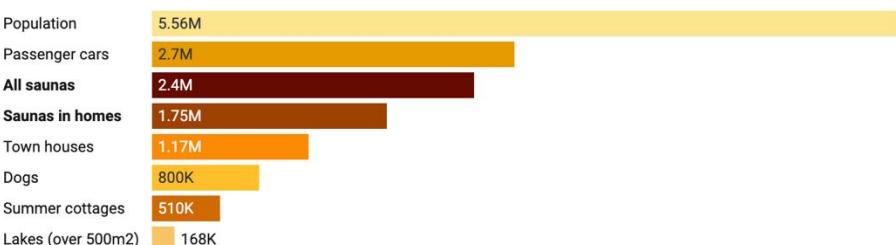
My childhood summers were spent on the countryside of Eastern Finland, at my grandparents' old house turned into a summer cottage. Mostly I remember the summers as sunny and bright, but the best part was going to sauna on a rainy day. My sisters and I would sit with our aunt in the dimly lit, humid, wood paneled room that Finns call "löyly", usually in about 80 degrees Celcius. We threw water on the steaming hot rocks above the wood burning stove called "kiuas" and enjoyed the scents of the bath whisk made by our grandmother out of fresh birch branches. After many rounds of löyly and going outside to cool off, we got to wash our hair outside in the rain.

I will never feel as refreshed as I did during those nights after sauna, drinking redcurrant juice at the cottage with my family.

Millions of Finns have similar fond memories of sauna, where you spend time naked with strangers, family, friends, and the occasional business acquaintance. In fact, there are millions of saunas in Finland, more than town houses and dogs combined.

The land of saunas

Number of saunas in Finland compared to the amount of other things found in the country.



Category	Value
Population	5.56M
Passenger cars	2.7M
All saunas	2.4M
Saunas in homes	1.75M
Town houses	1.17M
Dogs	800K
Summer cottages	510K
Lakes (over 500m2)	168K

Chart: Laura Kukkonen • Source: Statistics Finland, Automotive Information Center, Finnish Kennel Association • [Get the data](#) • Created with

Guide

Below are some of the most important AP Style rules for a data journalist to use when editing their work.

- Capitalization: Only capitalize proper nouns, titles directly before a name, and the first word in a sentence.
- Numbers: Spell out numbers one through nine; use figures for 10 and above, except in specific contexts (ages, dimensions, etc.).

...(many more rules).....

Instructions

Provide suggestions for improving the text of the work below as a list of bullet points. The text is provided under the 'PAGE TO BE EDITED' heading.

Copy editing guidelines

- Only address the copy of the piece.
- Do not nest bullet points.
- Only use the AP style guide to make suggestions.
- Every bullet point must be something that needs to be fixed.

Sauna



lauramiina.github.io/suomi-sauna/

Replace "there are millions of sauna's in Finland" with "there are millions of saunas in Finland" to correct the misuse of an apostrophe.

Millions of Finns have similar fond memories of sauna, where you spend time na
strangers, family, friends, and the occasional business acquaintance. In fact, the
saunas in Finland, more than town houses and dogs combined.

The land of saunas

Number of saunas in Finland compared to the amount of other things found in the country.

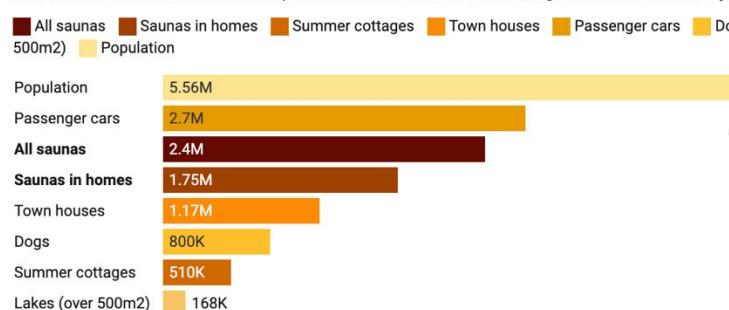
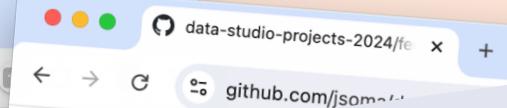
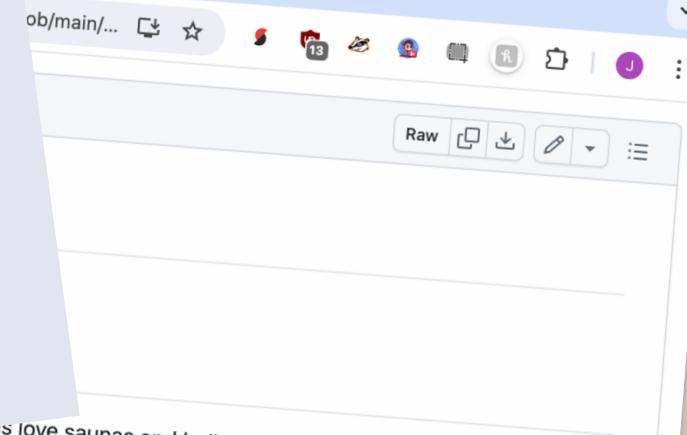


Chart: Laura Kukkonen • Source: Statistics Finland, Automotive Information Center, Finnish Kennel Association • [Get the data](#) • Created with



data-studio-projects-2024/fe

github.com/jsonoma...



ob/main/...

Raw

Download

Edit

View

More

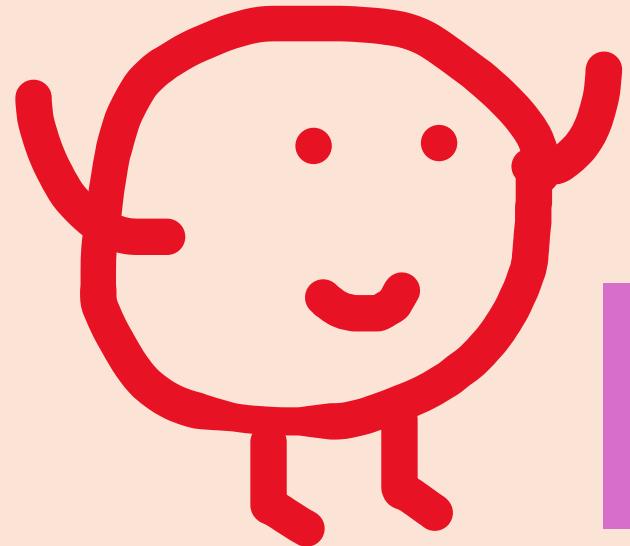
...is love saunas and build them..."
... degrees Celsius."

- Change "at least 80 degrees celcius" to "at least 80 degrees Celsius" for consistency and correct spelling of Celsius.
- Amend "causionary tales" to "cautionary tales" for correct spelling.
 - Change "like nights with friends and definitely bachelorette parties." to "including nights with friends and

Change "two" to "2" and "75" to "75" for consistency with AP style guidelines on ages.

- Amend "All of them go to sauna at least once a week, most of them more often." to "All of them go to sauna at least once a week, most more frequently." for conciseness and clarity.
- Correct "Details about this project can be found onGitHub." to "Details about this project can be found on GitHub." adding a space for correct branding and readability.
- Change "My childhood" to "My Childhood" if it's intended as a subtitle for consistency with capitalizing major words in titles.

in titles



Suggestions!

not demands!
if it's silly, ignore it

JS

You

Write a sonnet about my fat, mean cat Mulberry.

**ChatGPT**

Beneath the moon, Mulberry takes her throne,
A cat of girth, with fur like midnight's shade.
Her eyes, twin emeralds, cold as stone,
Within their depths, her fierce commands are made.

Her claws, like daggers, glint in soft moonlight,
A queen who rules with strict and stern decree.
Her growl can turn the bravest heart to fright.
Her hiss a symphony of

