

Jonathan Soma
js4571@columbia.edu
@dangerscarf

AI for Investigative Reporting

tricks for the newsroom

The background is a complex, repeating pattern on a dark blue field. It features stylized birds in shades of yellow, orange, and white, interspersed with various flowers in pink, white, and green. Small red strawberries with green leaves are also scattered throughout the design. The overall style is reminiscent of traditional folk art or textile patterns.

A.l.i.s

magic **patterns**

I`m

????????????

sleepy?

hungry?

Soma?

I haven't eaten since

yesterday, I'm starving

Dearest friend, I daresay I
have not partaken of food
in ages. I`m positively

famished



Playground

Complete

Your presets

Save

View code

Share



Dearest friend, I daresay I have not partaken of food in ages. I'm positively



i Completion models are now considered legacy. [Try our latest models](#)



Submit



21

Playground

Complete 

Your presets 

Save

View code

Share



Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished! Would you care to join me for a meal? I know of a charming little bistro just down the street. Their pastries are simply divine. What do you say? Shall we indulge in a feast fit for kings and queens?



Playground

Complete

Your presets

Save

View code

Share



Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished!



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 fury and delight,

```
# 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

low-quality output is ignorable

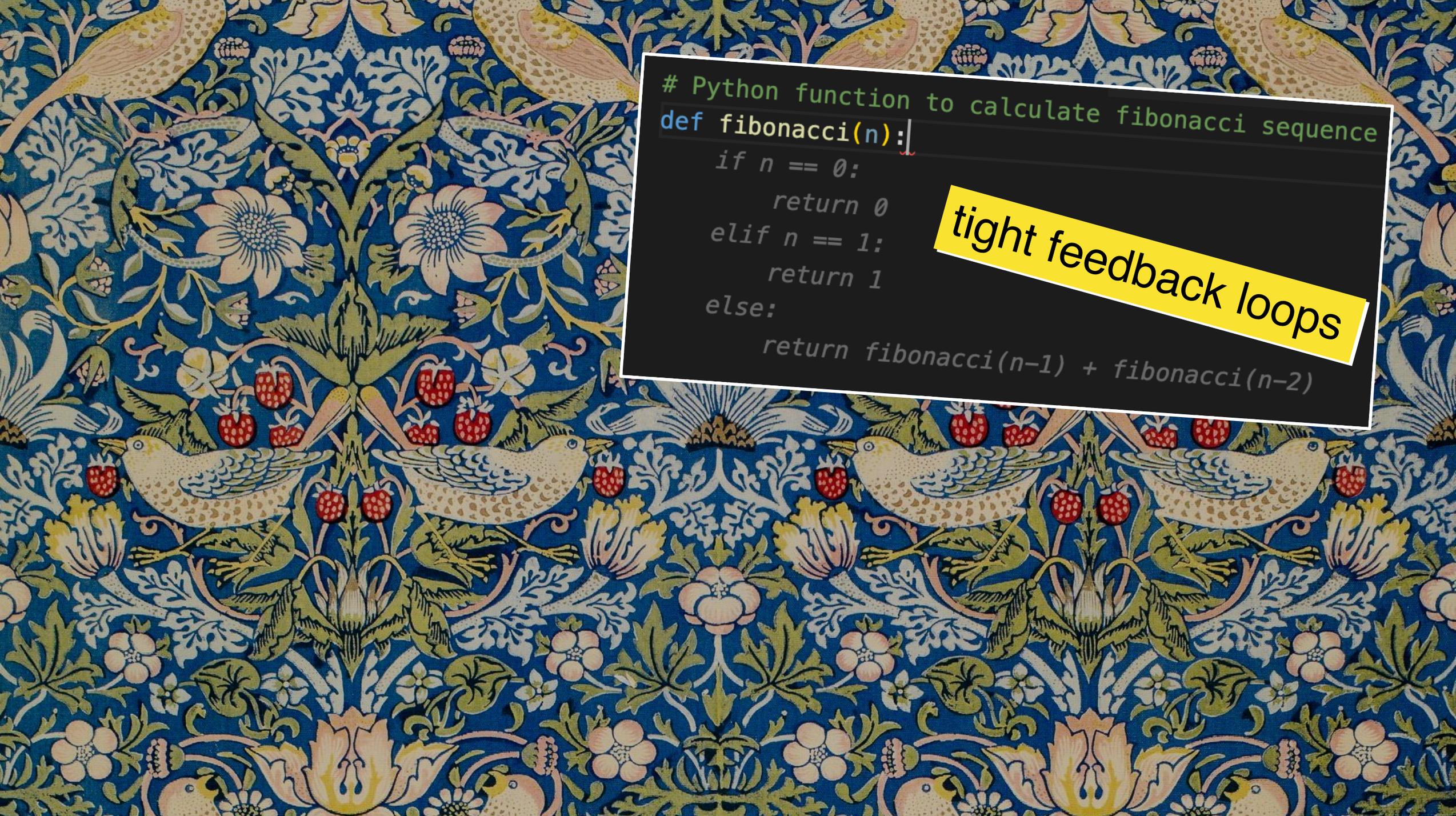
errors are expected

New York Sessions

Year	Session	Introduced	Completed	Legislation
2023-2024	General Assembly	20091	3380	Browse

(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



Handling Pagination

Since the results span multiple pages, we'll need

```
python

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

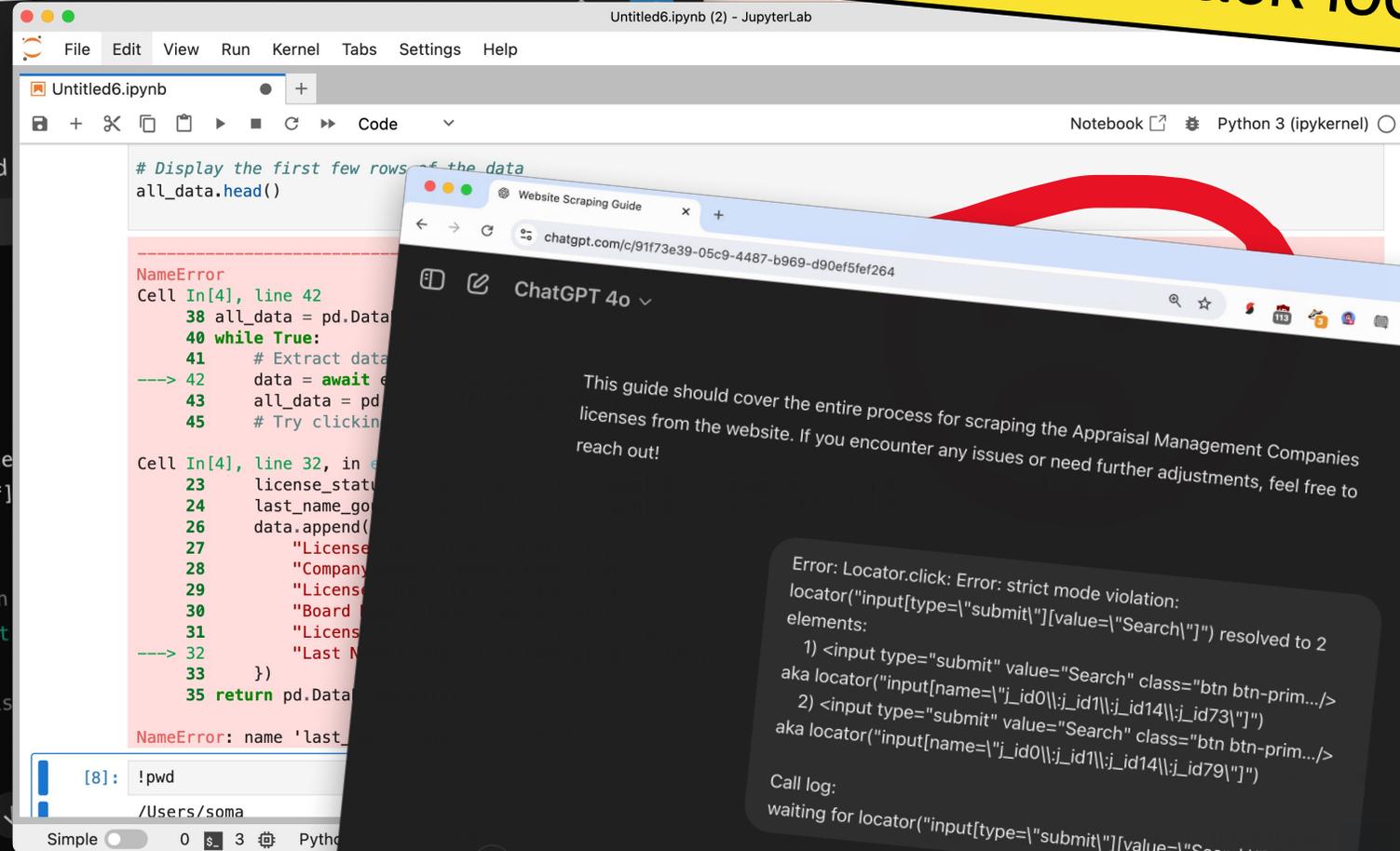
while True:
    # Scrape the data from the current
    df = await scrape_current_page(page)
    all_data = pd.concat([all_data, df])

    try:
        # Click the "Next Page" button
        await page.locator("a:has-text(
    except:
        # If the "Next Page" button is
        break

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

Message ChatGPT

ChatGPT can make mistakes. Check important info.



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-prim.../>
  aka locator("input[name='j_id0\\;j_id1\\;j_id14\\;j_id73']")
  2) <input type="submit" value="Search" class="btn btn-prim.../>
  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.

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 rage and spite.

