

Alspiration

Reimagining the Humanities and Art with Generative AI

Jonathan Soma

Knight Chair in Data Journalism

Columbia University Graduate School of Journalism

js4571@columbia.edu

investigate.ai • normalai.org • aifaq.wtf

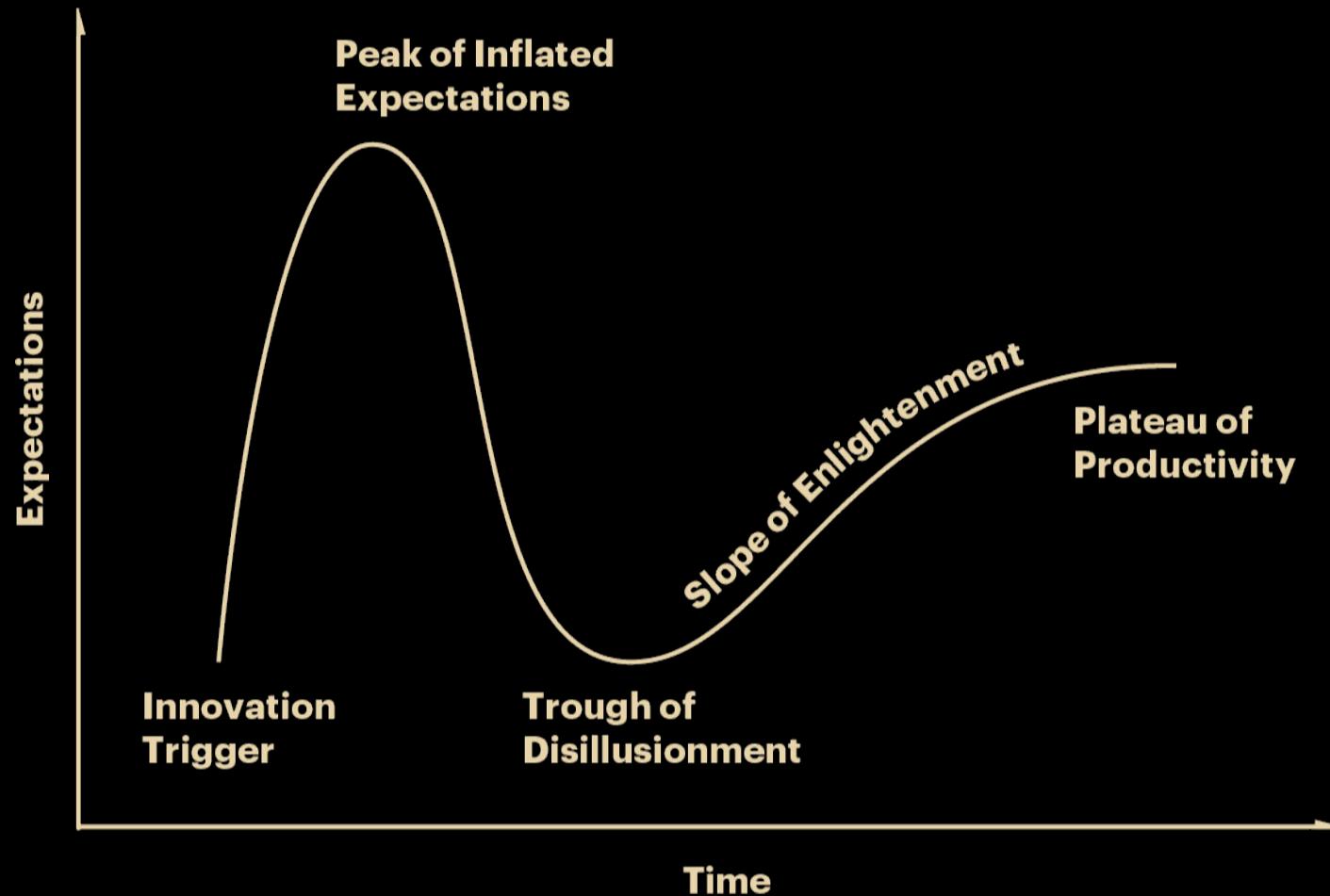
Your questions are always
more interesting than my talks!