low-quality output is ignorable

when shadows dance and silent

She curls beside me, softening I

Revealing tender moments, warr

For though she's fat and often se

In dreams, sweet Mulberry's a gei



```
# Python function to calculate fibonacci sequence
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

tight feedback loops

New York Sessions

Year Session

2023-
2024

General Assembly

20091

3380

Browse

Introduced

Completed

Legislation

(a.k.a. "intern work")



New York Sessions

Year Session

2023-
2024

General Assembly

errors are expected

Introduced

Completed

Legislation

20091

3380

Browse

(a.k.a. "intern work")

Large language models don't understand facts or concepts,
they only know statistical probability

Chatbots May ‘Hallucinate’ More Often Than Many Realize

Can be dangerous for translation and summarization

When summarizing facts, ChatGPT technology makes things up about 3 percent of the time, according to research from a new start-up. A Google system’s rate was 27 percent.

 Share full article



By Cade Metz

Cade Metz has been watching chatbots h...

Published Nov. 6, 2023 Updated Nov. 16, 2023

Hughes Hallucination Evaluation Model (HHEM) leaderboard

This leaderboard (by [Vectara](#)) evaluates how often an LLM introduces hallucinations when summarizing a document.

T	Model	Hallucination Rate (%)
●	openai/GPT-4-Turbo	3
●	openai/GPT-4	3
?	microsoft/Orca-2-13b	3.2
●	openai/GPT-3.5-Turbo	3.5
●	llama-2-70b	4.8
●	llama-2-13b	5.1
●	llama-2-13b-Pro	5.6

Brazilian Restaurant Delivery | X +

claude.ai/chat/5804fabb-fc5c-4730-9db4-97d9fd63e929

ide Brazilian Restaurant Delivery Menu ▾

**Translating a restaurant menu?
Errors are probably okay!**

CARDÁPIO EXCEPCIONAL DE DELIVERY

SÁBADO

ALMOÇO

- MOQUECA DE OVO COM CAMARÃO..... 34,00
- MOQUECA SHIMEJI..... 32,00
- MOQUECA DE FRUTOS DO MAR..... 62,00
- MOQUECA DE PEIXE..... 62,00

Acompanhamentos: arroz branco, pirão e farofa de acarajé ou farofa de dendê.

DOMINGO

ALMOÇO

- CARANGUEJADA..... C/ 1UN37,00/2UN49,00
- XINXIM DE FRANGO, VATAPÁ E FAROFÁ..... 48,00
- COZIDO BAIANO..... 49,00
- BAIÃO DE DOIS TRADICIONAL E VEGANO..... 35,00
- CALDO DE VÔNGOLE..... 24,00

TODOS OS DIAS

QUITOTES

- ACARAJÉ..... 25,00
- ABARÁ..... 30,00
- COXINHA DE PAELLA..... UN12,00/6UN55,00
- BOLO DE TAPIOCA (RENATINHO)..... 12,00

RESERVAS E PEDIDOS PELO WHATSAPP
(11) 9 7956-4494
ENTREGA GRÁTIS ATÉ 800M

JS explain this restaurant menu to me

This is a menu for a Brazilian restaurant called "Acará Fazz" offering delivery services. The menu is divided into three main sections:

1. Saturday (Sábado) Lunch:
 - Features various types of "moqueca" (a traditional Brazilian fish stew):
 - Egg and shrimp moqueca (34.00)

The menu includes sections for Saturday Lunch (SÁBADO), Sunday Lunch (DOMINGO), and All Days (TODOS OS DIAS). It lists several traditional Brazilian dishes such as Moqueca de Ovo com Camarão, Caranguejada, and Acarajé, along with their prices. The prices are listed in reais (R\$).

Section	Item	Price
SÁBADO (Saturday Lunch)	MOQUECA DE OVO COM CAMARÃO	34,00
	MOQUECA SHIMEJI	32,00
	MOQUECA DE FRUTOS DO MAR	62,00
	MOQUECA DE PEIXE	62,00
DOMINGO (Sunday Lunch)	CARANGUEJADA	C/ 1UN37,00/2UN49,00
	XINXIM DE FRANGO, VATAPÁ E FAROFÁ	48,00
	COZIDO BAIANO	49,00
	BAIÃO DE DOIS TRADICIONAL E VEGANO	35,00
	CALDO DE VÔNGOLE	24,00
	TODOS OS DIAS (All Days)	ACARAJÉ
ABARÁ		30,00
COXINHA DE PAELLA		UN12,00/6UN55,00
BOLO DE TAPIOCA (RENATINHO)		12,00
RESERVAS E PEDIDOS PELO WHATSAPP		(11) 9 7956-4494
ENTREGA GRÁTIS ATÉ 800M		

AMENDED IN SENATE JUNE 27, 2024
AMENDED IN SENATE JUNE 10, 2024
AMENDED IN SENATE JUNE 10, 2024

Writing a story about a law?
Errors are NOT okay!

AB 886

Introduced by Assembly Member Wicks
(Coauthors: Assembly Members Lowenthal and E
(Coauthors: Senators McGuire and Skinner) McGuire, Si
Stern)

February 14, 2023

An act to add Title 23 (commencing with Section 3273 of Division 3 of the Civil Code, relating to civil law.

LEGISLATIVE COUNSEL'S DIGEST

AB 886, as amended, Wicks. California Journalism Protection Act. Existing law generally regulates online platforms, requiring, on a semiannual basis and as specified, a company to submit to the Attorney General a terms of service contract that includes, among other things, the current version of the terms of service of the social media platform.

This bill, the California Journalism Preservation Act, requires a covered platform, as defined, to either pay at least \$ _____ to compensate digital journalism providers, as defined, for accessing the

20230AB886_93.pdf
Page 1 of 18

20230AB886_93.pdf
Page 1 of 18

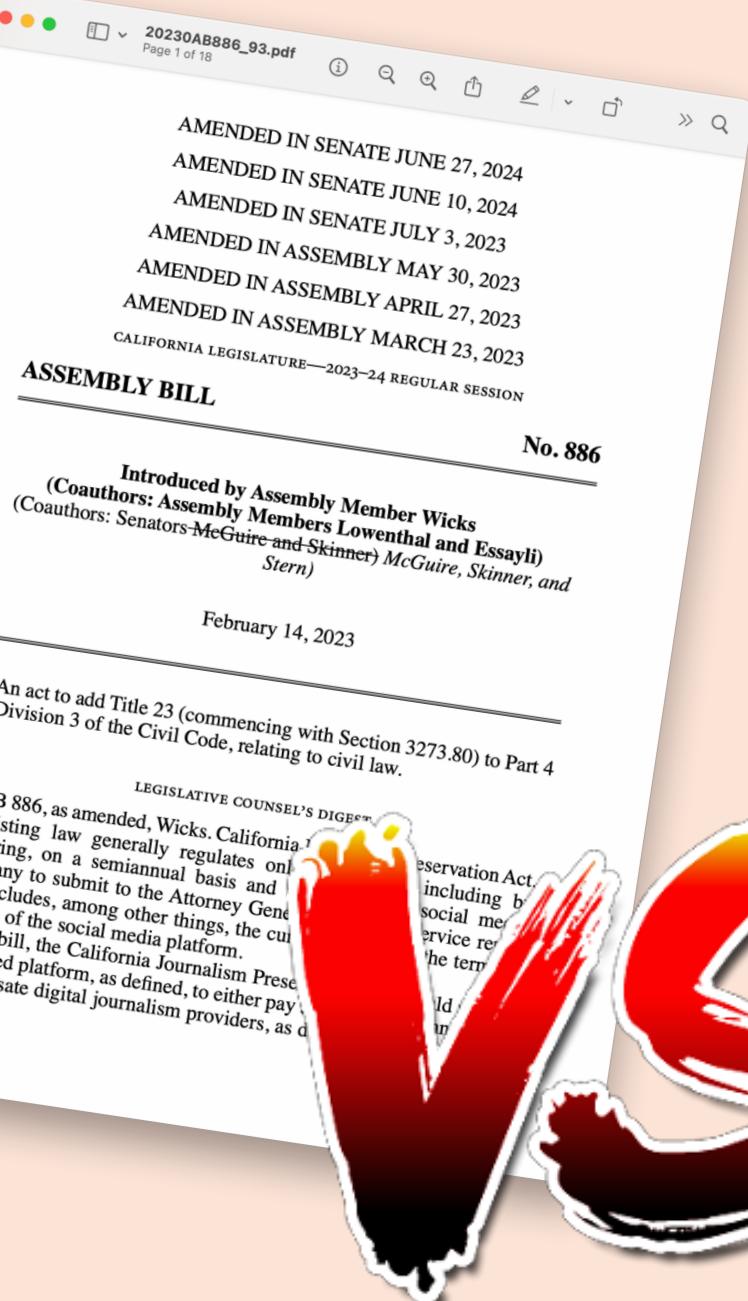
California Journalism Preservation Act advances through California Senate Judiciary Committee

June 26, 2024 • The NewsGuild-CWA

The California Senate Judiciary Committee passed the [California Journalism Preservation Act](#) (AB 886) on Tuesday with a vote of 9-2 during a scheduled hearing. The bill, introduced by Assembly Member Buffy Wicks, now advances to the Appropriations committee before a final vote on the California Senate floor.

1 2 3 4 8 12 16

17 18



When we have many documents, maybe we want to...

Find some examples of similar pieces of news legislation

Find what percent of legislation is about journalism

Get an alert each time journalism legislation is introduced

Find what legislators proposed legislation about journalism

In these situations, maybe a small error rate is okay?

sometimes you have many
documents and want

a little information

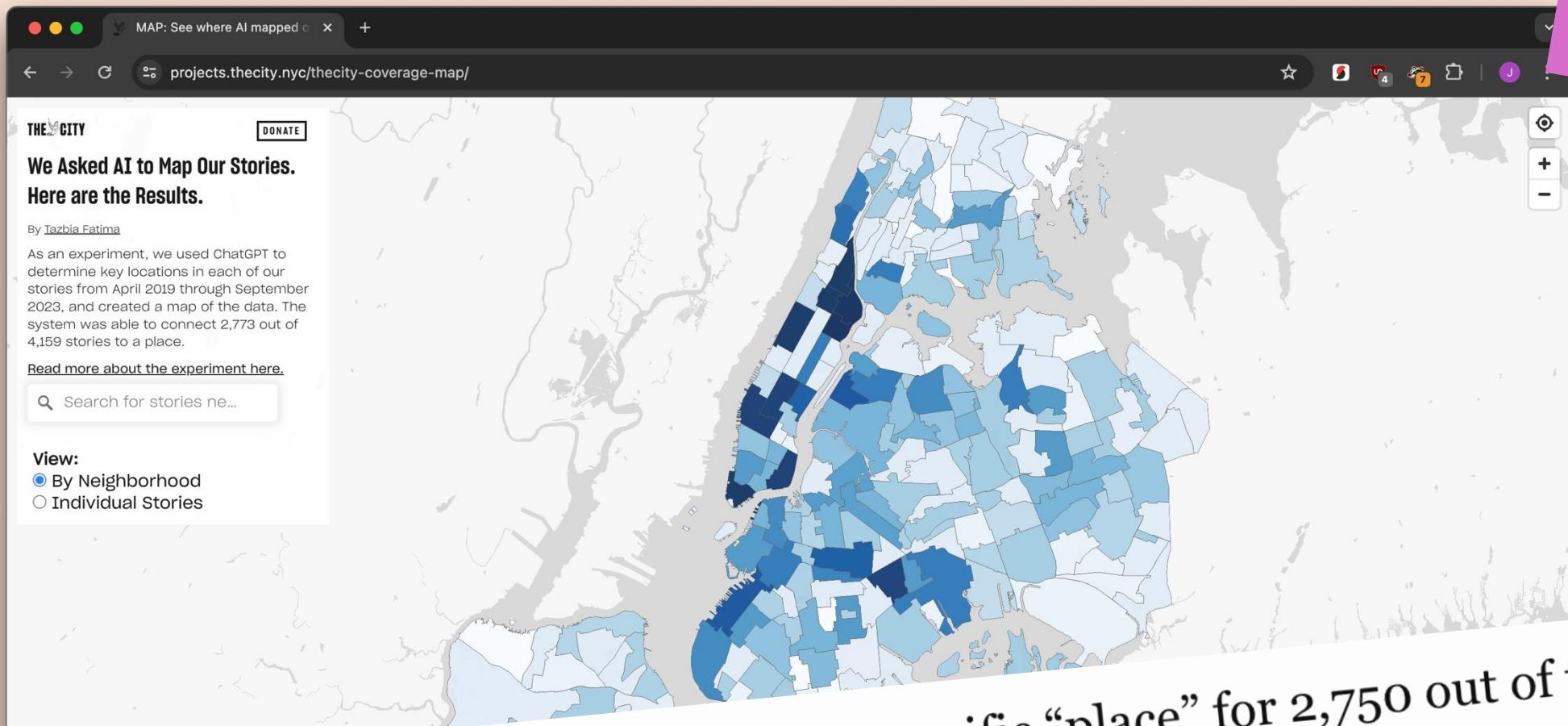
from or about each of them

an example

A screenshot of a web browser displaying an article from The Washington Post. The title of the article is "Apple says its App Store is 'a safe and trusted place.' We found 1,500 reports of unwanted sexual behavior on six apps, some targeting minors." Below the title, a subtitle reads "The prevalence of unwanted sexual content raises questions about whether Apple can continue to offer a protective cocoon to its customers as its platform grows". The author's name is Reed Albergotti and Al Johri, and the date is November 22, 2019 at 7:00 a.m. EST. A sidebar on the left features a quote from a user: "all this app hate and Nothing I had one". The Washington Post logo is at the top center, and a user profile for "jonathan.soma" is on the right.