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

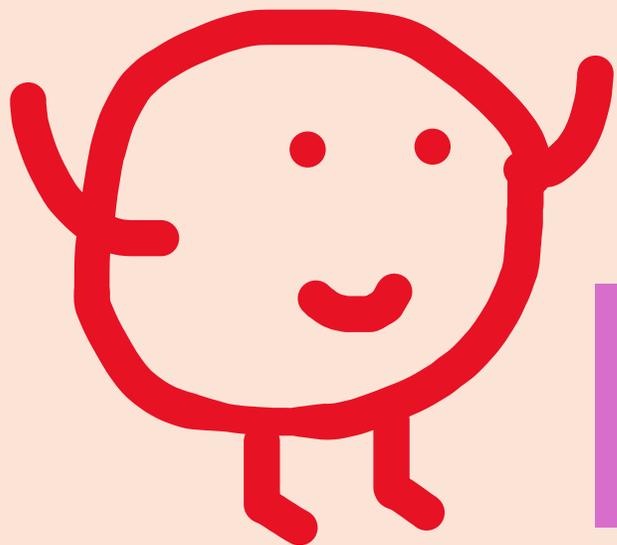
A bit bland



Delivery

Slightly off





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 fury in the night,

```
# 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

low-quality output is ignorable

errors are expected

New York Sessions

Year	Session	Introduced	Completed	Legislation
2023-2024	General Assembly	20091	3380	Browse

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





New York Sessions

errors are expected

Year	Session	Introduced	Completed	Legislation
2023-2024	General Assembly	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
	ai-Pro	4.8
		5.1
		5.6

CARDÁPIO EXCEPCIONAL DE DELIVERY

ALMOÇO

SÁBADO

MOQUECA DE OVO COM CAMARAO.....	34,00
MOQUECA SHIMEJI.....	32,00
MOQUECA DE FRUTOS DO MAR.....	62,00
MOQUECA DE PEIXE.....	52,00

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

ALMOÇO

DOMINGO

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

QUITUTES

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

o_acarajazz • Follow
Acarajazz - Sabores da Bahia

o_acarajazz Dia meu povo, peça e reserve pelo link da bio. 😊
19w See translation

loretacolucci Olá! Indo no restaurante é outro preço? Obrigada
3d 1 like Reply See translation
View replies (1)

fragmentado.ttt 🙄🙄🙄🙄
19w 1 like Reply
View replies (1)

maccs.carol Tem acarajé vegetariano ou vegano?
1w 1 like Reply See translation
View replies (1)

guxtavotamura É o mesmo cardapio pra almoçar ai?
19w Reply See translation
View replies (1)

jafamg A chuva na hora do almoço não me impediu de devorar a moqueca de frutos do mar do Acarajazz.....q delícia.....uma manjar para deuses!!! Vida longa para o Acarajazz 🍴🍴🍴🍴🍴🍴

Liked by **argentinosmurf** and others
February 24

Add a comment... **Post**

Claude

Professional Plan

☀️ Good morning, j

How can Claude help you today?

Claude 3.5 Sonnet ▾

Add content Use a project ▾

Generate excel formulas Polish your prose Generate interview questions

JS

Your recent chats ^ View all →

examples



Screenshot
2024-0...5 AM.png

Screenshot 2024-07-11 at 10.0

1.31 TB available

JS "Acarajé, abará, moquecas, compotas, licores... Essas e outras delícias do nordeste." what's the first few words mean

Let me translate and explain those first few words for you:

"Acarajé, abará, moquecas, compotas, licores..."

These are all types of food and drink, primarily associated with northeastern Brazilian cuisine:

1. Acarajé - A deep-fried ball made from mashed black-eyed peas, typically filled with shrimp and other ingredients.
2. Abará - Similar to acarajé, but steamed rather than fried, often wrapped in banana leaves.
3. Moquecas - Traditional Brazilian fish stews.
4. Compotas - Fruit preserves or jams.
5. Licores - Liqueurs, often made with local fruits.

The phrase as a whole is listing some of the specialties and delicacies of northeastern Brazilian cuisine that the restaurant likely serves. The ellipsis (...) at the end suggests there are more items beyond these examples.

The full sentence "Essas e outras delícias do nordeste" means "These and other delights from the Northeast," indicating that these items represent the northeastern Brazilian culinary traditions offered by the restaurant.

JS is northeast food bahian?

Reply to Claude...

Claude 3.5 Sonnet

Copy Retry

JS is northeast food bahian?

Northeast Brazilian cuisine and Bahian cuisine are related but not exactly the same:

1. Northeast Brazil (Nordeste) includes nine states: Bahia, Pernambuco, Ceará, Maranhão, Paraíba, Rio Grande do Norte, Alagoas, Sergipe, and Piauí.
2. Bahia is one state within the Northeast region, known for its particularly distinctive and influential cuisine.
3. Bahian cuisine is a subset of Northeastern cuisine, but it's often considered the most famous and representative of the region due to its unique flavors and African influences.
4. Many dishes mentioned (like acarajé, abará, and moqueca) are strongly associated with Bahian cuisine specifically.
5. However, other Northeastern states also have their own culinary specialties that may differ from Bahian cuisine.

So while Bahian food is Northeastern, not all Northeastern food is specifically Bahian. That said, in popular culture and in many restaurants outside the region, "Northeastern cuisine" often heavily features or focuses on Bahian dishes due to their popularity and distinct flavors.

JS ok this place is bahian, what should i expect and what should i get? i don't speak portugusse so i want to be prepared

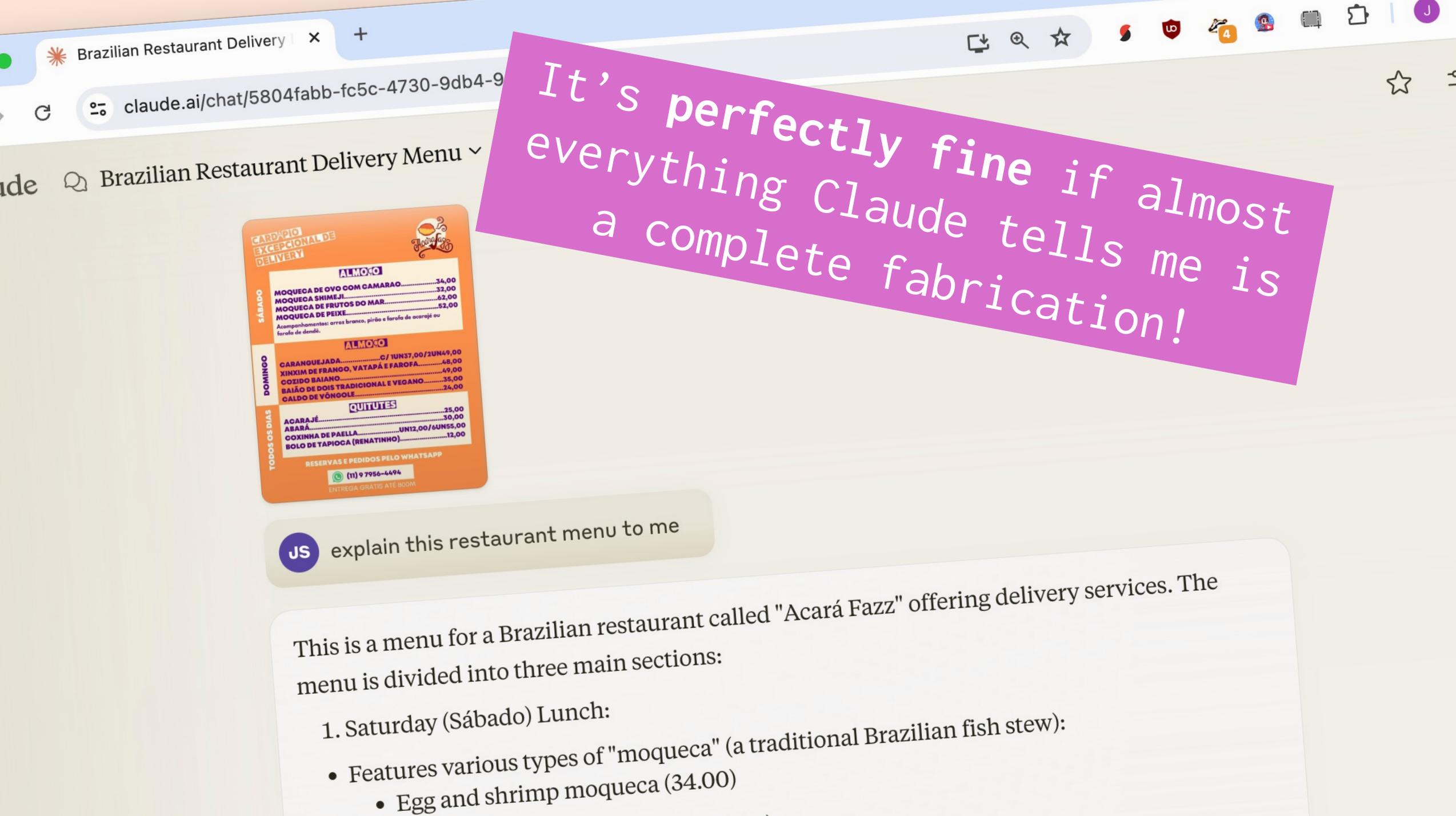
Reply to Claude...

Claude 3.5 Sonnet



JS





It's perfectly fine if almost everything Claude tells me is a complete fabrication!

CARDÁPIO EXCEPCIONAL DE DELIVERY	
SÁBADO	ALMOÇO
	MOQUECA DE OVO COM CAMARÃO..... 34,00
	MOQUECA SHIMEJI..... 32,00
	MOQUECA DE FRUTOS DO MAR..... 62,00
DOMINGO	ALMOÇO
	CARANQUEJADA..... C/ 1UN37,00/2UN49,00