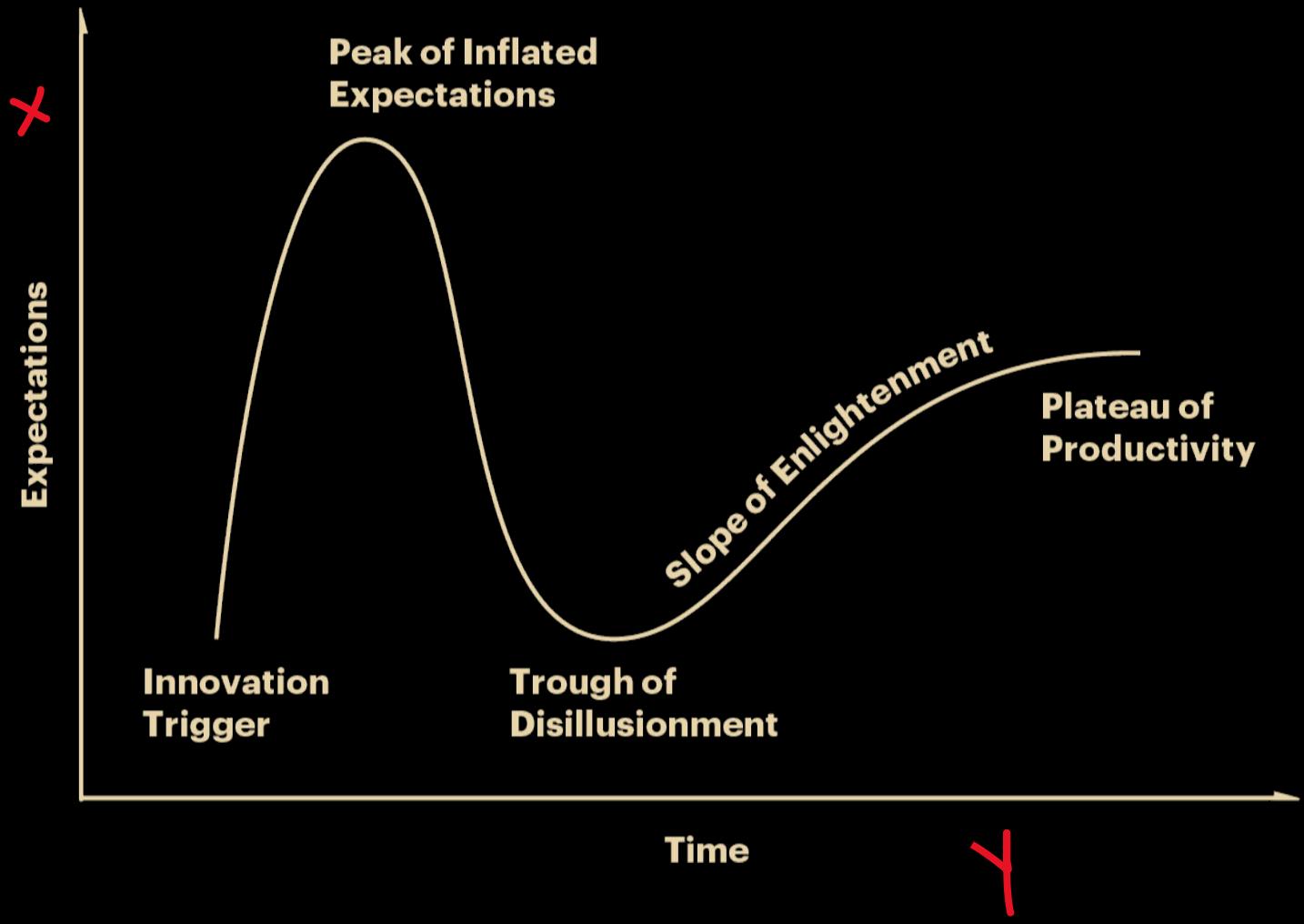
If I could do this talk in *five minutes* I definitely would.