Using a machine learning algorithm to identify App Store reviews containing reports of unwanted sexual content, racism and bullying, The Post sifted through more than 130,000 reviews of six random chat apps, all but one of which were ranked in the top 100 for social networking by Apple earlier this month. The Post manually inspected the more than 1,500 reviews that made mention of uncomfortable sexual situations.

an example



ChatGPT was able to pick a specific “place” for 2,750 out of the 4,159 stories we published from our April 2019 launch through September 2023. We were also able to match 2,129 stories to a given neighborhood. We used the

Claude

claude.ai/new

Professional Plan

Good morning, j

How can Claude help you today?

Claude 3.5 Sonnet ▾

Add content

Generate excel formulas

Polish your prose

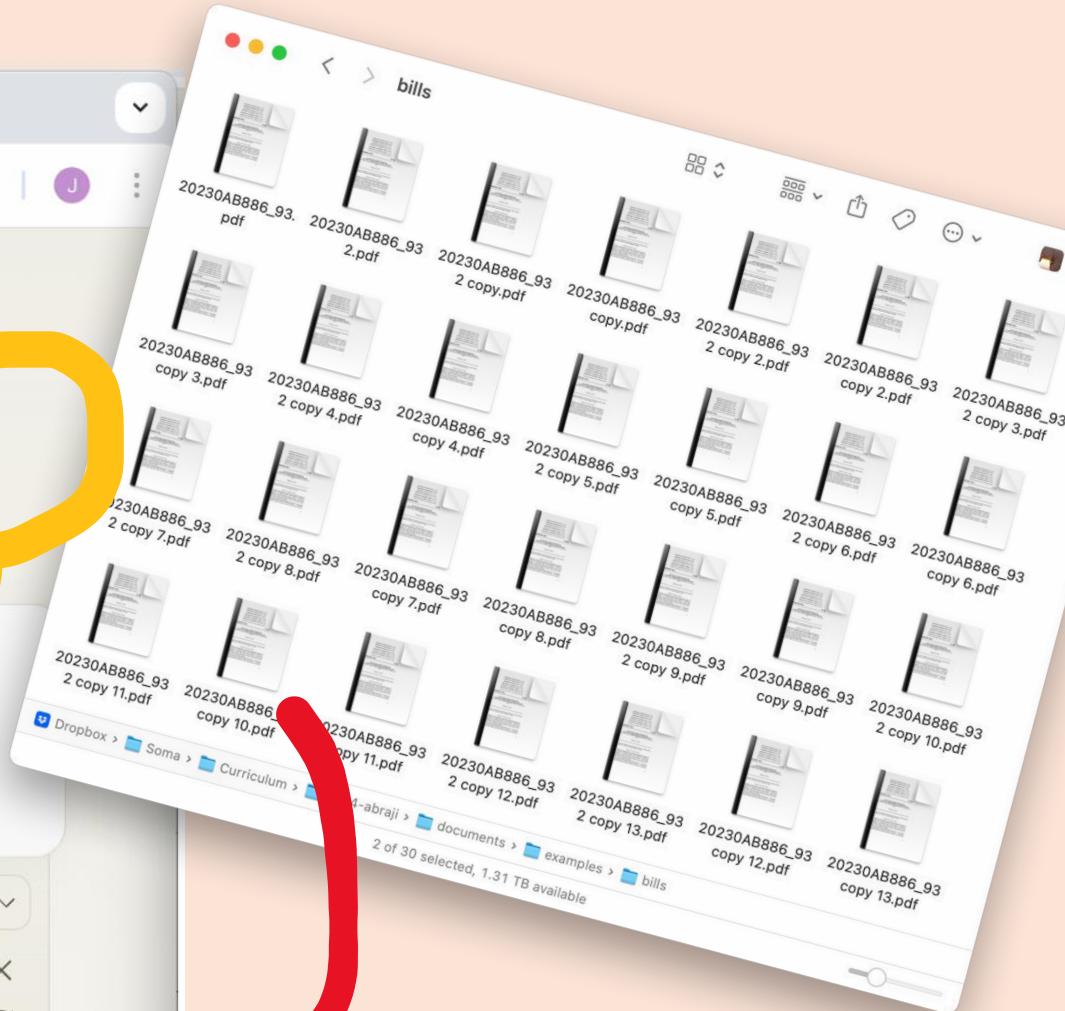
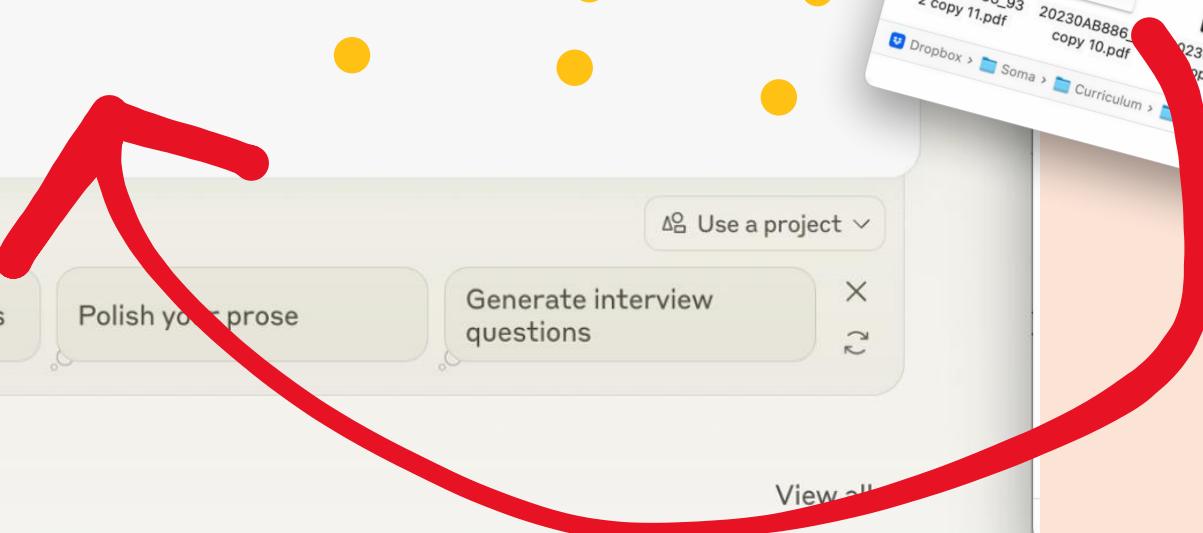
Generate interview questions

Use a project ▾

Your recent chats ^

View all

JS



Claude for Sheets™

Claude for Sheets™ brings your friendly AI assistant Claude from Anthropic right there in your spreadsheets.

By: [Anthropic](#)

Listing updated: March 11, 2024

Install

★★★★☆ 14 ⓘ ↓ 30K+

Works with:



AI in Sheets™ .Spread...

The most powerful AI tool for Google Sheets™ on the planet. Built for true scale. Safe for the enterprise. Unbeatable pricing model.

By: [Spreadsheet Daddy](#)

Listing updated: March 18, 2024

Install



GPT for Sheets™ and ...

ChatGPT in Google Sheets™ and Docs™. Get the full power of AI for inspiration data cleaning, classification, extraction, translation, edition, summarization, writing directly in you...

By: [Talarian](#)

Listing updated: March 6, 2024

Independent security verification

Uninstall

★★★★☆ 318 ⓘ ↓ 6M+

Claude for Sheets™ - Google workspace.google.com/marketplace/app/clause_for_sheets/909417792257

Google Workspace Marketplace Search apps

Home Editor's choice Intelligent apps Work from everywhere Business essentials Apps to discover

Built by Google Top charts

Most popular Top rated Categories Admin managed

Claude for Sheets™

Uninstall

Claude for Sheets™ brings your friendly AI assistant Claude from Anthropic right there in your spreadsheets.

By: [Anthropic](#) Listing updated: March 11, 2024

Works with:

★ ★ ★ ★ ★ 17 i 34K+

Overview Permissions Reviews

This sheet lets you compare a set of related prompts to Claude

Step 1: type your prompts in this column

In one sentence, what is good about the color blue?
Suppose you have a table called "Customers" with the following column names:
CustomerID, CustomerName, ContactName, Address, City, PostalCode, Country
Can you write a SQL query that finds all customers in San Francisco with a customer ID of greater than 1000?

Step 2: Select your model:

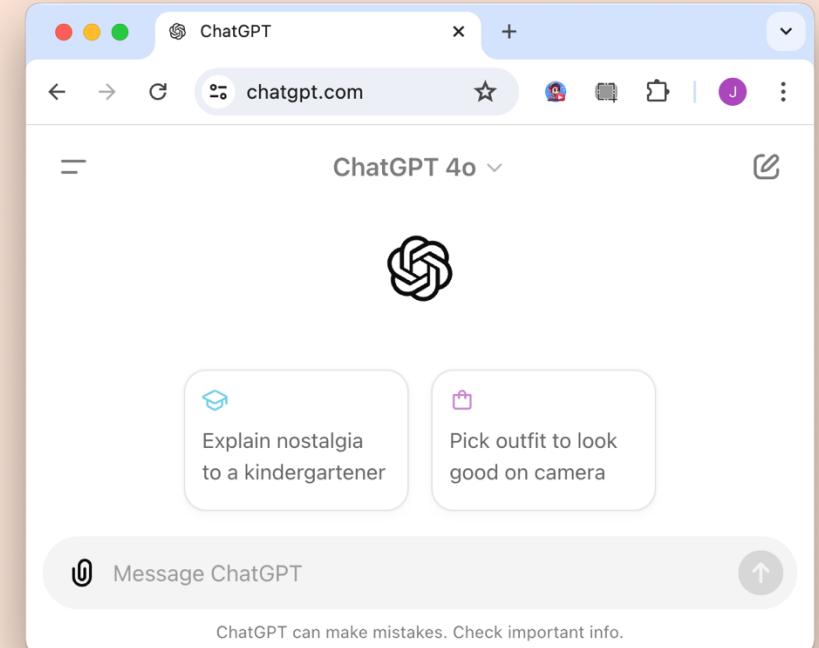
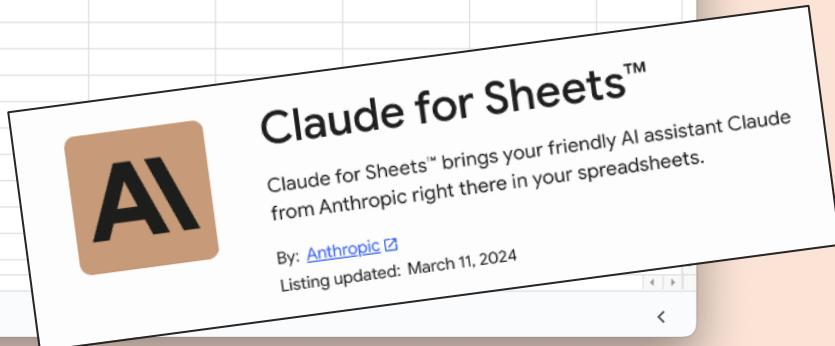
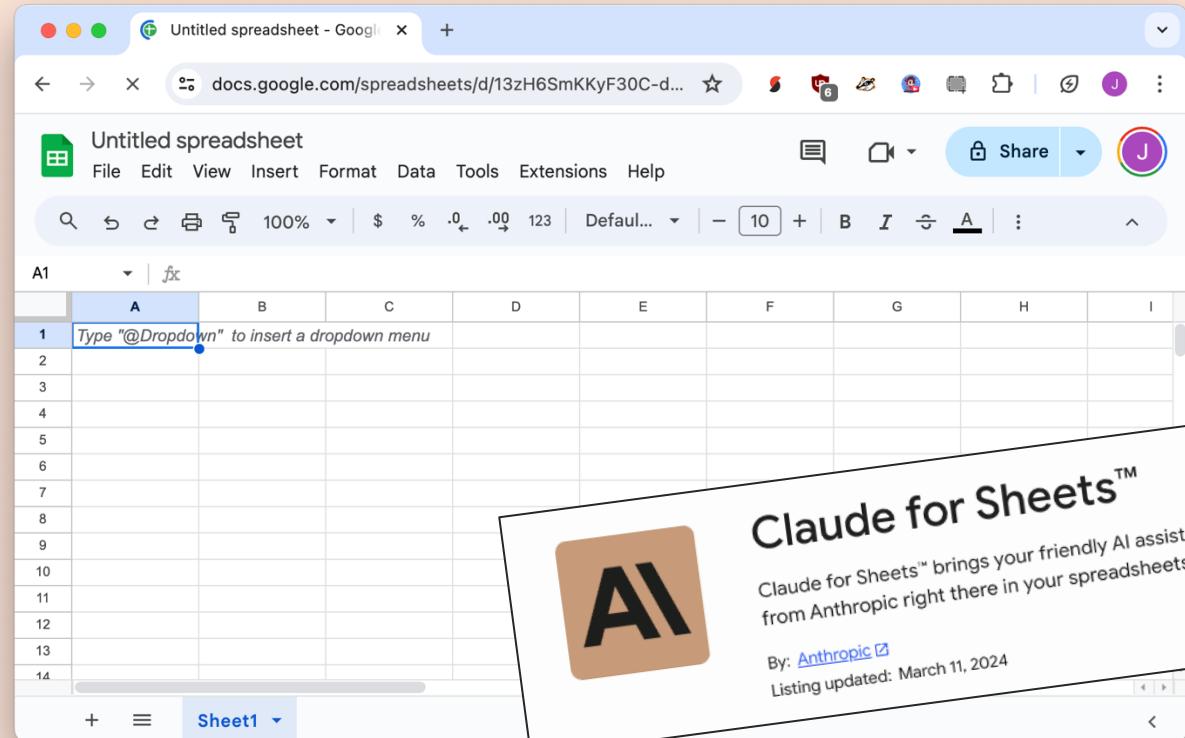
claudie-v1.3
claudie-2

Step 3: review Claude responses in this column

Blue is a calming and soothing color.
Here is a SQL query to find all customers in San Francisco with a customer ID 1000 from the Customers table:
...sql
SELECT *
FROM Customers
WHERE City = 'San Francisco' AND CustomerID > 1000
...
This query selects all columns (*) from the Customers table where the City co

Extensions Help
Add-ons Macros Apps Script AppSheet Claude in Sheets

asks question



gives answer

Key successfully added!

You must keep a record of the key below as you will not be able to view it again.

sk-ant-api03-Wriod0W3KzsL_-osV4lnJCMfdZ6...
nPxe-cpCdewNwV2sRKL_pfrkj1mQCTqoDcEo1fJX...
cld7exevD-Q-y5y_KAAA

API keys

platform.openai.com/api-keys

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them. Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that has leaked publicly. Enable tracking to see usage per API key on the [Usage page](#).

NAME	SECRET KEY	TRACKING	CREATED	LAST USED	PERMISSIONS
Secret key	sk-...MJ0b	+ Enable	Mar 6, 2023	Apr 13, 2023	All
Secret key	sk-...e67A	+ Enable	Mar 18, 2023	Apr 19, 2023	All
Secret key	sk-...1PWe	+ Enable	Mar 18, 2023	Mar 18, 2023	All
Secret key	sk-...MiLX	+ Enable	Mar 19, 2023	Mar 19, 2023	All
Secret key	sk-...2ABE	+ Enable	Mar 23, 2023	Never	All

Close

console.anthropic.com/settings/keys

ANTHROPIC

Dashboard Workbench Settings

Create Key

Your Organization Members & Invites Plans & Billing API Keys Logs Usage

API Keys

NAME	KEY	CREATED AT	UPDATED AT
claude for sheets	sk-ant-api03-1Sk...iAAA	Apr 9, 2024	...
sample class	sk-ant-api03-g0n...rgAA	Apr 10, 2024	...

200% | \$ % .0 .00 123 | Default... | - 10 + | B I A |

A1

| fx

A**B**

1

2

3

4

5

6

7

8

E8

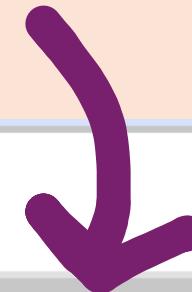
fx

	A	B	C
1	object	color	
2	sky		
3	strawberry		↗
4	basketball		
5			
6			
7			
8			

=CLAUDEEXTRACT("cor (em português)", A2)



object	color
céu	céu
morango	Não há "cor (em português)" nessa string.
basquetebol	None

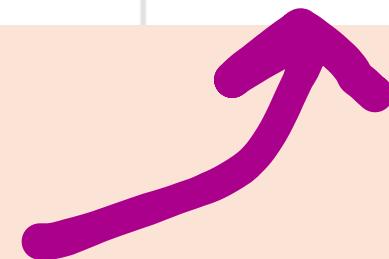


=CLAUDEEXTRACT("ngjyra (нë shqip)", A2)



object	color
qielli	blu
Iuleshtrydhe	kuqe
basketbolli	nuk ka informacion për ngjyrën në këtë fjalë

object	color
небо	блакитний
полуница	червоний
баскетбол	None

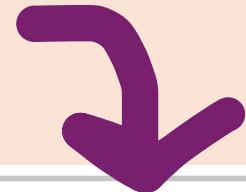


=CLAUDEEXTRACT("колір (українською мовою)", A2)

=CONCATENATE("Qual é a cor de '", A2, "'?
Responda em uma ou duas palavras.")



=CLAUDE(B2)



object	full prompt	response
céu	Qual é a cor de 'céu'? Responda em uma ou duas palavras.	Azul claro.
morango	Qual é a cor de 'morango'? Responda em uma ou duas palavras.	Vermelho.
basquetebol	Qual é a cor de 'basquetebol'? Responda em uma ou duas palavras.	Laranja.

Louisiana State Legislation (Completed in Class) ★ Saved to Drive

File Edit View Insert Format Data Tools Extensions Help

150% \$.00 123 Default... 10 B I A

G3 fx

	A	E	F	G
1	bill text	about retirement?	about employment?	
2	> HLS 24RS-94 **[ENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 8 > > BY REPRESENTATIVES MELERINE AND HUGHES > > SCHOOLS/HIGH SCHOOL: Prohibits the use of an appeals pro > certain students > > HLS 24RS-88 **[REENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefits c > the >			
3				

Louisiana State Legislation (Completed in Class) ★ Saved to Drive

File Edit View Insert Format Data Tools Extensions Help

150% \$.00 123 Default... 10 B I A

G3 | fx

	A	E	F	G
1	bill text > HLS 24RS-94 **[ENGROSSED]{.underline}** > > 2024 Regular Session	about retirement?	about employment?	