	XINXIM DE FRANGO, VATAPÁ E FAROFA..... 48,00
	COZIDO BAIANO..... 49,00
TODOS OS DIAS	QUITUTES
	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)

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

20230AB886_93.pdf
Page 1 of 18

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

LEGISLATIVE COUNSEL'S DIGEST

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 Preservation 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 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 \$_____ annually to compensate digital journalism providers, as defined, for accessing the

93

20230AB886_93.pdf
Page 1 of 18

1 2 3 4 8 12 16 17 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.

AMENDED IN SENATE JUNE 27, 2024
AMENDED IN SENATE JUNE 10, 2024
AMENDED IN SENATE JULY 3, 2023
AMENDED IN ASSEMBLY MAY 30, 2023
AMENDED IN ASSEMBLY APRIL 27, 2023
AMENDED IN ASSEMBLY MARCH 23, 2023
CALIFORNIA LEGISLATURE—2023–24 REGULAR SESSION

ASSEMBLY BILL

No. 886

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

February 14, 2023

An act to add Title 23 (commencing with Section 3273.80) to Part 4 of Division 3 of the Civil Code, relating to civil law.

LEGISLATIVE COUNSEL'S DIGEST

AB 886, as amended, Wicks. California Existing law generally regulates on requiring, on a semiannual basis and company to submit to the Attorney General that includes, among other things, the service of the social media platform. This bill, the California Journalism Preservation Act, requires a covered platform, as defined, to either pay or compensate digital journalism providers, as d



New York Sessions

Year	Session	Introduced	Completed	Legislation	Search	Dataset
2023-2024	General Assembly	20091	3380	Browse	Search	Download

Texas Sessions

Year	Session	Introduced	Completed	Legislation	Search	Dataset
2023	88th Legislature 4th Special Session	399	220	Browse	Search	Download
2023	88th Legislature 3rd Special Session	598	292	Browse	Search	Download
2023	88th Legislature 2nd Special Session	169	71	Browse	Search	Download
2023	88th Legislature 1st Special Session	117	46	Browse	Search	Download
2023-2024	88th Legislature	11807	4464	Browse	Search	Download

California Sessions

Year	Session	Introduced	Completed	Legislation	Search	Dataset
2023-2024	Regular Session	5287	1198	Browse	Search	Download

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

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.

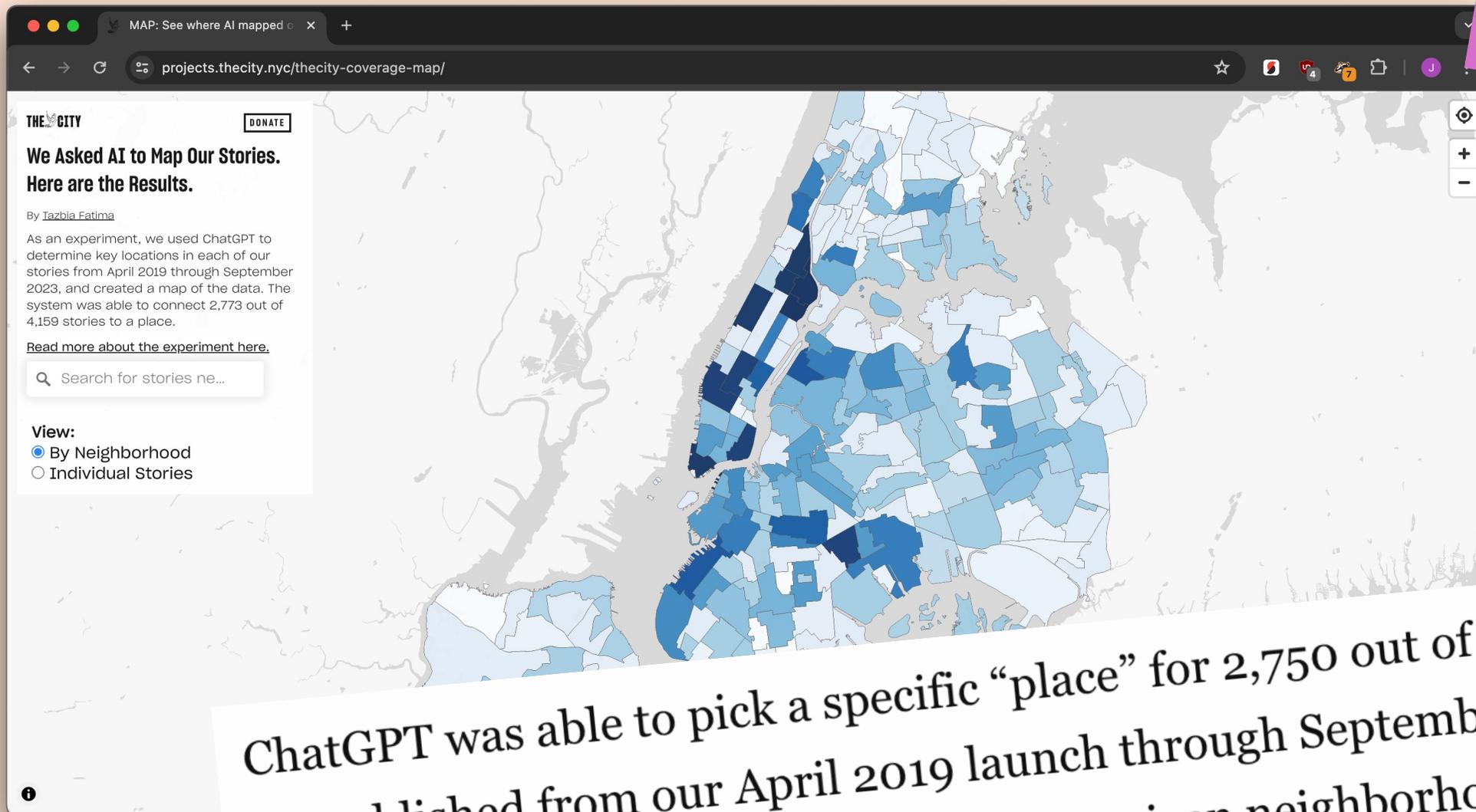
The prevalence of unwanted sexual content raises questions about whether Apple can continue to offer a protective cocoon to its customers as its platform grows

By [Reed Albergotti](#) and [Al Johri](#)
November 22, 2019 at 7:00 a.m. EST

all this ap
hate and
Nothing
I had one

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



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

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

Your recent chats

bills

20230AB886_93.pdf

20230AB886_93 copy 2.pdf

20230AB886_93 copy 3.pdf

20230AB886_93 copy 4.pdf

20230AB886_93 copy 5.pdf

20230AB886_93 copy 6.pdf

20230AB886_93 copy 7.pdf

20230AB886_93 copy 8.pdf

20230AB886_93 copy 9.pdf

20230AB886_93 copy 10.pdf

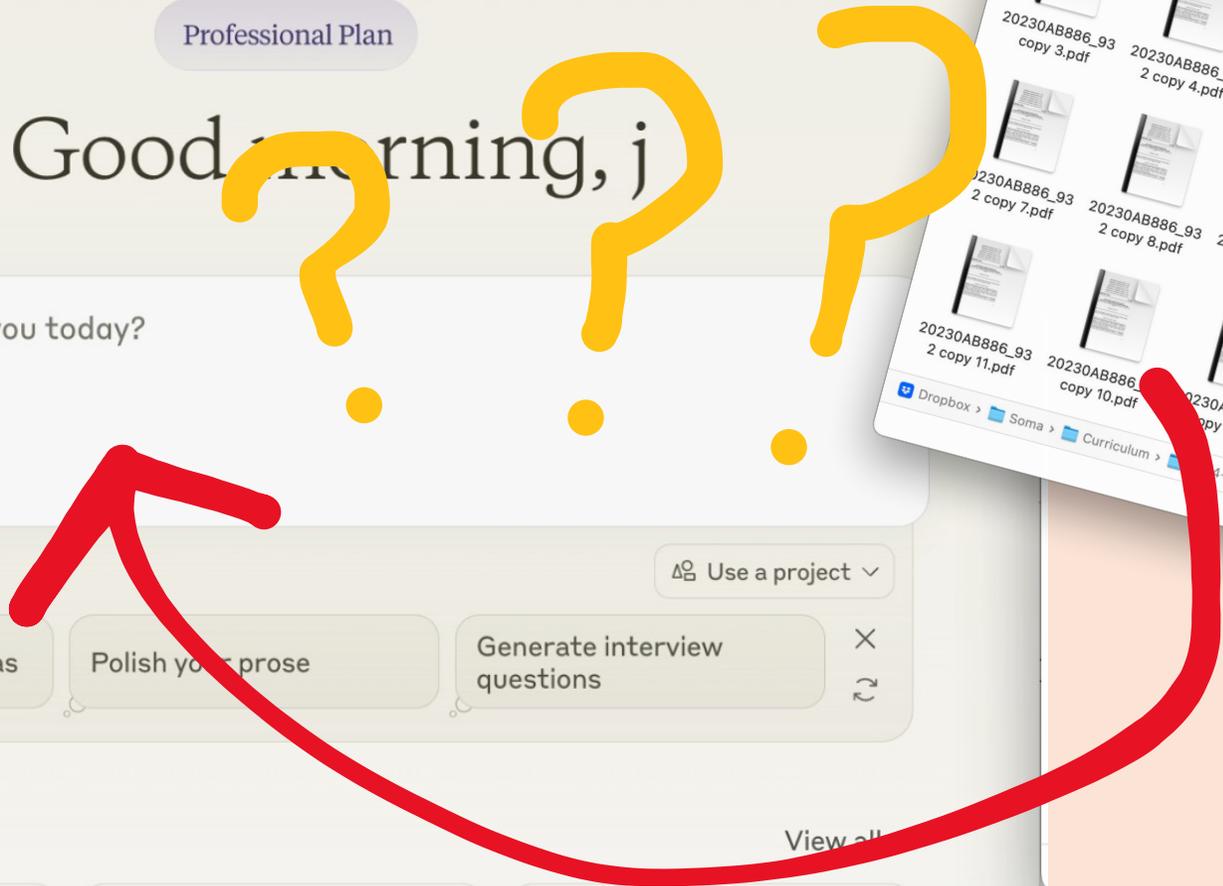
20230AB886_93 copy 11.pdf

20230AB886_93 copy 12.pdf

20230AB886_93 copy 13.pdf

Dropbox > Soma > Curriculum > 4-abraji > documents > examples > bills

2 of 30 selected, 1.31 TB available





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: 



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

Works with:  

★★★★☆ 318 ⓘ ↓ 6M+



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

Browser tabs: Claude for Sheets™ - Google

Address bar: workspace.google.com/marketplace/app/claude_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 ⓘ ↓ 34K+

Overview Permissions Reviews

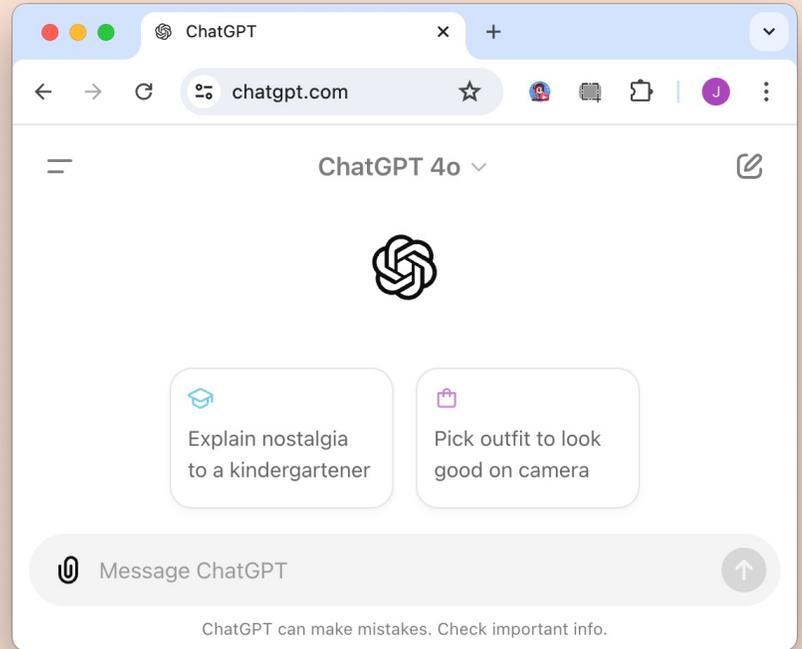
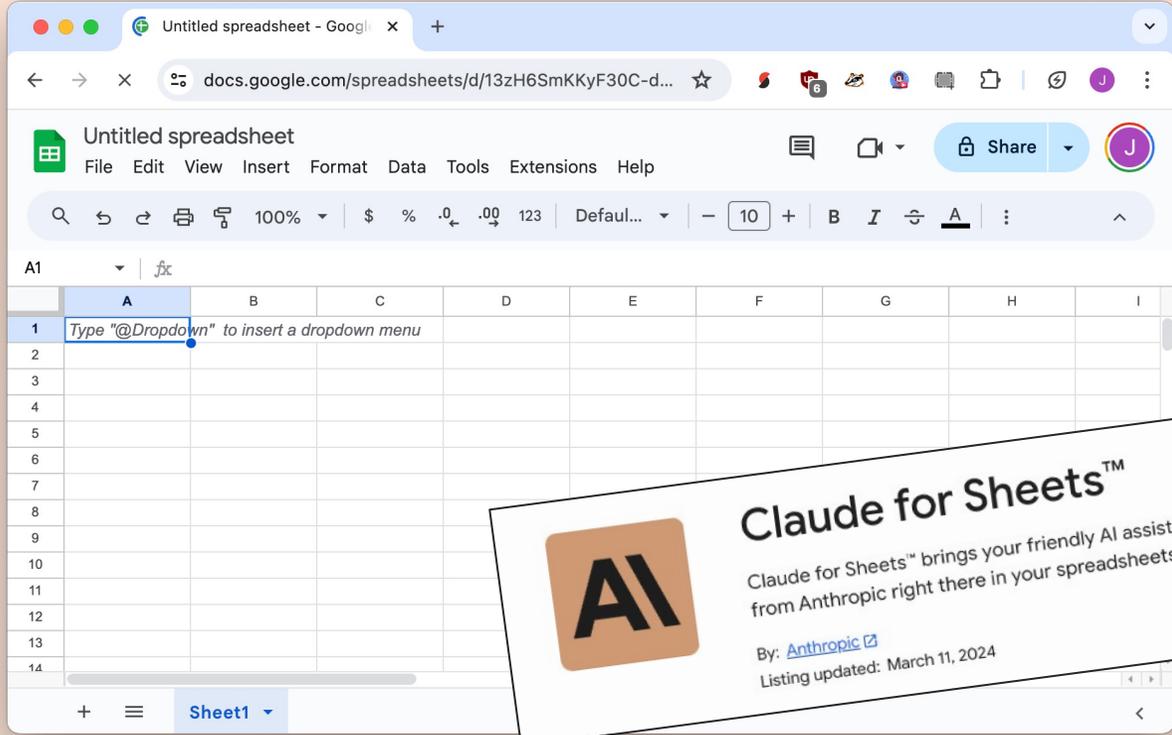
This sheet lets you compare a set of related prompts to Claude

Step 1: type your prompts in this column	Step 2: Select your model:	Step 3: review Claude responses in this column
In one sentence, what is good about the color blue?	claude-v1.3	Blue is a calming and soothing color.
Suppose you have a table called "Customers" with the following column names: <i>CustomerID, CustomerName, ContactName, Address, City, PostalCode, Country</i> Can you write a SQL query that finds all customers in San Francisco with a customer ID of greater than 1000?	claude-2	Here is a SQL query to find all customers in San Francisco with a customer ID 1000 from the Customers table: <pre>''sql SELECT * FROM Customers WHERE City = 'San Francisco' AND CustomerID > 1000 ...</pre> 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



AI **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

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  
nPx-cpCdeWV2sRKL_pfrkj1mQctqoDcEo1fJX  
cld7exevD-Q-y5y_KAAA
```