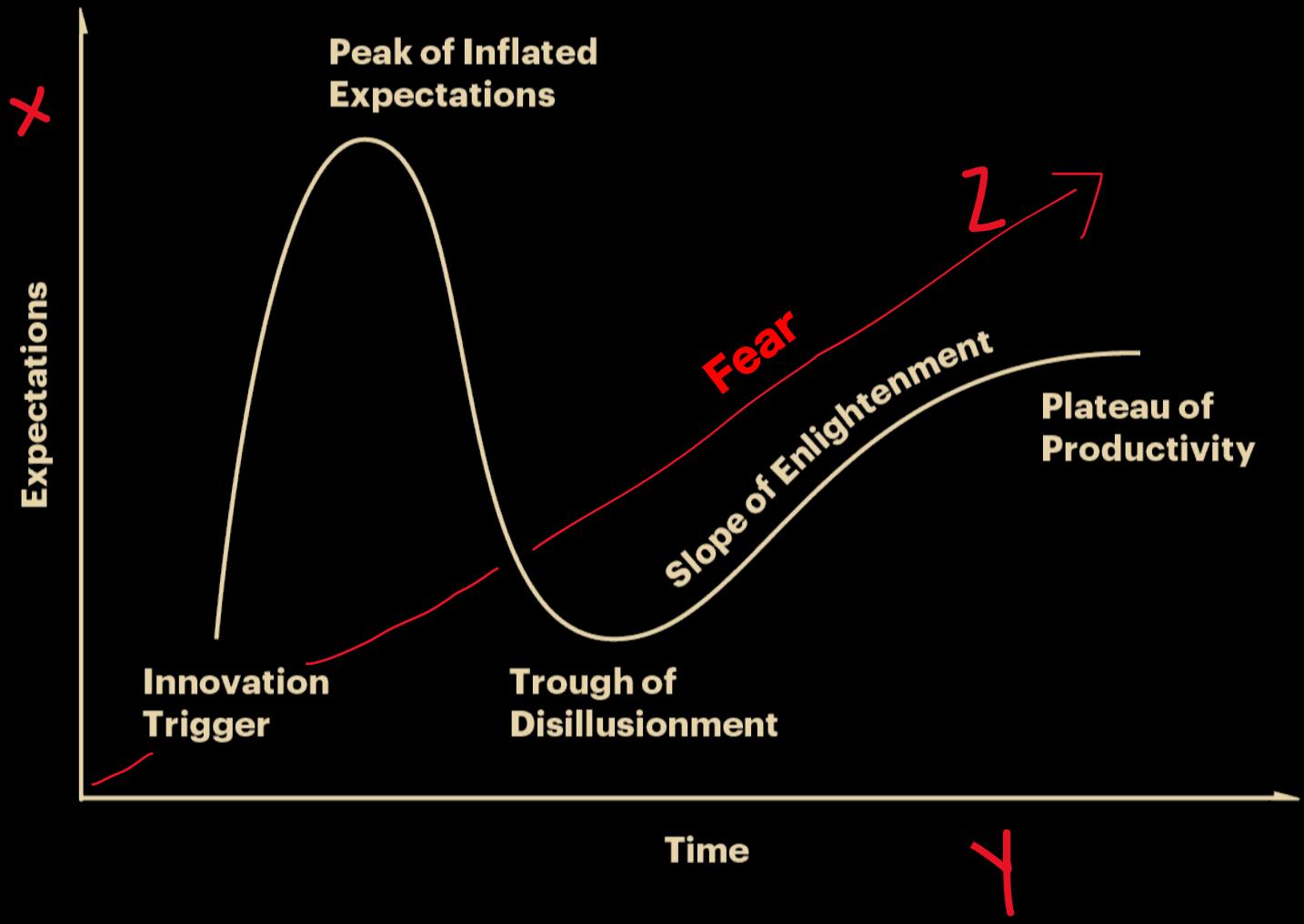
The image shows two overlapping web browser windows against a black background.

The left window displays the Columbia University M.S. in Data Journalism page at journalism.columbia.edu/ms-data-journalism. The title "M.S. in Data Journalism" is prominently displayed in blue. Below it, a call-to-action button says "Apply for the M.S. in Data Journalism degree [HERE](#)". A large section of text describes the program's focus on 21st-century journalism skills like data analysis and presentation. At the bottom, there's an orange "APPLY" button and a "JUMP TO:" menu with links to "Who Should Apply", "Curriculum", and "Classes".

The right window displays the The Lede Program for Data Journalism page at ledeprogram.com. The title "The Lede Program" is in large, bold letters. Below it, a sub-headline reads "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 break out of Excel". A photograph of a laptop screen with a notebook and pen is visible. At the bottom, a section titled "Not just for journalists" includes the text "Pulling insights from data isn't just for journalists" and a "Have questions about ..." button with an envelope icon.







Foundations of Computing

During this introduction to the ins and outs of the Python programming language, students build a foundation upon which their later, more coding-intensive sections will depend. Dirty, real-world data sets will be cleaned, parsed and processed while recreating modern journalistic projects. The course will focus heavily on how to use public resources such as Google and StackOverflow to build self-reliance.