=CLAUDEEXTRACT("Is this piece of legislation about retirement? Answer YES or NO", A)

fx

	A			
3	> certain students > > HLS 24RS-88 **[REENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefit > the >	no	no	yes

I am irate about the broccoli incident, I am never shopping at your store again. Take me off of your broccoli mailing list immediately.

name

Jackary Baloneynose
jackary.baloney@example.com

email

email	email	name
FROM: Mulberry Peppertown (mulbs@example.com) When I pick up the cans of beans they are all so light! At first I thought they were empty, but it turns out they are just futuristic beans that are not heavy like the old style beans I was used to. It is incredible.		=CLAUDEEXTRACT(...)
I am irate about the broccoli incident, I am never shopping at your store again. Take me off of your broccoli mailing list immediately.	mulbs@example.com	Mulberry Peppertown
Jackary Baloney nose jackary.baloney@example.com	jackary.baloney@example.com	Jackary Baloney nose
Your flour is ground too finely. I do not go here to buy 00 flower I want big, fat grainy level-2 flour. I want it so coarse that it can be mistaken for grains of sand scattered over my countertop. You can find me at bfids@example.com my name is Boxcar Fiddleworth and I would like to know when you get the good correct stuff.	bfids@example.com	Boxcar Fiddleworth
Cleo here, reporting live: I am not sure whether to go with cinnamon or sugar. I love sugar, I hate cinnamon. cleo@example.com .	cleo@example.com	Cleo

“email”

“name”

Elections européennes : Emm... +

lemonde.fr/elections-europeennes/article/2024/04/17/elections-europeenne... 🔍 🌐 ⚡ 🎧 📁 🎯 🎪 🎭 🎮 🎵 🎶 🎩 🎪 🎭 🎮 🎵 🎶 🎩

Le Monde | Élections européennes 2024 | Résultats des européennes | Carte des résultats | Les 81 députés français | S'abonner

ÉLECTIONS EUROPÉENNES • EMMANUEL MACRON

Élections européennes : Emmanuel Macron dénonce « l'hypocrisie » du Rassemblement national

Le chef de l'Etat a déploré, mercredi depuis Bruxelles, que les « peurs profitent aux réponses les plus simplistes » dans la campagne pour le scrutin du 9 juin, fustigeant ceux qui dénoncent l'Union européenne tout en engrangeant les « dividendes silencieux ».

Le Monde avec AFP

Publié le 17 avril 2024 à 20h55, modifié le 18 avril 2024 à 03h01 • ⏳ Lecture 2 min.

Ajouter à vos sélections



Les plus lus

- En direct, la situation politique en France : Elisabeth Borne siégera au sein du groupe Renaissance, son président n'est toujours pas élu
- L'avertissement de Gérard Larcher à Emmanuel Macron : « On ne joue pas avec la France »

Interrogé en marge de son déplacement à Bruxelles sur la campagne des élections européennes, Emmanuel Macron a ciblé, sans le nommer, le Rassemblement national (RN) dont la liste conduite par Jordan Bardella caracole en tête des sondages.

« Je pense qu'il y a beaucoup de peurs, d'inquiétudes dans le moment que nous vivons et que ces colères profitent toujours aux réponses les plus simplistes », a déclaré le chef de l'Etat français lors de son arrivée, mercredi 17 avril, dans la capitale belge avant la tenue d'un sommet européen, au cours d'un bref échange avec la presse aux côtés de la tête de liste de son camp aux européennes, Valérie Heyer. « Il y a aujourd'hui une espèce d'hypocrisie du débat et j'espère qu'en rentrant dans les prochaines semaines dans celui-ci, cette hypocrisie sera levée », a ajouté M. Macron, visant en creux le RN.

Il s'agissait de la première apparition publique du chef de l'Etat et de M^{me} Hayer depuis le début de la campagne, avant qu'ils n'assistent ensemble à une réunion du groupe Renew Europe.

Lire aussi | [Elections européennes 2024 : le calendrier du scrutin](#)

La candidate du camp présidentiel peine pour l'instant à imprimer sa marque dans la campagne. « On s'organise, on a la bonne candidate (...) On a surtout les bonnes idées, celles qui se sont mises en place en Europe ces dernières années », a lancé M. Macron, interpellé par un journaliste.

- =CLAUDEEXTRACT("people mentioned in the article, with titles if available. One per line.", B5)
- =CLAUDEEXTRACT("brief one-sentence summary", B5)
- =CLAUDEEXTRACT("English translation", E5)
- =CLAUDEEXTRACT("article category, one of politics, sports, international, other", B5)

B	C	E	F	G	H	I
content	sources	summary	language code	is it English?	English translation	article category
Ante inversores en Washington, Caputo defendió la estrategia fiscal y reafirmó el compromiso para levantar el cepo El jefe del Palacio de Hacienda participó de un seminario organizado por el banco J.P. Morgan WASHINGTON.- El ministro de Economía, Luis Caputo, llegó al seminario del banco de inversión J.P. Morgan en el Park Hyatt hotel poco después de las tres de la tarde, <small>hora local, un poco despu�s de comenzar en Washington</small> Elections europ�ennes : Emmanuel Macron dnonce « l'hypocrisie » du RN Le chef de l'Etat a dplor� mercredi depuis Bruxelles que les « peurs profitent aux r�ponses les plus simplistes » dans la campagne pour le scrutin du 9 juin, fustigeant ceux qui dnoncent l'Union europ�enne tout en engrangeant les « dividendes silencieux ».	Luis Caputo, Ministro de Econom�a Pablo Quirno, Secretario de Finanzas Javier Milei, gobierno Mauricio Macri, gobierno Nicol�s Posse, Jefe de Gabinete Alberto Fern�ndez, gobierno Pierre-Olivier Gourinchas, Economista Jefe del FMI	The economy minister defended the government's fiscal adjustment strategy and reiterated the intention to lift currency controls, but said they cannot do so with the current level of reserves <small>unless they receive a fresh injection of money.</small>	en	yes	The economy minister defended the government's fiscal adjustment strategy and reiterated the intention to lift currency controls, but said they cannot do so with the current level of reserves unless they receive a fresh injection of money.	international
Vous pouvez partager un article en cliquant sur les ic�nes de partage en haut de chaque article. 岸田首相、韓国・尹大統領に日米首脳会談の成果説明　日韓連携も確認 岸田文雄首相は17日、韓国の尹錫悦（ユンソンニヨル）大統領と電話で協議した。首相によると、10日にワシントンで行われた日米首脳会談での合意事項などについて説明したという。 協議後、首相は首相官邸で記者団に「日韓、日米韓の連携をより深化させ、日韓首脳間を含め緊密に意思疎通を図つて」の如きを述べた。	Emmanuel Macron, President of France Val�rie Heyer, Head of list for Macron's camp in European elections Jordan Bardella, Head of list for National Rally (Rassemblement National, RN) Rapha�l Glucksmann, Head of list for Socialist Party (Parti Socialiste) and Place Publique Gabriel Attal, Prime Minister of France	Emmanuel Macron a dplor� que les "peurs profitent aux r�ponses les plus simplistes", mercredi 17 avril.	fr	no	Emmanuel Macron lamented that "fears profit from the simplest answers", Wednesday April 17.	politics