Close

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

ANTHROPIC

Dashboard Workbench Settings

Create Key

API Keys

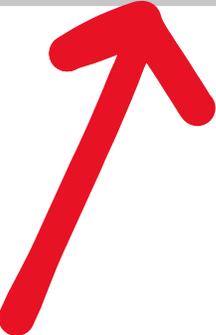
NAME	KEY	CREATED AT	UPDATED AT
claude for sheets	sk-ant-api03-1Sk...iAAA	Apr 9, 2024	...
sample class	sk-ant-api03-g0n...rgAA	Apr 10, 2024	...

A1	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			

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

G3 | fx

	A	E	F	G
1	bill text	about retirement?	about employment?	

> HLS 24RS-94 **[ENGROSSED]**
>
> 2024 Regular Session

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

> HIGH SCHOOL: Prohibits the use of an appeals
> certain students
>
> HLS 24RS-88 **[REENGROSSED]**
>
> 2024 Regular Session
>
> HOUSE BILL NO. 9
>
> BY REPRESENTATIVE ILLG
>
> RETIREMENT/DISTRICT ATTY: Provides for survivor benefit
> the

no no

yes no

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.	mulbs@example.com	Mulberry Peppertown
I am irate about the broccoli incident, I am never shopping at your store again. Take me off of your broccoli mailing list immediately.		
Jackary Baloneynose jackary.baloney@example.com	jackary.baloney@example.com	Jackary Baloneynose
Your flour is ground too finely. I do not go here to buy 00 flour 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

=CLAUDEEXTRACT(...)

“email”

“name”

ÉLECTIONS EUROPÉENNES * EMMANUEL MACRON

Elections 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

- 1 En direct, la situation politique en France : Elisabeth Borne siègera au sein du groupe Renaissance, son président n'est toujours pas élu
- 2 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 Hayer. « 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
<p>Ante inversores en Washington, Caputo defendió la estrategia fiscal y reafirmó el compromiso para levantar el cepo</p> <p>El jefe del Palacio de Hacienda participó de un seminario organizado por el banco J.P. Morgan</p> <p>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, luego de haber estado desayunando en un restaurante de Washington.</p> <p>Elections en France: Emmanuel Macron défend l'hypocrisie</p> <p>Le chef de l'Etat a déclaré que les « peurs dans la campagne présidentielle profitent aux plus simplistes », ceux qui dénoncent l'hypocrisie en dérangeant les « dividendes silencieux ».</p> <p>Vous pouvez partager un article en cliquant sur les icônes de réseaux sociaux à droite de celui-ci.</p> <p>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</p> <p>岸田文雄首相は17日、韓国の尹錫悦（ユンソンニョル）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。</p> <p>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っていく」と述べ、尹大統領と電話で協議した。</p>	<p>Luis Caputo, Ministro de Economía Pablo Quirno, Secretario de Finanzas Javier Milei, gobierno Mauricio Macri, gobierno Nicolás Posse, Jefe de Gabinete Alberto Fernández, gobierno Pierre-Olivier Gourinchas, Economista Jefe del FMI</p> <p>Raphaël Glucksmann, Head of list for Socialist Party (Parti Socialiste) and Place Publique Gabriel Attal, Prime Minister of France</p> <p>Fumio Kishida, Prime Minister; Yoon Suk-yeol, President of South Korea</p>	<p>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</p> <p>deplora que les « peurs profitent aux réponses les plus simplistes », mercredi 17 avril.</p> <p>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。</p>	<p>en</p> <p>fr</p> <p>ja</p>	<p>yes</p> <p>no</p> <p>no</p>	<p>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.</p> <p>Emmanuel Macron lamented that "fears profit from the simplest answers", Wednesday April 17.</p> <p>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.</p>	<p>international</p> <p>politics</p> <p>other</p>