- **Focus:** Familiarize yourself with the data-driven programmatic landscape
- **Topics & tools include:** Python, basic statistical analysis, pandas, CSVs, APIs, git/GitHub, cron, web scraping, StackOverflow, data cleaning, command line tools, and more

Columbia University's Lede Program is a 10-week

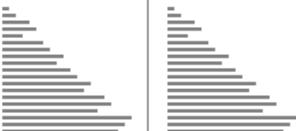
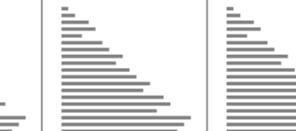
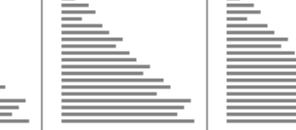
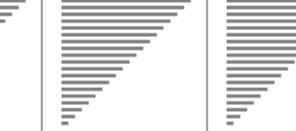
Not just for journalists

Pulling insights from data isn't just for journalists

✉ Have questions about ...

Computer Science:

It's just putting numbers in order!

	 Insertion	 Selection	 Bubble	 Shell	 Merge	 Heap	 Quick	 Quick3
								
 Random								
 Nearly Sorted								
 Reversed								
 Few Unique								

Here's another solution more fleshed out, taken from [Chris Albon's](#)

87 Create "long" dataframe

```
raw_data = {'patient': [1, 1, 1, 2, 2],  
            'obs': [1, 2, 3, 1, 2],  
            'treatment': [0, 1, 0, 1, 0],  
            'score': [6252, 24243, 2345, 2342, 23525]}
```

```
df = pd.DataFrame(raw_data, columns = ['patient', 'obs', 'treatment', 'score'])
```

	patient	obs	treatment	score
0	1	1	0	6252
1	1	2	1	24243
2	1	3	0	2345
3	2			

When using this method (from the book) I get "ValueError: Index contains duplicate entries, cannot reshape" even after I've used df.drop_duplicates() - [d8aninja](#) Oct 25, 2014 at 16:09
@d8aninja I have the same issue, can you tell me how you fixed this? - [Salih](#) Nov 25, 2022 at 5:52

@Salih 8 years later, unfortunately, and I cannot. Certainly with another solution from this post, mixed-in? I'm assuming the issue may have been some form of NaN (etc) in my pivot? - [d8aninja](#) Jan 2 at 13:54

I have the same issue with all the methods I try to convert from long to wide, it simply does not work.
Did any of you manage to fix it? - [baskat](#) Feb 6 at 7:43

```
...count() + 1  
...Salesman', 'Height'],  
values=['product', 'price'], :)
```

```
df = df.sort_index(axis=1, level=1)  
df.columns = [f'{x}_{y}' for x,y in df.columns]  
df = df.reset_index()
```

Output:

	Salesman	Height	price_1	product_1	price_2	product_2	price_3
0	Knut	6	5.0	bat	1.0	ball	3.0
1	Steve	5	2.0	ben	NaN	NaN	NaN

Why are you still using Python 2.7? How about in Python 3? - [devinbst](#) Jan 25, 2018 at 4:16
For python3, you do `from io import StringIO` and use print as a function and all is well. Basic idea of setting the index and unstacking works identically. - [Nate](#) Mar 19, 2018 at 20:16
I find this the only working solution. `pivot` fails because it is unable to carry over columns that are neither index, variable or value and is unable to make use of a multi index. `pivot_table` fails due to its low performance, it can't handle anything larger than a few thousands rows. - [deenees](#) Jan 18, 2022

df['idx']

```
lesman',columns='idx')[['product','price']]
```

	price
2	0 1 2
and	5 1 3
aN	2 NaN NaN

wand	3
NA	NA

Do you really care that the two pivoted variables be interlaced: `product_1 price_1 product_2 price_2 product_3 price_3`? Can they just be `product_1 product_2 ... price_1 price_2`...? - [smci](#) May 24, 2018 at 11:26
Yeah that doesn't matter. - [Luke](#) May 24, 2018 at 17:25
This is a similar long-to-wide question that uses `pivot` and `join`. stackoverflow.com/a/65839968/7434285. - [Toto](#) ...