	Fumio Kishida, Prime Minister; Yoon Suk-yeol, President of South Korea	岸田首相は17日、韓国の尹大統領と電話で協議し、日米首脳会談の成果を説明、日韓連携も確認した。	ja	no	Prime Minister Kishida held talks by phone with South Korean President Yoon on the 17th, explaining the results of the Japan-U.S. summit and confirming cooperation between Japan and South Korea.	other



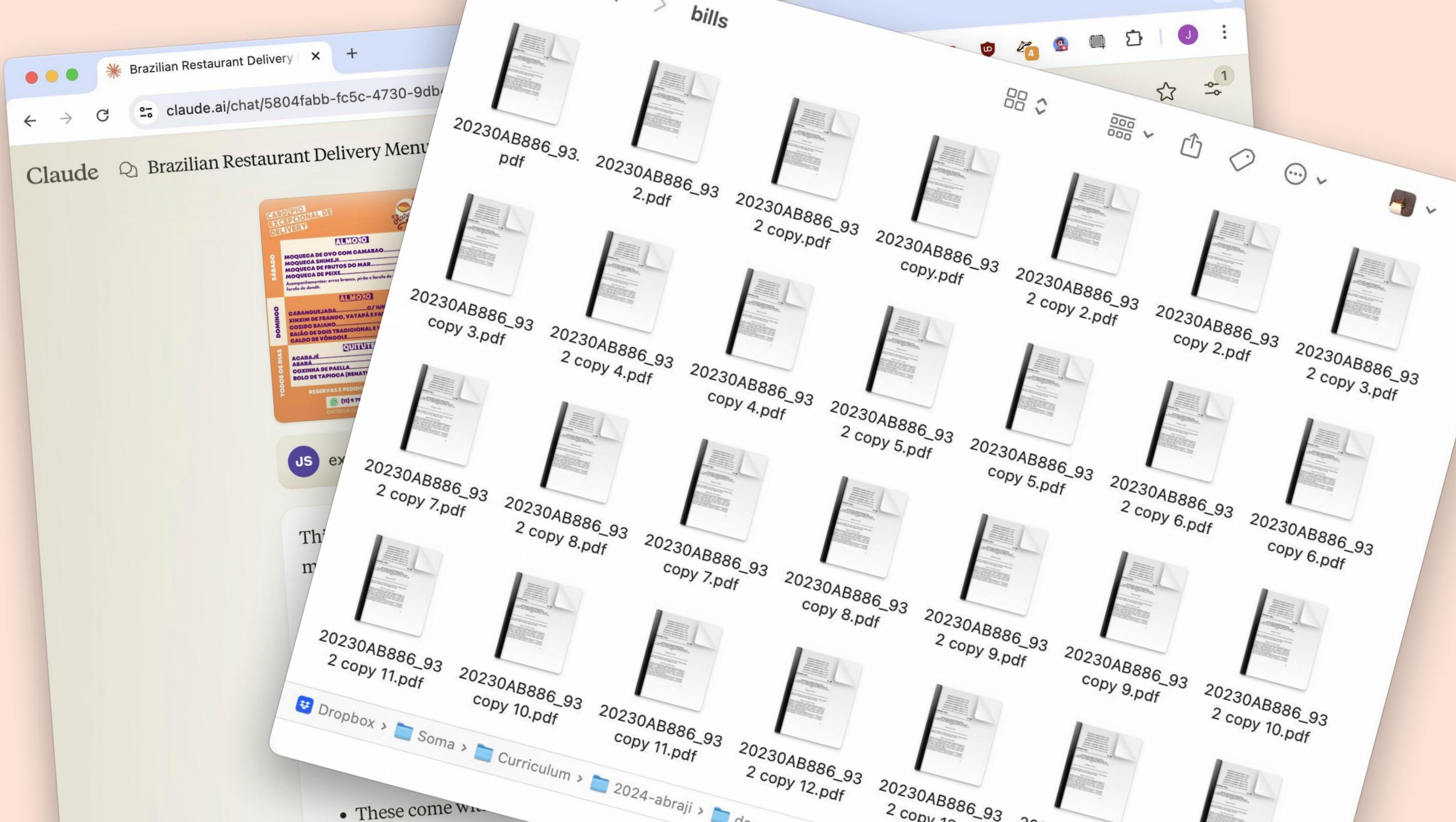
but yes, you must

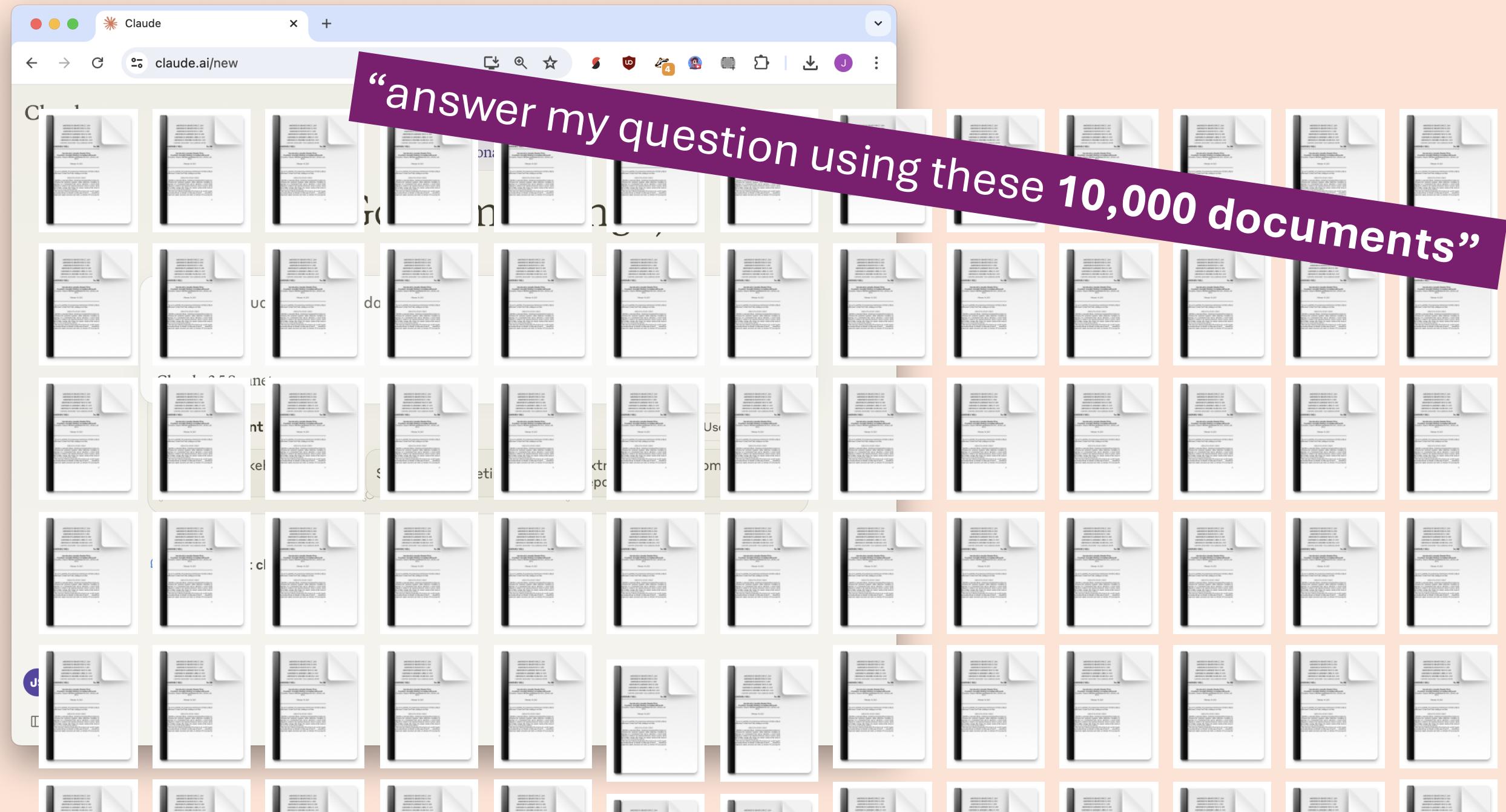
spot check results

sometimes you just want

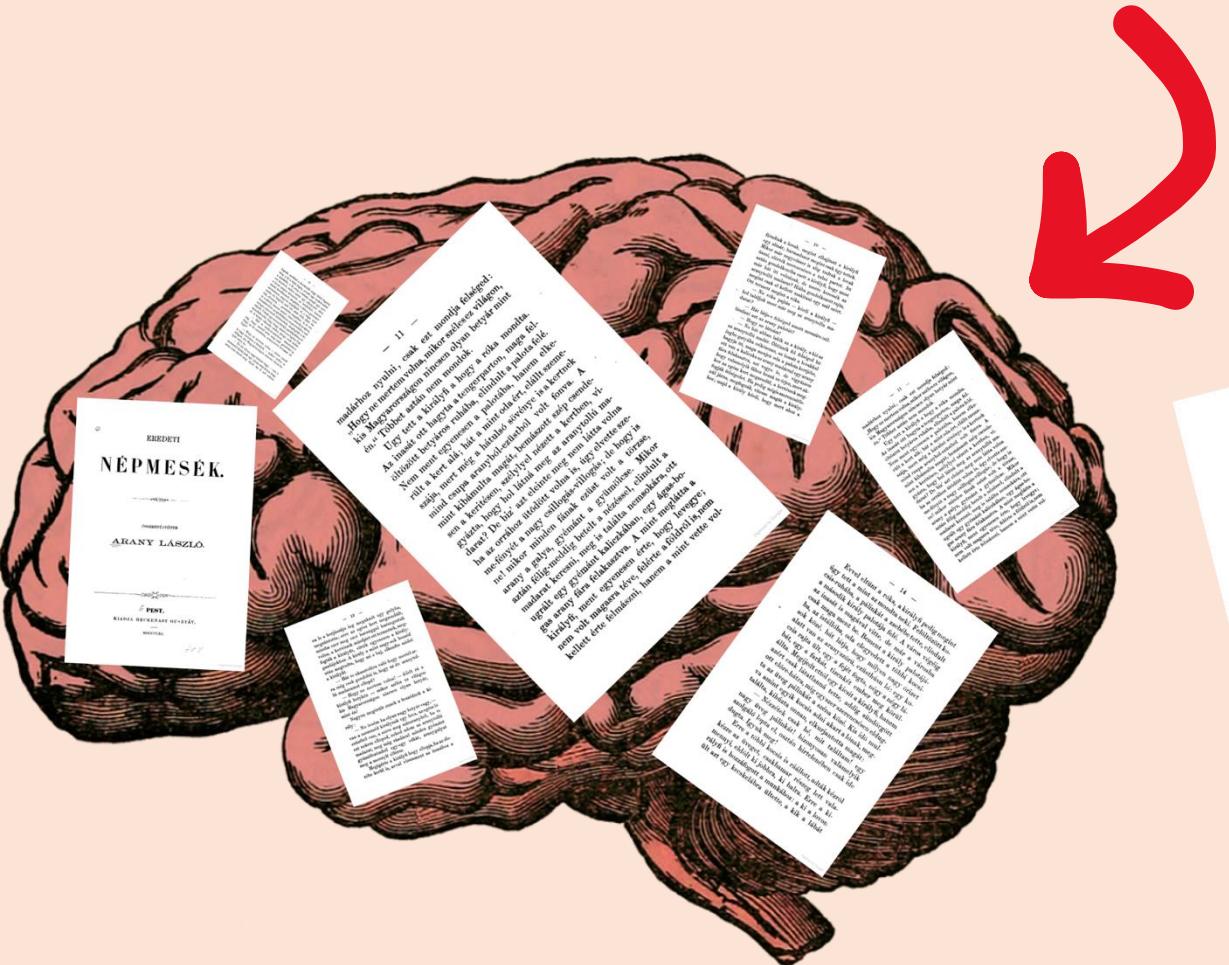
a single answer

from a lot of documents





Context window



— 19 —
akármint lesz, el kell neki lopni a kisa szonyt,
jobb hát ha most elszükteti, azután majd taná-
csot kérhet a rókától.

El is szökkéte szerencsén, ki-
zöre a királyfi inásához, ott, — mig az inas a
lovakat nyergelte, — a királyfi férj ment, el-
szakította az ötödik szál szörtsége is. Ott termett
a róka: „... bai felséges királyfi?

— Soh'sc' bújsájon azon a királyi, majd
az én azt, most majd együtt megyek fel.

Mentek aztán együtt négyen: a királyfi, a
szégeggel.

indent, az egész
szeméterek egymás-
ra is elbeszélte, hogy
igen gazdag király
kedve hozzá menni,
gy a hamut is manu-
az apjáinak mind nem
neuenj hozzá; „Most hár-
te az enyim, én a tied,
hanem az ejjel szoktunk
d országában megesküd-

írta két kezzel a tanácsos, után hármasnak, hogy még azon az ej- den mindenki lenyűgöz a királyszéknek, reggelre pedig, ha újra megugrik a határon. Ezzel a kissasszony szobrájából, várakor akkor jutott eszébe, hogy nekonyt a szomszéd kirdynak kell lovat megsokkolja, mert ha azt meg az aranytollal madár se lesz az óvédig nem akar az apja szeme elől, mert ezen szörnyű módon, már azt is meggy érte se meggy a kissasszonycér, eszébe jutott hogy akárhogyan lesz,

ellopop, neked adom az aranyoszörű lovat, a gyémánt szersámot, meg még ráadásul tiz köböl arany abrakot.

Jól van; megigérte a királyfi hogy ellopja, avval kiment a város végére az inasához. Itt megint mit volt mit tenni? el kellett szakítani a negyedik száj szörtsége. Ott termert a róka:
— No mi a baj felséges királyfi?
— Hát... —

— Hát csak e' meg e. Itt elbeszélte a királyi hogy már meg királykisasszonnyt kellenek lopni.

— No királyfi — mondja a róka, — most öltözök fel szakács ruhábá, menjen el ahol a királyhoz, szegődék be szakácsnak. Déliben majd felséleg vigye be a levest, hanem az előtőben ejts el. Többet nem.

hol vert

talking to a chatbot about a
lot of documents is called

retrieval augmented
generation

tech people call it R.A.G.

But I'm a little worried the answer
is going to be wrong!!!!

Did the AI actually find **relevant documents**?

Did we send **too few** sections, so it doesn't have the answer?

Did we send **too many** sections, confusing the AI?

Did it **hallucinate** even if it had the right content?

Chatbots May ‘Hallucinate’ More Often Than Many Realize

When summarizing facts, ChatGPT technology makes things up about 3 percent of the time, according to research from a new start-up. A Google system's rate was 27 percent.

MyCity Chatbot

chat.nyc.gov

NYC MyCity Official website of the City of New York

ChatbotBeta

FEEDBACK

MyCity Chatbot Beta

We are continuously working to improve the MyCity Chatbot, which uses business-related NYC agency data and AI to answer your questions. Your feedback is invaluable for refinement.

Example Questions

"How do I avoid noise violations and complaints for my construction company?"

"What information can you help me with?"

"How do I apply for the MWBE program?"

Capabilities

Trained to provide you with official NYC Business information.

Will not use the contents of your chat history to learn new information.

Responds to languages required by Local Law 30.

Limitations

May occasionally produce incorrect, harmful or biased content.

Limited knowledge of the world beyond NYC Business topics.

Trained to decline inappropriate requests.

© 2024 City of New York. All Rights Reserved.

Terms of Use Privacy Policy

Ask MyCity Chatbot business-related questions only

IMPORTANT: Responses may occasionally produce inaccurate or incomplete content. Validate answers on NYC.gov and MyCity Business.



Menu

The Mar

Challenging technology to ser

News

NYC's AI Chatbot Tells Business the Law

The Microsoft-powered bot says bosses can take
can discriminate based on source of income

By [Colin Lecher](#)

March 29, 2024 06:00 ET



? Question Submitted

Are buildings required to accept section 8 vouchers?

NYC Chatbot Answer (Emphasis by The Markup)

"No, buildings are not required to accept Section 8 vouchers."

Reality

Landlords [cannot discriminate by source of income](#), with a minor exception for small buildings where the landlord or their family lives.

Do landlords have to accept tenants on rental assistance?

"No, landlords are not required to accept tenants on rental assistance."

Landlords [cannot discriminate by source of income](#), with a minor exception for small buildings where the landlord or their family lives.

Can I keep my funeral home pricing confidential?

"Yes, you can keep your funeral home pricing confidential. There are no specific regulations that mandate the disclosure of funeral home pricing in New York City."

[The Federal Trade Commission](#) has outlawed concealing funeral prices.

Can I make my store cashless?

"Yes, you can make your store cashless in New York City. There are no specific requirements that prohibit businesses from operating as cashless establishments in the city."

Stores have been [required to accept cash](#) as payment since 2020.

RAG is very irresponsible
(especially facing readers!)
but it's *also* very easy and fun

so let's do it!

Llamaindex - Llamaindex

docs.llamaindex.ai/en/stable/#getting-started

Llamaindex

Getting Started

Llamaindex is available in Python (these docs) and [TypeScript](#). If you're not sure where to start, we recommend reading [how to read these docs](#) which will point you to the right place based on your experience level.

30 second quickstart

Set an environment variable called `OPENAI_API_KEY` with an [OpenAI API key](#). Install the Python library:

```
pip install llama-index
```

Put some documents in a folder called `data`, then ask questions about them with our famous 5-line starter:

```
from llama_index.core import VectorStoreIndex, SimpleDirectoryReader
documents = SimpleDirectoryReader("data").load_data()
index = VectorStoreIndex.from_documents(documents)
query_engine = index.as_query_engine()
response = query_engine.query("Some question about the data should go here")
print(response)
```

If any part of this trips you up, don't worry! Check out our more comprehensive starter tutorials using [remote APIs like OpenAI](#) or [any model that runs on your laptop](#).

LlamaCloud

Table of contents

- Introduction
 - What is context augmentation?
- Llamaindex is the Data Framework for Context-Augmented LLM Apps
- Use cases
 - Who is Llamaindex for?
- Getting Started
- 30 second quickstart
- LlamaCloud
- LlamaParse
- Community
- Getting the library
- Contributing
- Related projects

AnythingLLM | The ultimate AI business intelligence tool.

useanything.com

AnythingLLM

AnythingLLM Cloud Docs Github Download AnythingLLM

The last chatbot you will ever need

*The ultimate AI business intelligence tool.
Any LLM, any document, full control, full privacy.*

[Download AnythingLLM for Desktop](#)

for MacOS, Linux, and Windows operating systems

AnythingLLM: The Only Document Chatbot You Need



[New Workspace](#)[test](#)

upload documents!

Name

 custom-documents everything adding-zip-codes-to-lat-long.md

Cached

 annotations

Cached

 annotations.md

Cached

 arcgis-rest.md

Cached

 auto-updating-scaper-viz.md

Cached



Click to upload or drag and drop

supports text files, csv's, spreadsheets, audio files, and more!

or submit a link

Fetch website

These files will be uploaded to the document processor running on this AnythingLLM instance.
These files are not sent or shared with a third party.



GitHub Repo

Import an entire public or private Github repository in a single click.



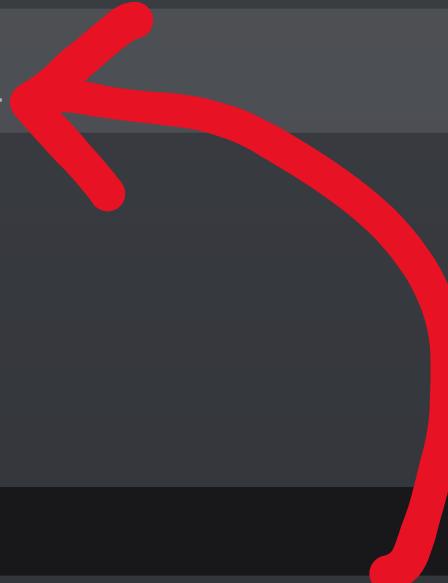
YouTube Transcript

Import the transcription of an entire YouTube video from a link.



Bulk Link Scraper

Scrape a website and its sub-links up to a certain depth.



Website URL

URL of the website you want to scrape.

https://example.com

Depth

This is the number of child-links that the worker should follow from the original page.

1

Max Links

Maximum number of links to scrape.

20

scrape websites!

Submit

do a better job
than an NYC
agency that was
paid a lot of
money!

Anything LLM

+ New Workspace

Can I deny renters based on Section 8 vouchers

No, you cannot deny renters based on Section 8 vouchers. It is against the law to discriminate based on a current or prospective tenant's source of income, including housing vouchers.

Show Citations >

but can i consider credit scores?

If the voucher/subsidy covers 100% of the rent, you are prohibited from considering credit. However, if the tenant pays a portion of the rent, you may consider credit. Each application should be considered on a case-by-case basis.

Show Citations >

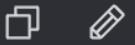
Send a message

/ @ T





Can I deny renters based on Section 8 vouchers



No, you cannot deny renters based on Section 8 vouchers. It is against the law to discriminate based on a tenant's source of income, including Section 8.



Hide Citations ▾



SourceOfIncomeFact...eet.pdf



FairH...



Section 8 Guide fo...nts.pdf

The screenshot shows a PDF document titled "FairHouse_FAQBroker-2021Eng.pdf". The document is dated March 2023 and is from NYC.gov/HumanRights (@NYCCHR). It contains a section titled "BEST PRACTICES FOR LICENSED SALESPERSONS AND BROKERS TO AVOID SOURCE OF INCOME DISCRIMINATION". The text discusses the responsibility of brokers for their agents' actions and advises against maintaining separate lists for voucher holders, as it constitutes aiding and abetting discrimination.

March 2023
NYC.gov/HumanRights
@NYCCHR

1

BEST PRACTICES FOR LICENSED SALESPERSONS AND BROKERS TO AVOID SOURCE OF INCOME DISCRIMINATION

Yes. You are responsible for the actions of anyone who works under your broker's license or holds themselves out as an agent of your brokerage company. What should I do if an owner refuses to accept housing vouchers/subsidies? You should not work with owners who discriminate based on a current or prospective tenant's source of income. Source of income discrimination must be treated just like any other kind of discrimination. It is against the law. Can I keep a separate list ("waitlist") of apartments for voucher holders? No. Maintaining separate lists of apartments for voucher/subsidy recipients is a discriminatory practice. Steering applicants to specific buildings not only constitutes discrimination, but may also be considered aiding and abetting the discrimination of owners.

60% match

Citations aren't reliable, so be sure to fact-check!

Evaluating Verifiability in Generative Search Engines

Nelson F. Liu* Tianyi Zhang Percy Liang

Department of Computer Science
Stanford University

nfliu@cs.stanford.edu



Nelson Liu
@nelsonfliu

...

Generative search engines are transforming how we find info, but are they trustworthy?

We evaluate Bing Chat, NeevaAI, [perplexity.ai](#) & YouChat—only 52% of statements are supported by cites and 75% of cites actually support their statements.

arxiv.org/abs/2304.09848



What are the latest discoveries from the James Webb Space Telescope?



The James Webb Space Telescope is designed to peer into the dusty clouds of gas where stars and planetary systems are born. Webb has captured the first direct image of an exoplanet, and the Pillars of Creation in the Eagle Nebula[1][2]. Additionally, the telescope will be used to study the next interstellar interloper[3].

(*Some generated statements may not be fully supported by citations, while others are fully supported.)

Cited Webpages

- [1]: nasa.gov (✖citation does not support its associated statement)
[NASA's Webb Confirms Its First Exoplanet](#)
... Researchers confirmed an exoplanet, a planet that orbits another star, using NASA's James Webb Space Telescope for the first time. ...
- [2]: cnn.com (⚠citation partially supports its associated statement)
[Pillars of Creation: James Webb Space Telescope ...](#)
... The Pillars of Creation, in the Eagle Nebula, is a star-forming region captured in a new image (right) by the James Webb Space Telescope that reveals more detail than a 2014 image (left) by Hubble ...
- [3]: nasa.gov (✓citation fully supports its associated statement)
[Studying the Next Interstellar Interloper with Webb](#)