<https://bit.ly/claude-sheets>



but yes, you must

spot check results

*Chatbots May 'Hallucinate'
More Often Than Many Realize*

sometimes you just want

a single answer

from a lot of documents



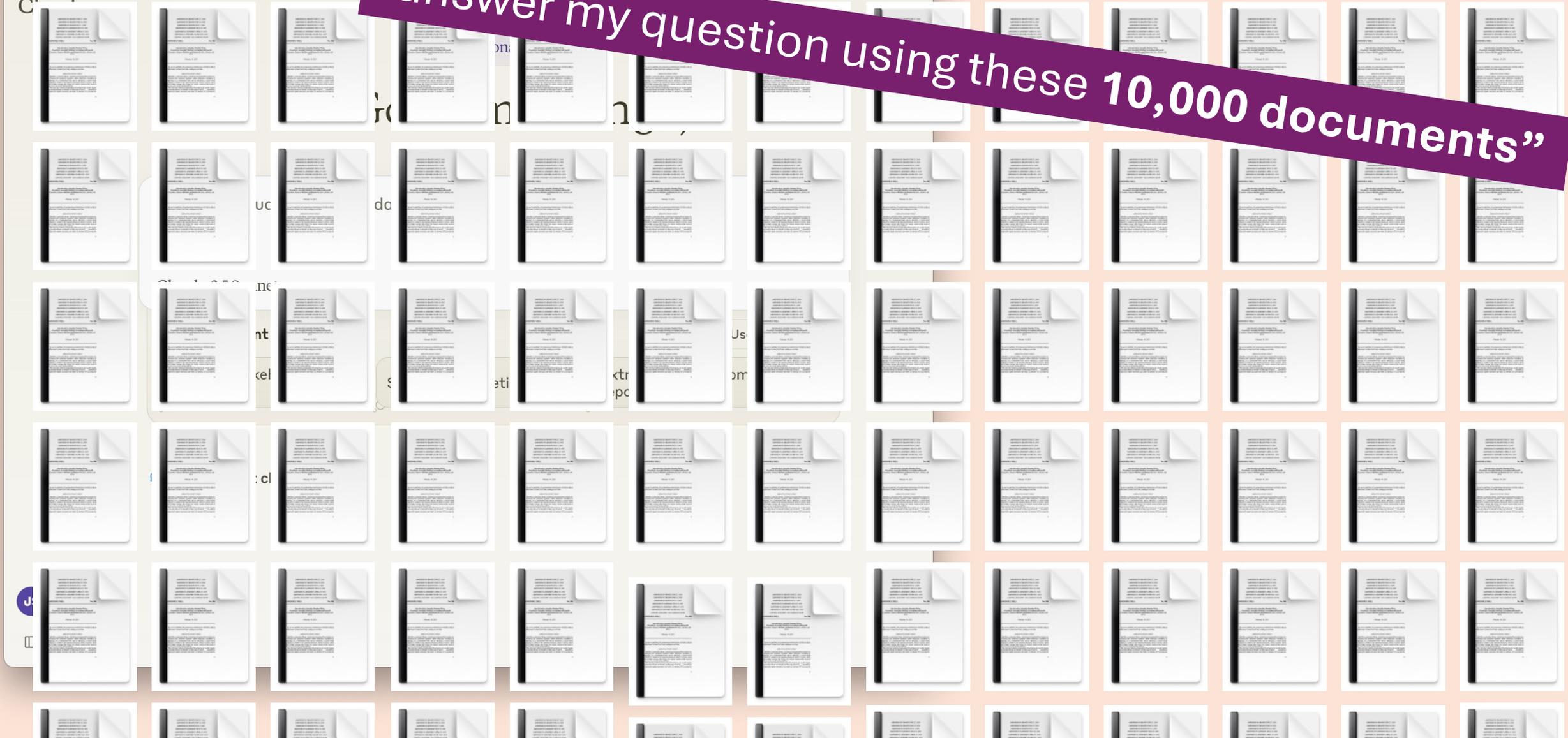
JS explain this restaurant menu to me

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)
 - Shimeji mushroom moqueca (32.00)
 - Seafood moqueca (62.00)
 - Fish moqueca (52.00)
 - These come with white rice, pirão (a fish gravy), and farofa made with either acarajé



“answer my question using these 10,000 documents”



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 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 Bus the Law

The Microsoft-powered bot says bosses can tak
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?

Do landlords have to accept tenants on rental assistance?

Can I keep my funeral home pricing confidential?

Can I make my store cashless?

🤖 NYC Chatbot Answer (Emphasis by The Markup)

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

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

“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.”

“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.”

🏠 Reality

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

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

The Federal Trade Commission has outlawed concealing funeral prices.

Stores have been required to accept cash as payment since 2020.

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.



You

where did he get his masters?



ChatGPT

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



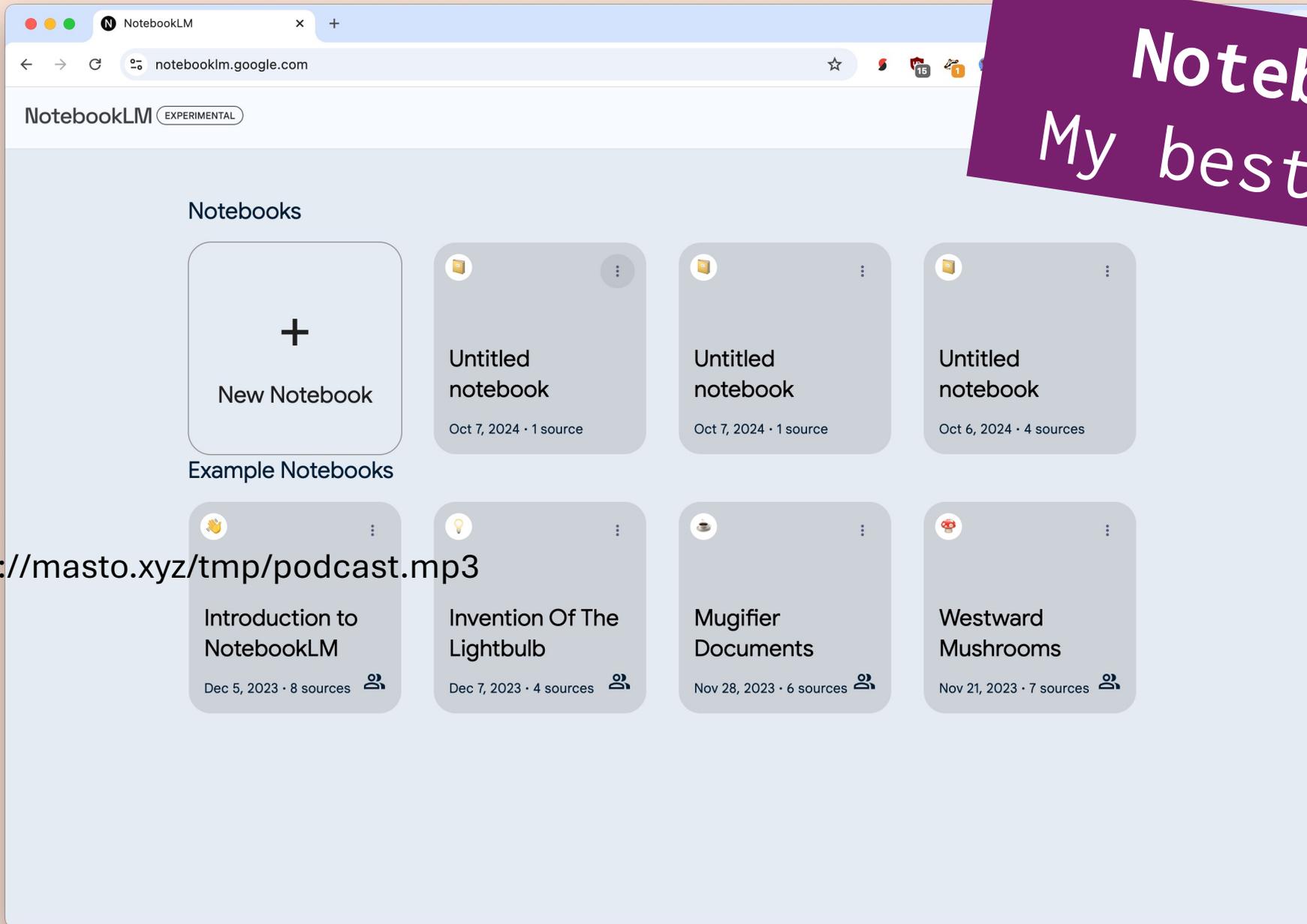
You

Where did he get his master's degree?



ChatGPT

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



NotebookLM,
My best friend!

<https://masto.xyz/tmp/podcast.mp3>

NYC Housing Rights: Source of Income Discriminator

Sources

- FairHouse_FAQBroker-2021Eng.pdf
- Section 8 Guide for Tenants.pdf
- SourceOfIncomeFactSheet.pdf
- Wages and pay _ New York State Attor...

Chat

Can I discriminate against poor people?

Based on the sources, you cannot discriminate against poor people in housing in New York City because the NYC Human Rights Law prohibits discrimination based on "lawful source of income" 1 ...

Here's a breakdown of what the sources say:

- "Lawful source of income" is a protected category under the NYC Human Rights Law 1 ... This includes various forms of public or housing assistance, Supplemental Security Income (SSI), HIV/AIDS Services Administration (HASA), CIT...
- This protection extends to deposits, move-in fees, or...
- Brokers, real estate agents, and landlords are prohibited from discriminating because they receive sub...
- You should not work with anyone who discriminates against the law 4 .
- Maintaining separate wait...

or no credit, requiring a specific credit score may trigger an investigation by the Commission. Each application should be considered on a case-by-case basis. Prior to rejecting an applicant based on credit, the applicant

The NYC Human Rights Law

Citations!

4 sources

Who is liable for discriminating under NYC law? What penalties exist for violating NYC law?

- 2012-bicycle-crash-data-report.pdf
- 5261.pdf
- Brochure79_7-24.pdf
- dining-dogs-rules.pdf
- FairHouse_FAQBroker-2021Eng.pdf
- mv529a.pdf
- noa-dining-out-nyc.pdf
- notice-public-hearing-dining-out-nyc.pdf
- nyc-screenline-report-2012.pdf
- nycchecklist_restaurantsigns.pdf
- Outdoor_Dining_Heating_Systems_ENGLISH.pdf
- Outdoor-Dining-FAQ.pdf
- rat-guidance-outdoor-dining.pdf
- Section 8 Guide for Tenants.pdf
- SourceOfIncomeFactSheet.pdf
- TDA-Loans-Code42_10-6-24.pdf
- Wages and pay _ New Yo...tate Attorney General.pdf
- young_mens_initiative_annual_report_2012.pdf

< > Searching "pdfs" >>

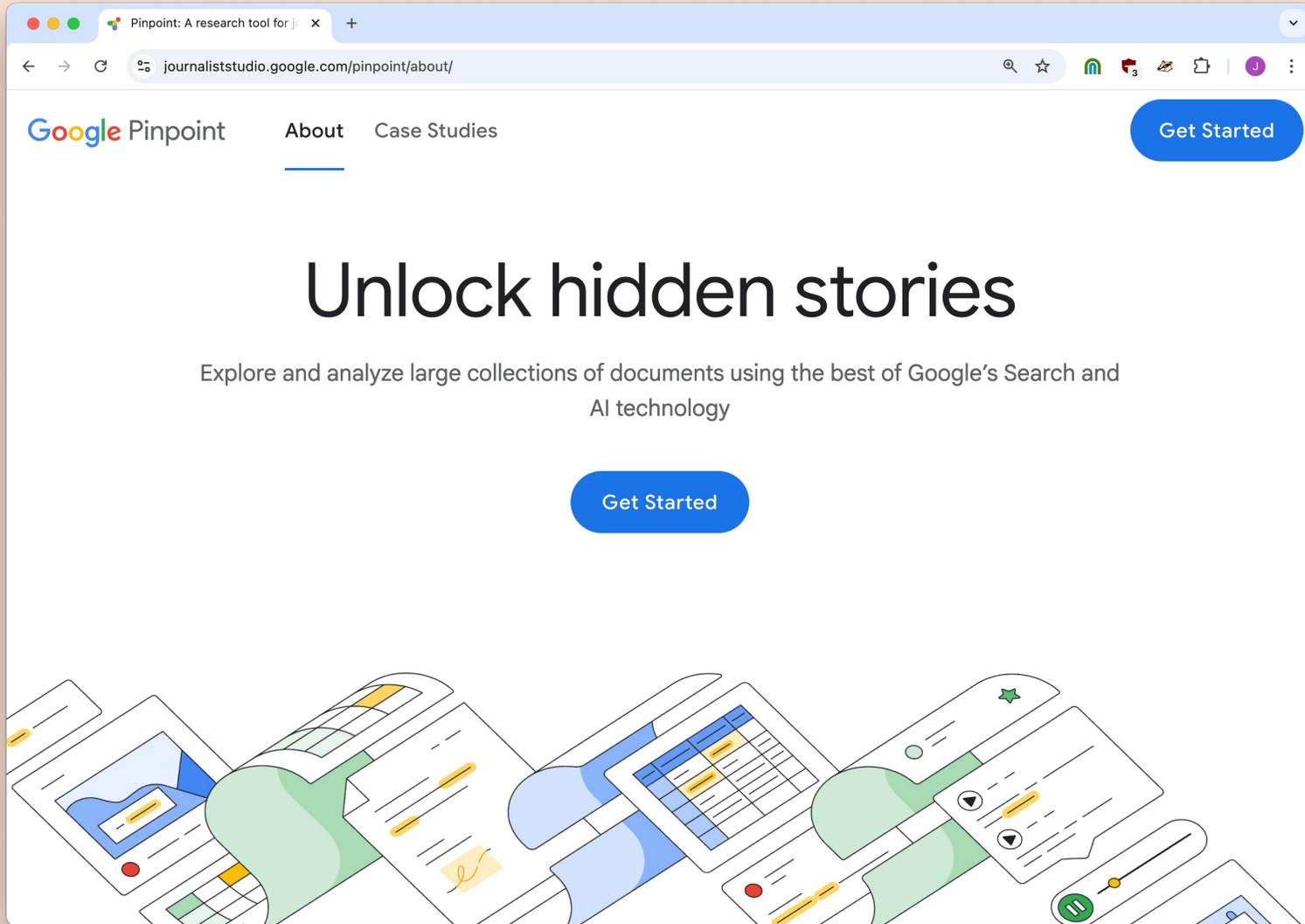
Search: This Mac "pdfs" Save +

Name	Date Modified	Size
 notice-public-...g-out-nyc.pdf	Mar 19, 2025 at 10:31 AM	

< > Searching "pdfs" >>

Search: This Mac "pdfs" Save +

Name	Date Modified	Size



Pinpoint - cji x +

journaliststudio.google.com/pinpoint/search?collection=1888004b82213322

Pinpoint Search documents ? Give us feedback

+ Add documents

Documents 1 - 18 of 18 Show info

- PDF 2012-bicycle-crash-data-report.pdf
- PDF 5261.pdf
- PDF Brochure79_7-24.pdf
- PDF dining-dogs-rules.pdf
- PDF FairHouse_FAQBroker-2
- PDF mv529a.pdf
- PDF noa-dining-out-nyc.pdf
- PDF notice-public-hearing-dini
- PDF nyc-screenline-report-201

Extract Structured Data BETA

Help

Give us feedback

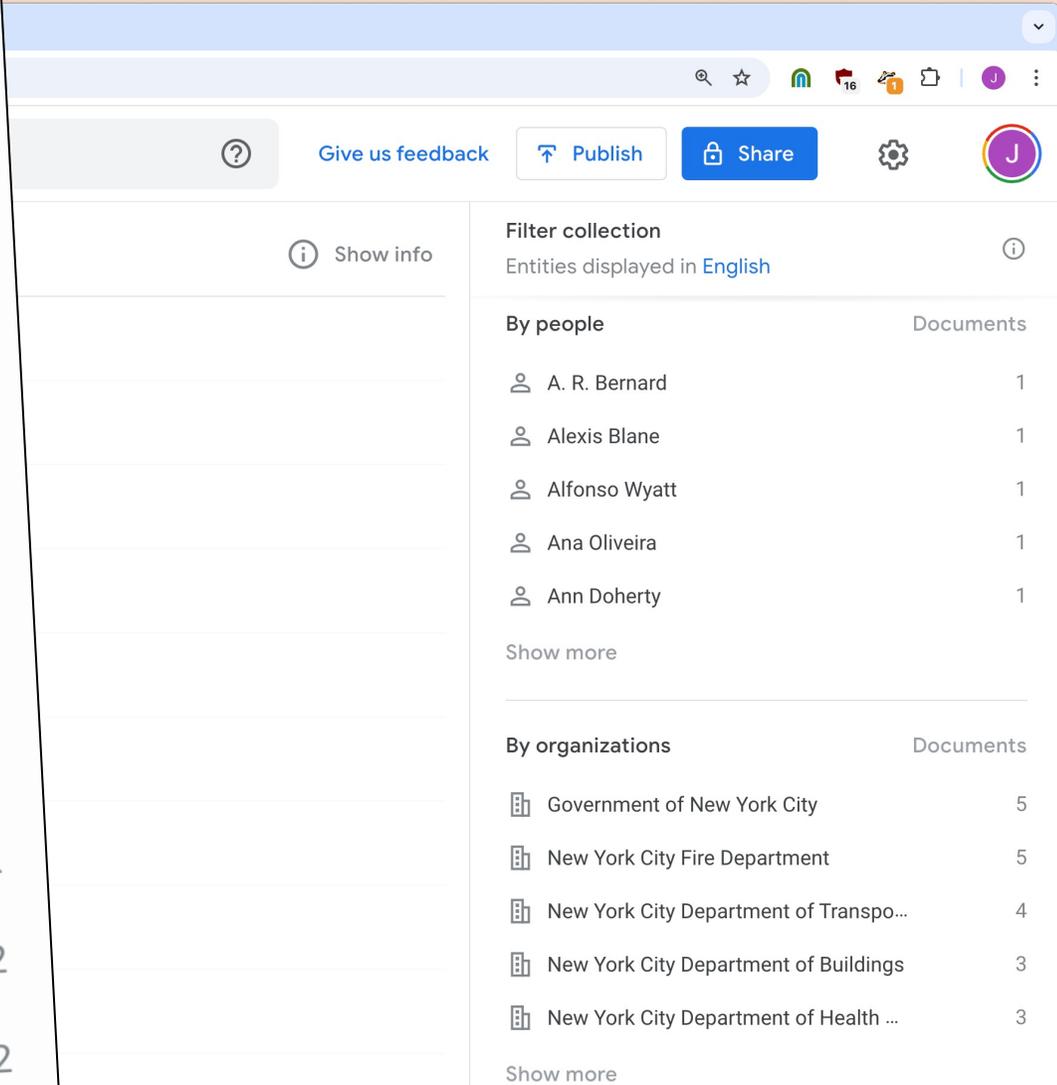
Search bike racks ? Give us feedback

Documents 1 - 2 of 2

- PDF noa-dining-out-nyc.pdf
Citi Bike/Bike Share Stations, Bike Corrals, 5 ...
the presence and composition of bicycle lanes and ...
shall not be used for storage, including containers
- PDF notice-public-hearing-dining-out-nyc.pdf
5 Feet Citibike/Bikeshare stations, Bike Corrals, Micromobility ...
and composition of bicycle lanes and ...
not be used for storage, including containers to