... Scientists have had only limited ability to study these objects once discovered, but all of that is about to change with NASA's James Webb Space Telescope...The team will use Webb's spectroscopic capabilities in both the near-infrared and mid-infrared bands to study two different aspects of the interstellar object.



How I convinced GPT to teach me about **Hungarian folktales** (without speaking a word of Hungarian)

This is both interesting and related to journalism, *I promise.*

Jonathan Soma

Knight Chair in Data Journalism, Columbia University

js4571@columbia.edu

@dangerscarf

<https://github.com/jsoma/mediaparty-folktales>

"What did **Zsuzska**
steal from the devil?"



Semantic search

De **Zsuzska** nem adta: „Tán bolond vagyok, hogy visszaadjam, mikor kívülvagyok már vele az udvaron!“ Az **ördög** kergette egy darabig, de séhogyse tudta utolérni, utoljára is visszafordult, **Zsuzska** pedig mentegyenesen a király előlbe, od'adta neki az arany fej káposztát. – No felséges király elhoztam már ezt is! A két nénjéi **Zsuzskának**, majd hogy még nem ütötte a guta, másnap megintbementek a királyhoz. – Jaj felséges király van még annál ördögnek egy arany kis gyermekes arany bőlcsoben, **Zsuzska** beszélt fűnek-tának, hogy ő azt is el tudná lopni. Megint behívta király **Zsuzskát**. – Fiam **Zsuzska**, azt hallottam, hogy van anna ördögnek egy arany kisgyermeke is, arany bőlcsoben, te azt tudod lopni, azt beszélted, azért ha az éjjel el nem lopod, halálával halsz meg.

Using the texts below, answer the question

"what did Zsuzska steal from the devil?"



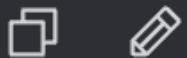
Hiába tagadta szegény Zsuzska, nem használt semmit, elindult hát nagyszomorúan. Épen éjfél volt, mikor az ördög házához ért, aludt az ördögis, a felesége is. Zsuzska csendesen belopódzott, ellopta a tenger-ütőpálczát, avval bekiáltott az ablakon. – Hej ördög, viszem ám már a tenger-ütő pálczádat is. – Hej kutya Zsuzska, megölettek három szép lyányomat, elloptad atenger-lépő czipőmet, most viszed a tenger-ütő pálczámat, de majd meglakolsz te ezért. Utána is szaladt, de megint csak a tengerparton tudott közel jutnihozzá, ott meg Zsuzska megütötte a tengert a tenger-ütő pálczával, kétfelé vált előtte, utána meg összecsapódott, megint nem foghatta megaz ördög. Zsuzska ment egyenesen a királyhoz.

De Zsuzska nem adta; „Tán bolond vagyok, hogy visszaadjam, mikor kivülvagyok már vele az udvaron?!” Az ördög kergette egy darabig, de sehogyse tudta utolérni, utoljára is visszafordult, Zsuzska pedig mentegyenesen a király elibe, od'adta neki az arany fej káposztát. – No felséges király elhoztam már ezt is! A két nénjét Zsuzskának, majd hogy meg nem ütötte a guta, mikormegtudták, hogy Zsuzskának most se' lett semmi baja, másnap megint bementek a királyhoz. – Jaj felséges király van még annak az ördögnek egy arany kis gyermekelis arany bölcsőben, Zsuzska azt beszéli fűnek-fának, hogy ő azt is eltudná lopni. Megint behívatta a király Zsuzskát. – Fiam Zsuzska, azt hallottam, hogy van annak az ördögnek egy arany kisgyermek is, arany bölcsőben, te azt is el tudod lopni, azt beszélted, azért ha az éjjel el nem lopod, halálnak halálával halsz meg.

– No felséges király, elhoztam már a tengerütő pálczát is. A király még jobban megszerette Zsuzskát, hogy olyan életre való, de anénjei még jobban irigykedtek rá, csakhamar megint avval árulták be, hogy van annak az ördögnek egy arany fej káposztája is, Zsuzska azt isel tudná lopni, azt mondta. A király megint ráparancsolt Zsuzskára erősparancsolattal, hogy ha a káposztát el nem lopja, halálnak halálával hal meg. Elindult hát szegény Zsuzska megint, el is ért szerencsésen épen éjfélre az ördög kertjibe, levágta az arany fej káposztát, avval bekiáltott az ablakon. – Hej ördög, viszem ám már az arany fej káposztádat is. – Hej kutya Zsuzska, megölettek három szép lyányomat, elloptad atenger-lépő czipőmet, elloptad a tenger-ütő pálczámat, most viszed az arany fej káposztámat, csak ezt az egyet add vissza, soha szemedre sevetem.



What did Zsuzska steal from the devil? Write a short list.



01. Tenger-lépő cipő (sea-stepping shoes)
02. Tenger-ütő pálca (sea-striking rod)
03. Arany fej káposzta (golden-headed cabbage)
04. Arany kisgyermek arany bölcsőben (golden child in a golden cradle)
05. Zsák arany dió (sack of golden nuts)



Show Citations >

A szegény kis macska meghunyászkodott, félre ült a kuczikba, a lyány meg a patkán eszegetett. A mint ott eszeget, a mint ott eszeget. egyszer csak megzörgetik az ajtót:

- Tanyér talpam, lompos farkam, szép lyány mátkám nyiss a j-

Akkor már oda szaladt ám a lyány a kis macsk

- Jaj cziczuskám-miczuskám mit csináljak?!

- Hát ne ereszd be.

A lyán
megi

- Hej ördög, viszem ám a tenger-lépő cipődet.

- Hej kutya Zsuzska, megölettek három szép lyányomat, most viszed a tenger-lépő cipőmet, hanem majd meglakolszmeg te ezért, fogadom azt az egyet!

Kergette is az ördög jó darabig, már im hogy utól nem érte, de épen egy tenger mellé jutottak, ott Zsuzska felhúzta a tenger-lépő cipőt, átlépte vele a tengert. Avval ment egyenesen a királyhoz.

- No felséges király, elhoztam már a tenger-lépő cipőt!

verify, verify, verify

ZSUZSKA ÉS AZ ÖRDÖG.

Egyszer volt, hol nem volt, még az óperenciás tengeren is túl volt, volt a sós tenger kellő-közepén egy lakatlan sziget, azon a lakatlan szigeten egy nagy hegy, annak a heaven beszédei,

tetején egy magas torony, abból szedtem én ezen szent

szony, annak volt három evelte, de hogy nagyon hanem elküldte

JS

You

where did he get his masters?



ChatGPT

Jonathan Soma earned his Master's degree from the University of Denver.



JS

You

Where did he get his master's degree?



ChatGPT

Jonathan Soma earned his Master's degree from the Columbia University Graduate School of Journalism [1].

and finally, time for

transcription

because you hate typing out
your interviews

The New York Times

Elon Musk Sues OpenAI and Sam Altman for Violating the Company's Principles

Musk said the prominent A.I. start-up had put profits and commercial interests ahead of seeking to benefit humanity.

Share full article

328



ChatGPT

ChatGPT 4

You

How many n's are in mayonnnnase?

ChatGPT

The word "mayonnaise" typically has two 'n's in it. However, in the way you've spelled it—"mayonnnnase"—there are three 'n's.

AI

GPT...

ChatGPT can make mistakes. Consider checking important information.

September 21, 2022

Introducing Whisper

We've trained and are open-sourcing a neural net called Whisper that approaches human level robustness and accuracy on English speech recognition.

[Read paper ↗](#)

[View code ↗](#) [View model card ↗](#)

Models - Hugging Face x +

huggingface.co/models?sort=trending&sear... ☆

Hugging Face Search models, datasets, users...