By organizations

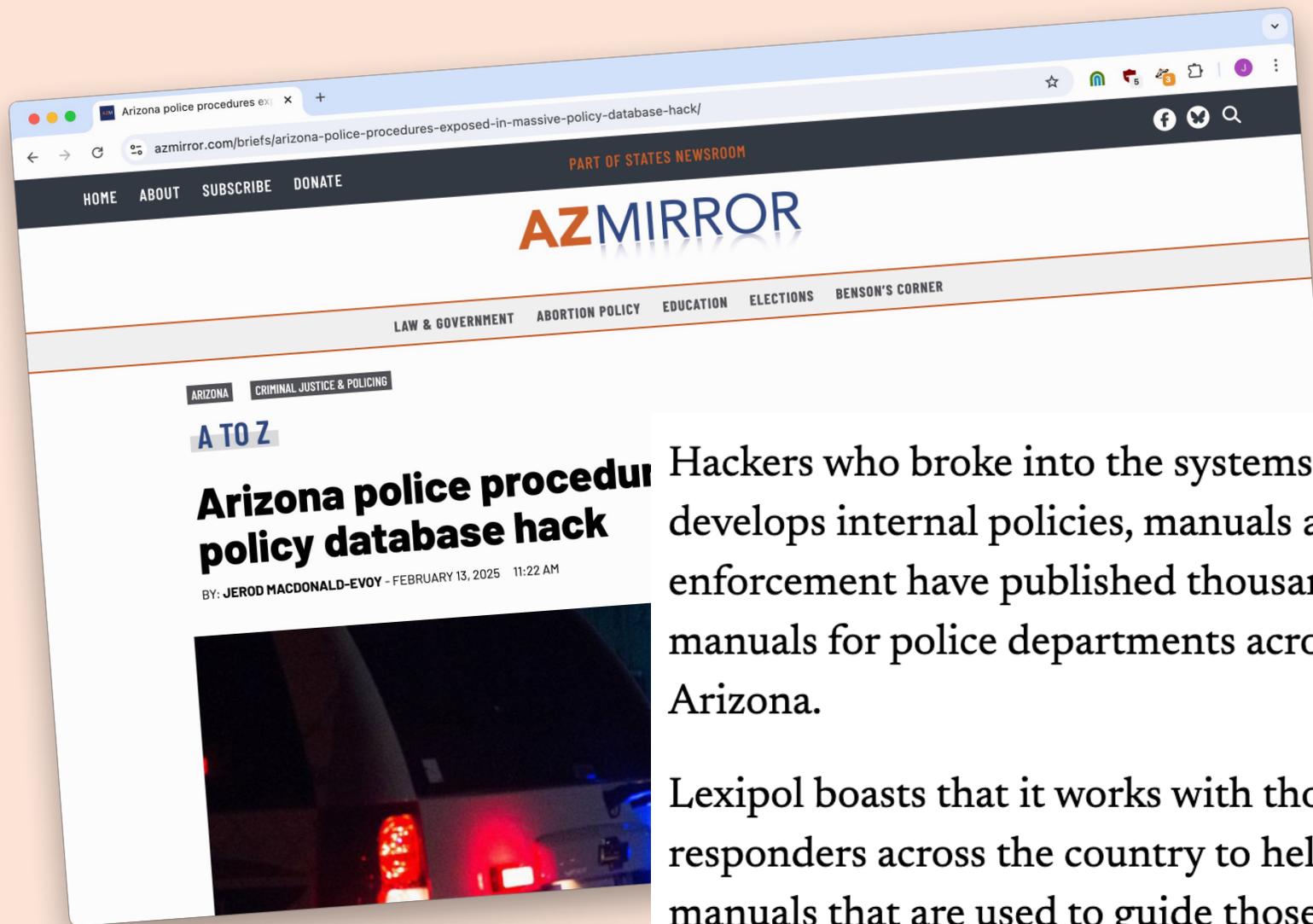
	Documents
 Government of New York City	5
 New York City Fire Department	5
 New York City Department of Transpo...	4
 New York City Department of Buildings	3
 New York City Department of Health ...	3
 Citi Bike	2
 MTA	2
 New York City Department of Finance	2
 New York City Department of Parks &...	2
 New York City Housing Authority	2
 New York City Landmarks Preservatio...	2
 New York City Sanitation Department	2



The screenshot shows a web interface with a navigation bar at the top containing a search icon, a star, and several utility icons. Below the navigation bar are buttons for "Give us feedback", "Publish", and "Share". The main content area is divided into two columns. The left column has a "Show info" button. The right column contains a "Filter collection" section with a sub-header "Entities displayed in English". Under this, there are two filter categories: "By people" and "By organizations".

By people	Documents
 A. R. Bernard	1
 Alexis Blane	1
 Alfonso Wyatt	1
 Ana Oliveira	1
 Ann Doherty	1
Show more	

By organizations	Documents
 Government of New York City	5
 New York City Fire Department	5
 New York City Department of Transpo...	4
 New York City Department of Buildings	3
 New York City Department of Health ...	3
Show more	



Hackers who broke into the systems of a Texas-based company that develops internal policies, manuals and other training material for law enforcement have published thousands of finished and draft procedure manuals for police departments across the country, including many in Arizona.

Lexipol boasts that it works with thousands of police departments and first responders across the country to help them create policy and procedure manuals that are used to guide those agencies on a myriad of topics.

The Arizona Mirror examined the over 8,500 files leaked by a hacking group calling itself “the puppygirl hacker polycule” and published by the nonprofit [DDoSecrets](#) that archives leaked and hacked material.

Pinpoint - Lexipol

journaliststudio.google.com/pinpoint/search?collection=cdf2b54c0e17728

Pinpoint Search documents ? Give us feedback Publish Share

+ Add documents

Lexipol

Documents 1 - 100 of 156 Show info

- RELEASE_20120203_T114307_Morrison County SO Custody Manual (pages 1 to 231).pdf
- RELEASE_20120203_T114307_Morrison County SO Custody Manual (pages 232 to 452).pdf
- RELEASE_20121226_T121300 (pages 1 to 234).pdf
- RELEASE_20121226_T121300 (pages 235 to 460).pdf
- RELEASE_20130605_T110000_Contra Costa County Probation Dept Policy Manual.pdf
- RELEASE_20140316_T081945 (pages 1 to 246).pdf
- RELEASE_20140316_T081945 (pages 247 to 493).pdf
- RELEASE_20140316_T082657 (pages 1 to 309).pdf

Search locations

- Massachusetts 9
- North Carolina 9
- Missouri 8
- Nevada 8
- North Dakota 8
- Wisconsin 8
- Arizona 7
- Idaho Peace Officers Standards and ... 6
- New York 6
- The U.S. Mission to the United Nations 6
- Utah 6
- C-ASSIST Family Health Clinic 5
- Department of Criminal Justice Infor... 5
- Department of Personnel & Training 5

Extract Structured Data BETA

Help

Give us feedback

Add documents

Washington Commanders X

Documents 1 - 44 of 44

RELEASE_20121226_T121300 (pages 235 to 460).pdf
data for the quarterly report should be gathered by the Responsible Physician via monthly meetings with all facility Shift Commanders and qualified

RELEASE_20140316_T082657 (pages 1 to 309).pdf
420 Criminal Street Gangs. - 421 Area Sergeants - 422 Area Commanders . - 423 - Mobile Audio Video Procedure . 424 - Use of Audio Recorders

RELEASE_20140316_T114053 (pages 1 to 247).pdf
Table of Contents - 5 Twin Rivers Unified School District Police Department Twin Rivers District PD CA Policy Manual 420 Watch Commanders

RELEASE_20140316_T114053 (pages 248 to 493).pdf
Policy 420 Twin Rivers Unified School District Police Department Twin Rivers District PD CA Policy Manual Watch Commanders 420.1 PURPOSE

RELEASE_20140316_T132956 (pages 1 to 304).pdf
Watch Commanders will ensure that at least one field supervisor is deployed during each

Filter collection
Entities displayed in English

By dates mentioned
Start date YYYY-MM-DD
End date YYYY-MM-DD
Filter

Table with 2 columns: By people, Documents. Rows include Jeanne Clery (30), James Byrd (11), Jane Doe (11), John Doe (10), Matthew Shepard (9).

By organizations Documents

Extract Structured Data BETA

Help

Give us feedback

and finally, time for

transcription

because you hate typing out
your interviews

nytimes.com/2024/03/01/technology/elon-musk-openai-sam-altman-la... ☆

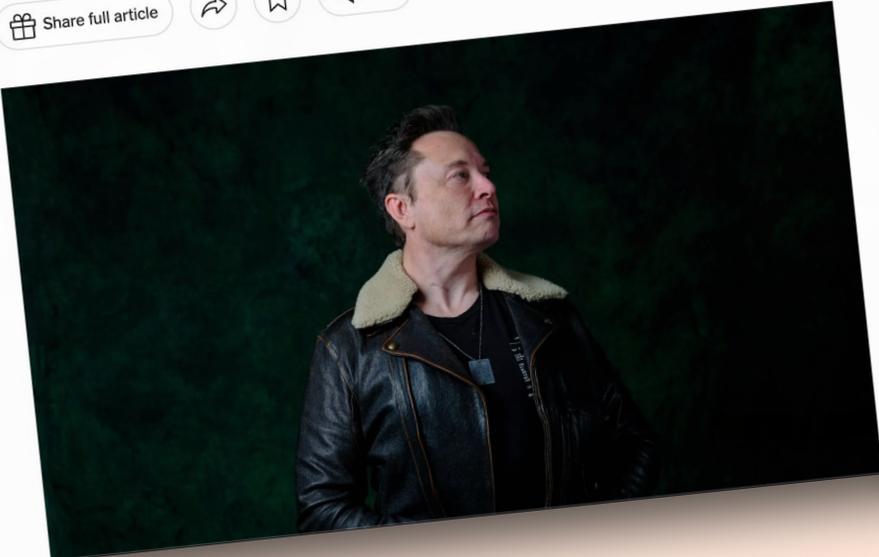
The New York Times

Artificial Intelligence > Harvesting Data for A.I. A.I. Data Race, Explained What Is 'Synthetic Data'? Key Figures in the Field A.I. Faces Quiz

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

chat.openai.com/c/e17b2a54-dd54-4c26-ae78-fdea741cc09f

ChatGPT 4

You: How many n's are in mayonnnnnaise?

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

ChatGPT...

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 ↗](#)

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

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

MacWhisper

goodsnooze.gumroad.com//mac...

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 optim one's out there trying to make a po their time or be ineffective. I think it we were previously optimizing for m the moment. Or maybe we've enter uptake. We kind of are still inheriting following and we haven't upgraded c really want is a little bit of th 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: 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](#).