Models 292

openai/whisper new Full-text search Add filters Sort: Trending

-  **openai/whisper-large-v3**
Automatic Speech Recognition • Updated Feb 8 • ↓ 2.01M • ❤ 2.41k
-  **openai/whisper-large-v2**
Automatic Speech Recognition • Updated Feb 29 • ↓ 306k • ❤ 1.56k
-  **openai/whisper-tiny**
Automatic Speech Recognition • Updated Feb 29 • ↓ 216k • ❤ 180
-  **openai/whisper-base**
Automatic Speech Recognition • Updated Feb 29 • ↓ 359k • ❤ 159
-  **openai/whisper-small**
Automatic Speech Recognition • Updated Feb 29 • ↓ 368k • ❤ 152
-  **openai/whisper-medium**
Automatic Speech Recognition • Updated Feb 29 • ↓ 188k • ❤ 174
-  **openai/whisper-tiny.en**
Automatic Speech Recognition • Updated Jan 22 • ↓ 170k • ❤ 77

Private, h 100K

potential? Cause it feels to me like previous generation learned. Yeah, of self-awareness, competing prior being clear about what you're optimizing one's out there trying to make a point their time or be ineffective. I think it we were previously optimizing for the moment. Or maybe we've entered uptake. We kind of are still inheriting following and we haven't upgraded our really want is a little bit different to be praised for or what the people are

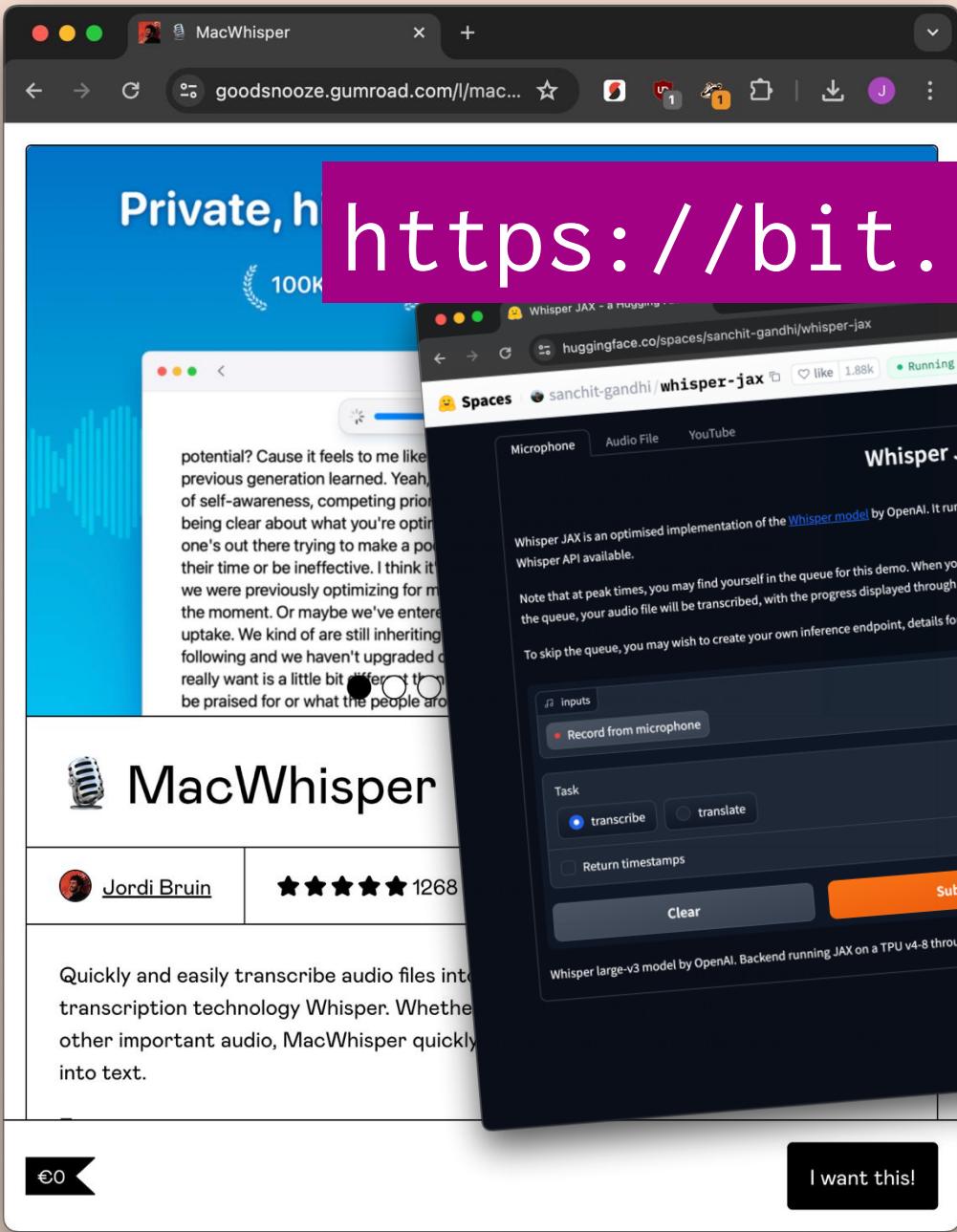
MacWhisper

Jordi Bruin ★★★★★ 1268

Quickly and easily transcribe audio files into transcription technology Whisper. Whether other important audio, MacWhisper quickly into text.

€0 ↗

I want this!



Whisper JAX - a Hugging Face Space

Spaces sanchit-gandhi/whisper-jax like 1.88k • Running

Microphone Audio File YouTube

Whisper JAX: The Fastest Whisper API ⚡

Whisper JAX is an optimised implementation of the [Whisper model](#) by OpenAI. It runs on JAX with a TPU v4-8 in the backend. Compared to PyTorch on an A100 GPU, it is over [70x faster](#), making it the fastest Whisper API available.

Note that at peak times, you may find yourself in the queue for this demo. When you submit a request, your queue position will be shown in the top right-hand side of the demo pane. Once you reach the front of the queue, your audio file will be transcribed, with the progress displayed through a progress bar.

To skip the queue, you may wish to create your own inference endpoint, details for which can be found in the [Whisper JAX repository](#).

Inputs
Record from microphone

Task
 transcribe translate

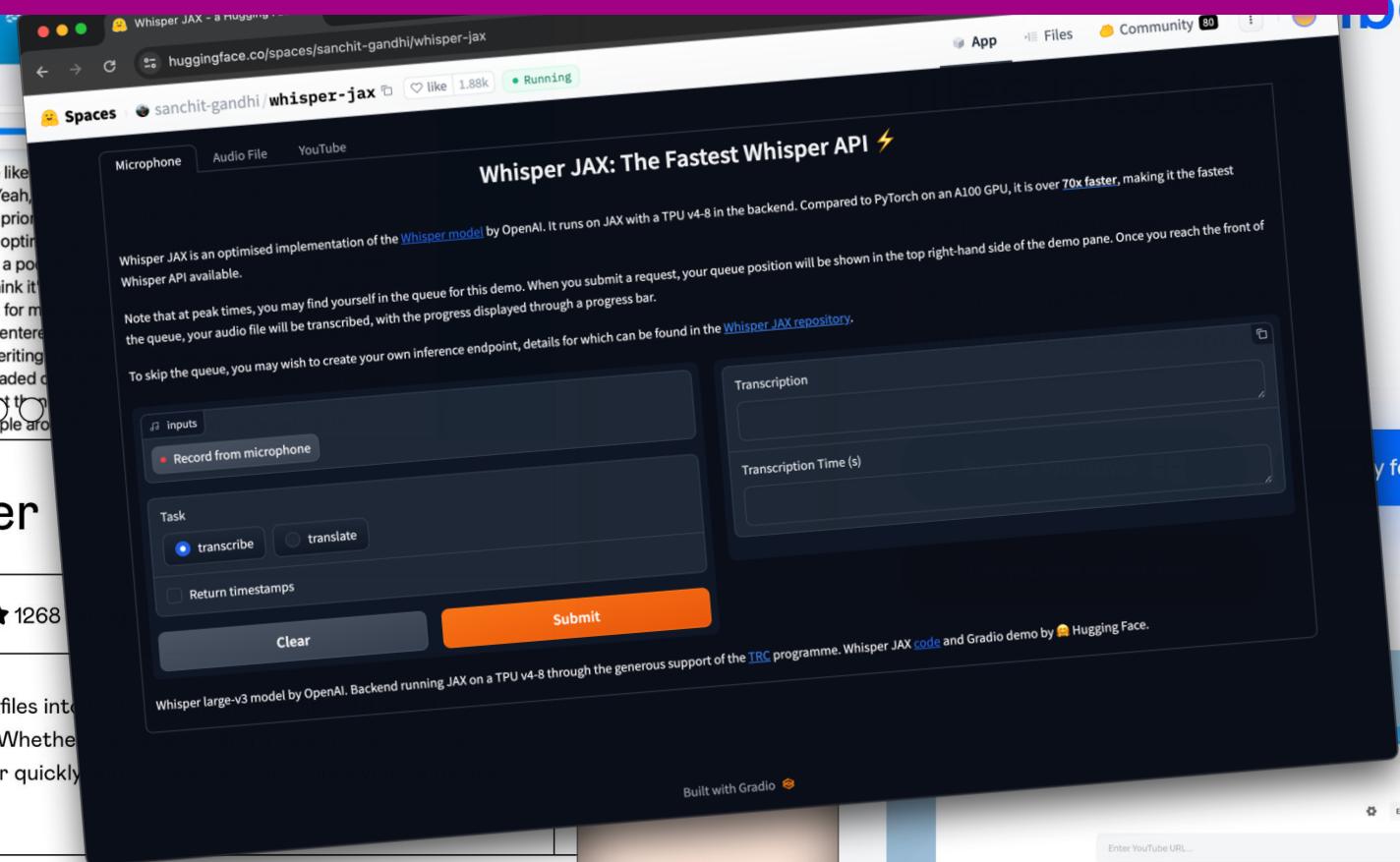
Return timestamps

Clear Submit

Transcription
Transcription Time (s)

Whisper large-v3 model by OpenAI. Backend running JAX on a TPU v4-8 through the generous support of the [TRC programme](#). Whisper JAX code and Gradio demo by 🤗 Hugging Face.

Built with Gradio 🤗



EasyWhisper | Easily transcript... easywhisper.io

Features Prices FAQs

Transcribe audio

Buy for Mac OS macOS

Enter YouTube URL... English Medium (English)

My files





MacWhisper



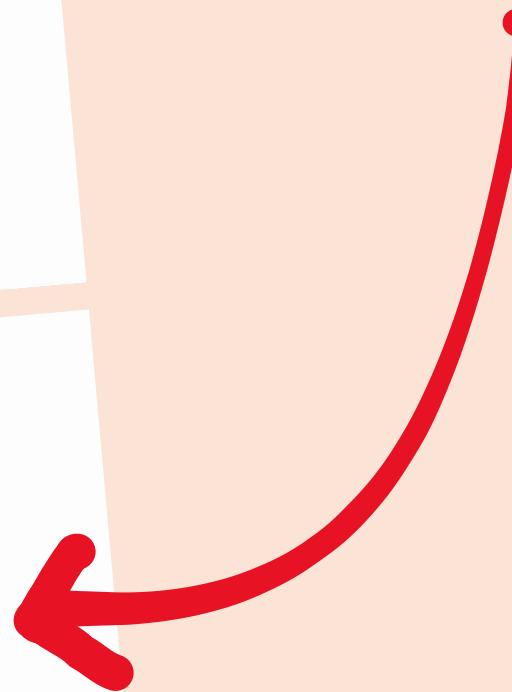
Jordi Bruin

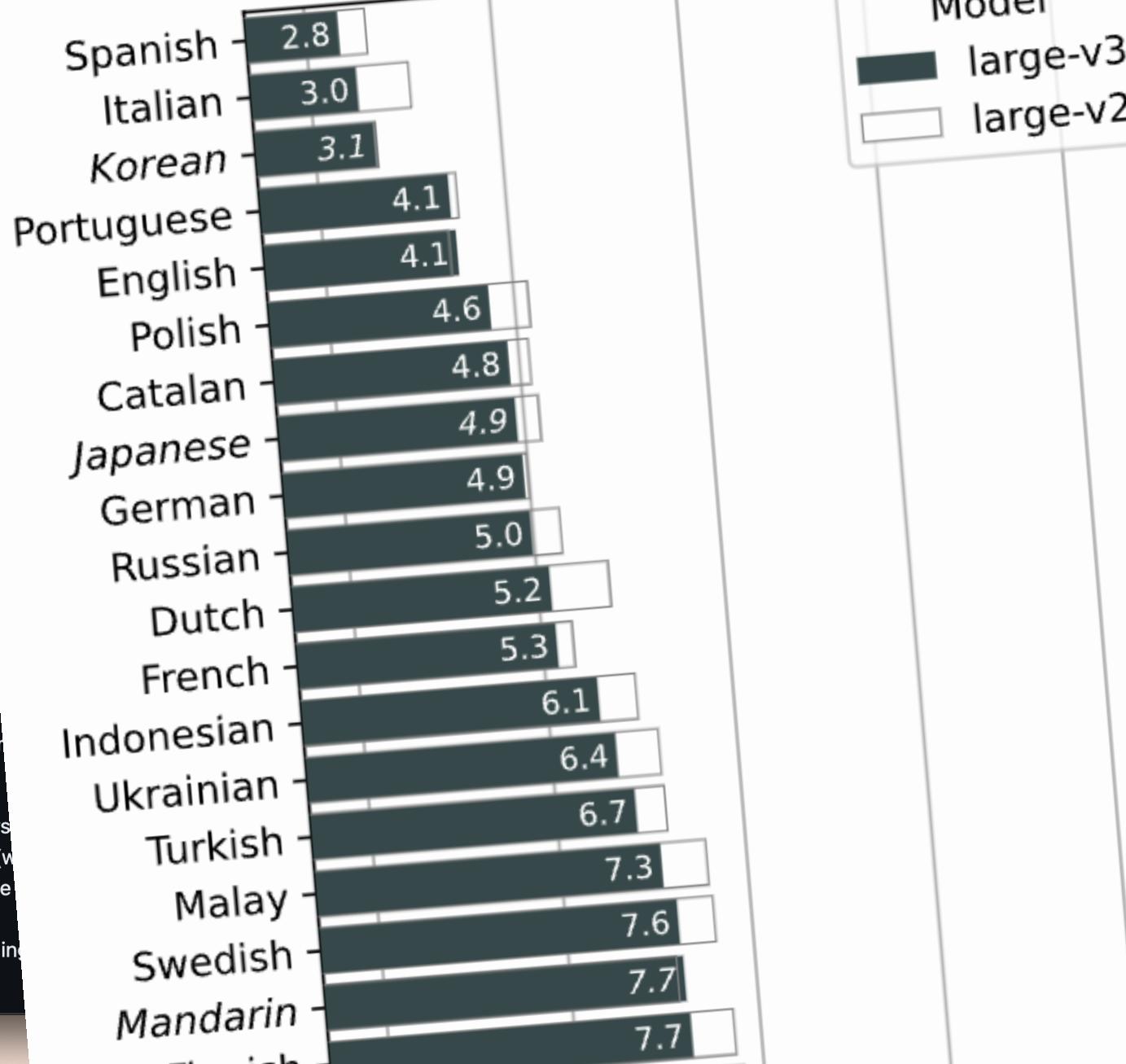
★★★★★ 1268 ratings

Quickly and easily transcribe audio files into text with OpenAI's state-of-the-art transcription technology Whisper. Whether you're recording a meeting, lecture, or other important audio, MacWhisper quickly and accurately transcribes your audio files into text.

Features

- Easily record and transcribe audio files on your Mac
- Just drag and drop audio files to get a high quality transcription
- Supported formats: mp3, wav, m4a, ogg, opus, mov and mp4 videos.
- Adjust whisper settings (beam search / greedy, beam size etc)
- Supports Tiny, Base and Small transcription qualities (and you can compare the transcription quality between all quality levels)





Available models and languages

There are five model sizes, four with English-only versions, offering speed and accuracy tradeoffs. Below are the names of the available models and their approximate memory requirements and inference speed relative to the large model; actual speed may vary on many factors including the available hardware.

Size	Parameters	English-only model	Multilingual model	Required VRAM	RAM
tiny	39 M	tiny.en	tiny	~1 GB	~100 MB
base	74 M	base.en	base	~1 GB	~100 MB
small	244 M	small.en	small	~2 GB	~100 MB
medium	769 M	medium.en	medium	~5 GB	~100 MB
large	1550 M	N/A	large	~10 GB	~100 MB

The .en models for English-only applications tend to perform better, especially for the tiny.en and base.en models. We observed that the difference becomes less significant for the small.en and medium.en models.

Whisper's performance varies widely depending on the language. The figure below shows the performance breakdown of large-v3 and large-v2 models by language, using WERs (word error rates) or CER (character error rates, shown in *italic*) evaluated on the Common Voice and Fleurs datasets. Additional WER/CER metrics corresponding to the other models and datasets can be found in Appendix D.1, D.2, and D.4 of [the paper](#), as well as the BLEU (Bilingual Evaluation Understudy) scores for translation in Appendix D.3.

Racial disparities in automated speech recognition

Allison Koenecke^a , Andrew Nam^b, Emily Lake^c, Joe Nudell^d, Minnie Quartey^e, Zion Mengesha^c, Connor Toups^c, John R. Rickford^c, Dan Jurafsky^{c,f}, and Sharad Goel^{d,1} 

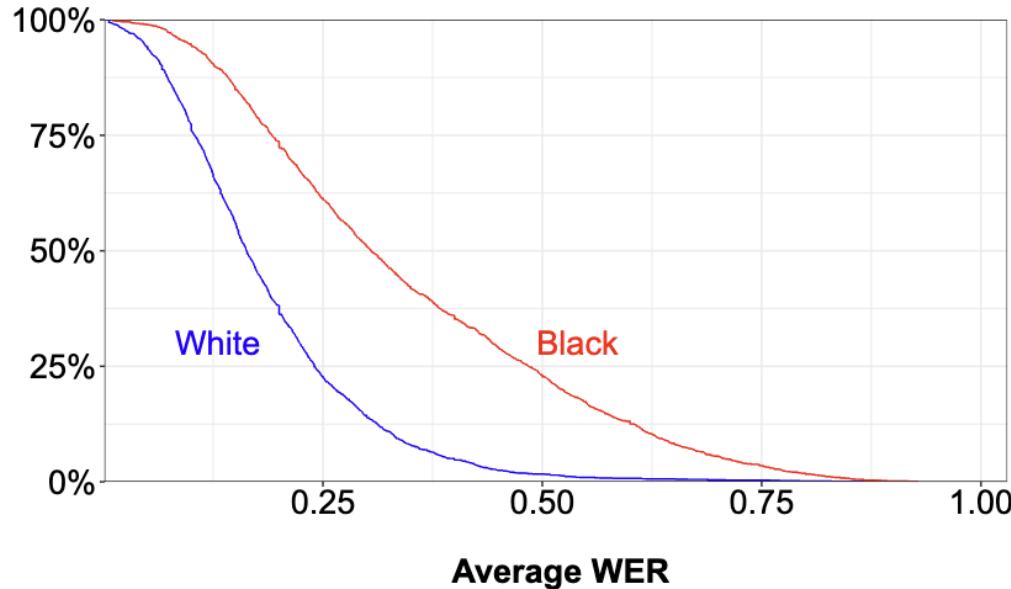


Fig. 2. The CCDF denotes the share of audio snippets having a WER greater than the value specified along the horizontal axis. The two CCDFs shown for audio snippets by white speakers (blue) versus those by black speakers (red) use the average WER across the five ASR services tested. If we assume that a $\text{WER} > 0.5$ implies a transcript is unusable, then 23% of audio snippets of black speakers result in unusable transcripts, whereas only 1.6% of audio snippets of white speakers result in unusable transcripts.



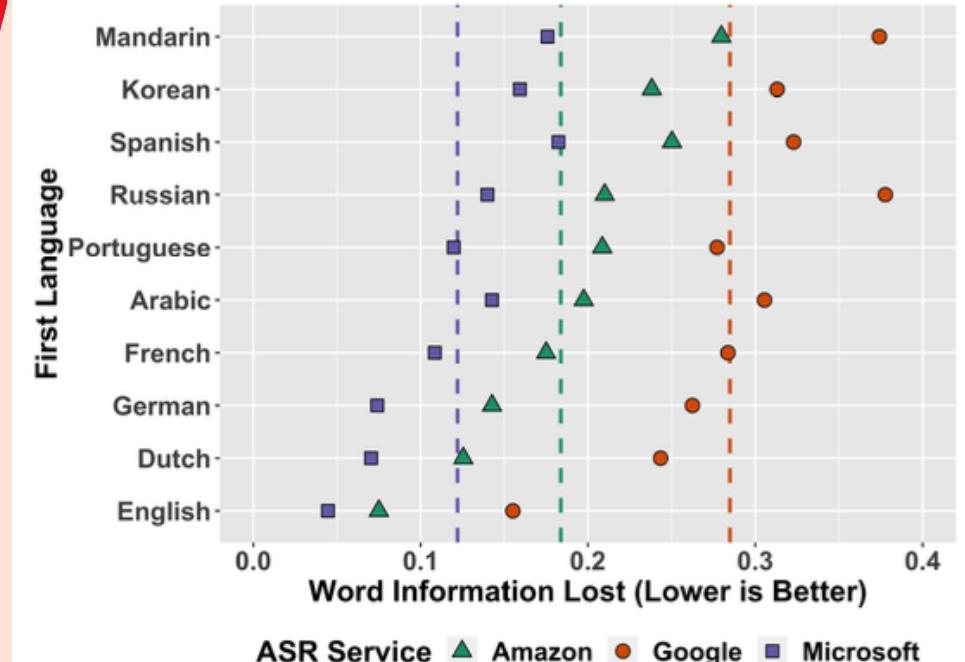
Performance Disparities Between Accents in Automatic Speech Recognition

Alex DiChristofano
a.dchristofano@wustl.edu
Division of Computational & Data Sciences
Washington University in St. Louis
St. Louis, Missouri, USA

Shefali Chandra
sc23@wustl.edu
Department of Women, Gender and Sexuality Studies and
Department of History
Washington University in St. Louis
St. Louis, Missouri, USA

Henry Shuster
henryshuster@wustl.edu
Department of Computer Science & Engineering
Washington University in St. Louis
St. Louis, Missouri, USA

Neal Patwari
npatwari@wustl.edu
Department of Electrical & Systems Engineering and
Department of Computer Science & Engineering
Washington University in St. Louis
St. Louis, Missouri, USA



<https://bit.ly/birn-data>

Jonathan Soma
Columbia University
js4571@columbia.edu
@dangerscarf

AI for Reporting

tricks for the newsroom