Microphone Audio File YouTube

inputs

Record from microphone

Task

transcribe translate

Return timestamps

Transcription

Transcription Time (s)

Clear Submit

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

Built with Gradio

EasyWhisper | Easily transcri...

easywhisper.io

Features Prices FAQs

Use audio

for Mac OS

Enter YouTube URL

My files



MacWhisper



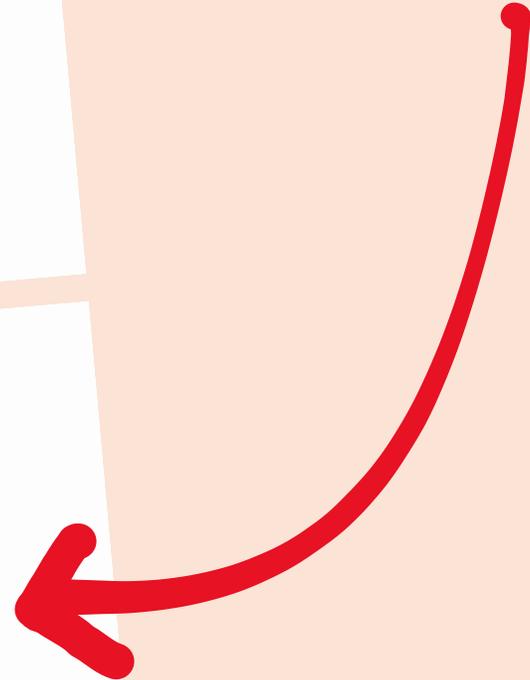
Jordi Bruin

★★★★★ 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)



openai/whisper: Robust Spee x +

github.com/openai/whisper

README MIT license

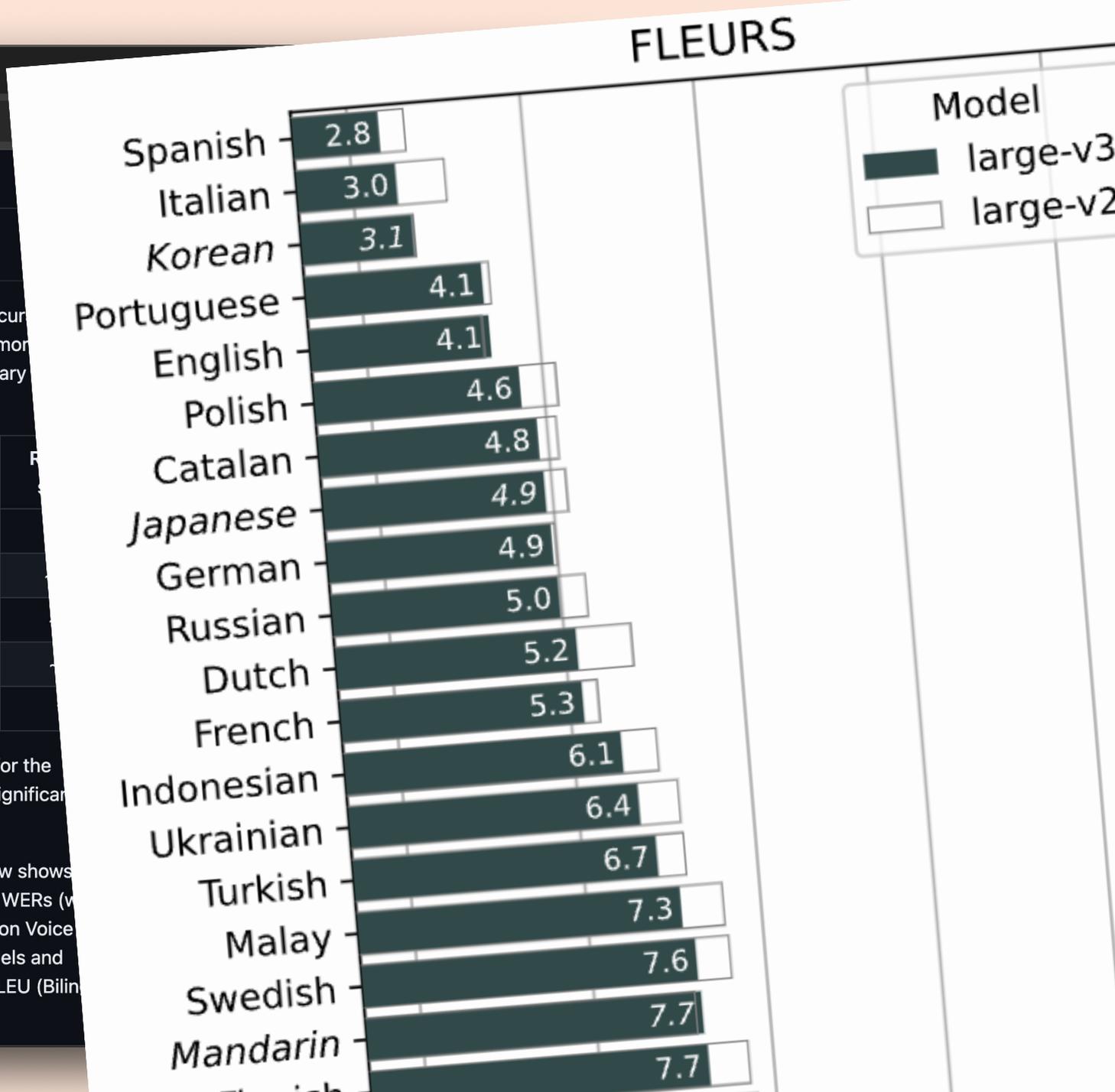
Available models and languages

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	Relative Speed
tiny	39 M	<code>tiny.en</code>	<code>tiny</code>	~1 GB	2.8
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	3.0
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	3.1
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	4.1
large	1550 M	N/A	<code>large</code>	~10 GB	4.1

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 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}

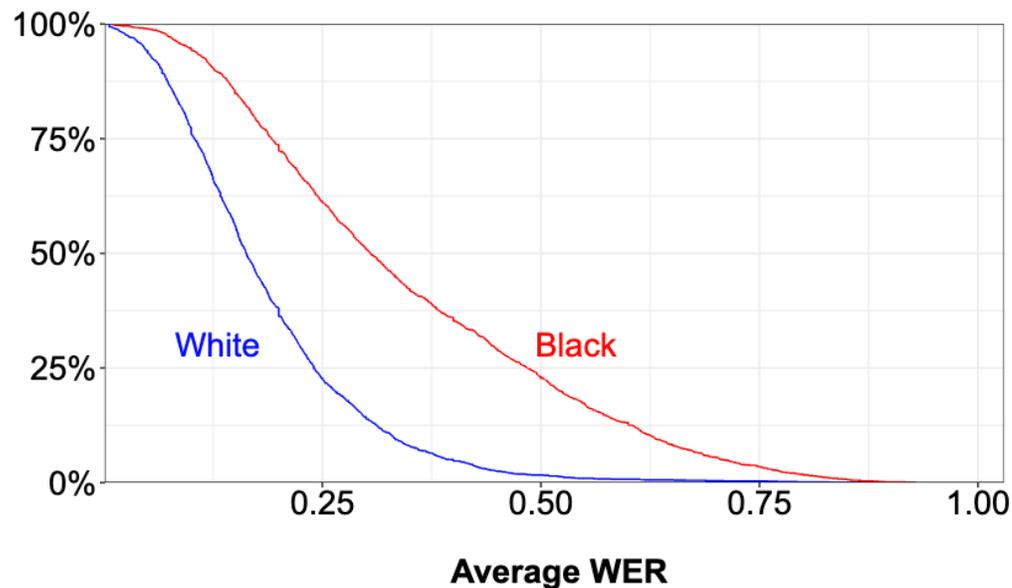


Fig. 2. The CCDF denotes the share of audio snippets having a WER greater than the value specified along the horizontal axis. The two CCDFs shown for audio snippets by white speakers (blue) versus those by black speakers (red) use the average WER across the five ASR services tested. If we assume that a 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.



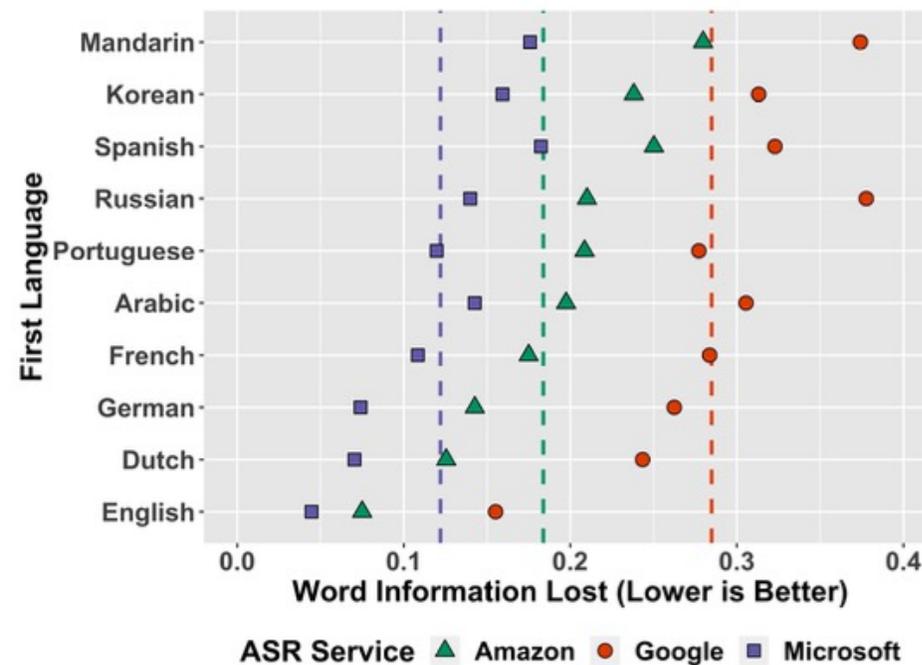
Performance Disparities Between Accents in Automatic Speech Recognition

Alex DiChristofano
a.dichristofano@wustl.edu
Division of Computational & Data Sciences
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

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

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



Jonathan Soma
js4571@columbia.edu
@dangerscarf

AI for Investigative Reporting

tricks for the newsroom