**THE ENTIRE
INTERNET**

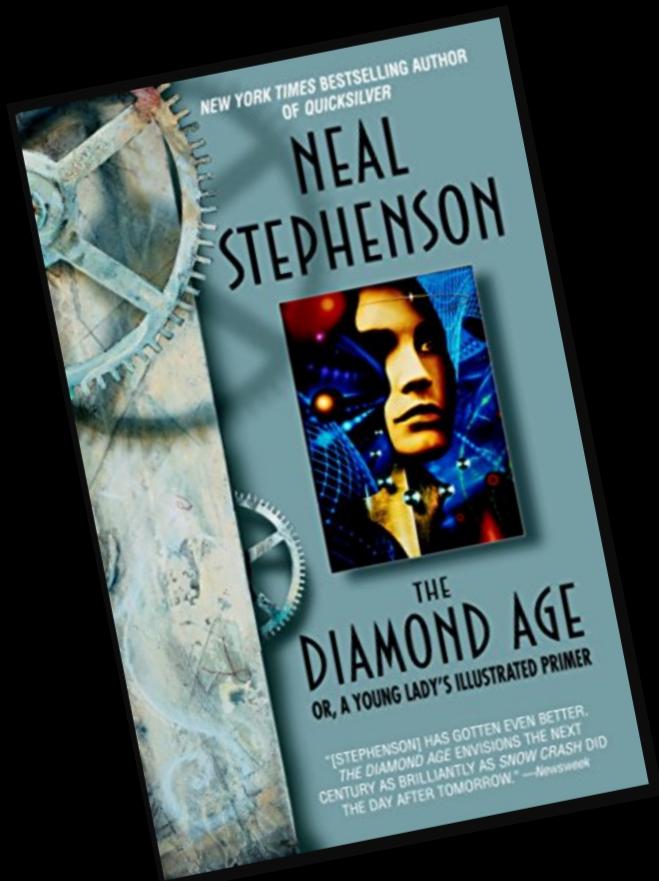
TRAINING
(to learn language)

**IT ALREADY
KNOWS EVERYTHING**
(maybe this is scary???)

**OKAY NOW
WE USE IT!!!!**

GPT and friends

Neal Stephenson – The Diamond Age



```
>>> 2 + '2'
```

```
Traceback (most recent call last):
```

```
File
```

```
TypeE
```

```
JS
```

How do I fix "TypeError: unsupported operand type(s) for +: 'int' and 'str'"?



The error message "TypeError: unsupported operand type(s) for +: 'int' and 'str'" arises when you try to add (using the `+` operator) an integer and a string in Python. This is not allowed because these data types are incompatible in the context of the addition operation.

Here are a few ways to resolve this error:

1. **Explicit Conversion to String:** If you want to concatenate the integer to the string, you need to convert the integer to a string using the `str()` function.

```
python
x = 5
y = "Hello"
result = y + str(x) # Converts the integer to string before concatenation
```

2. **Explicit Conversion to Integer:** If you want to perform a mathematical addition, ensure both operands are integers.

```
python
x = "5"
y = 10
result = int(x) + y # Converts the string to integer before addition
```

3. **Using f-strings (Python 3.6+):** If you want to create a formatted string that includes both strings and integers, you can use f-strings.

```
python
x = 5
y = f"Hello{x}" # The integer is automatically converted to string inside
```

4. **Using the `format()` Method:** This is another way to format strings.

```
python
x = 5
y = "Hello{}".format(x) # The integer is automatically converted to string
```

5. **Check Input:** If the error is arising due to user input or data from external sources, make sure to validate and sanitize the data before performing operations.

If you provide the specific code where you're encountering the error, I can give a more tailored solution.

I'm _____

Soma?

sleepy?

hungry?

I haven't eaten since this morning. I'm hungry

I haven't eaten since two
days ago. I'm **starving**

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

LLMs are not built for accuracy

LLMs cannot *reason*

- They do **NOT** have sources of information, they just keep predicting the next most likely word
- Don't confuse *style* for substance

LLMs can't understand things we (people) take for granted

- Trust, integrity, or relative importance: *not meaningful!*
- Humans care: we pay more attention to important things

But what are they good at doing is stating **the predictably obvious**

JS

write a bio for jonathan soma



JS

write a bio for jonathan soma



Jonathan Soma is a data wrangler, educator, and founder of Brooklyn Brainery, a community-driven education center in Brooklyn, New York. He has a diverse background, with degrees in anthropology and library science, and has worked in a variety of fields including journalism, advertising, and digital media.



In addition to running Brooklyn Brainery, Soma has also taught data visualization and analysis at Columbia University's Graduate School of Journalism and at New York University's Studio 20 program. He has also given talks on data visualization and analysis at conferences around the world, including SXSW, Eyeo, and Strata.

Soma is the author of "The League of Moveable Type," a collection of open-source fonts, and "Frankenfont," a guide to creating your own font. He has also written about data, technology, and education for publications such as The New Yorker, The Atlantic, and Wired.

JS

write a bio for jonathan soma



Jonathan Soma is a data wrangler, educator, and founder of Brooklyn Brainery, a community-driven education center in Brooklyn, New York. He has a diverse background, with degrees in anthropology and library science, and has worked in journalism, advertising, and digital media.



*all very statistically likely,
but all very incorrect*

In addition to running Brooklyn Brainery, Soma has also taught data visualization and analysis at Columbia University's Graduate School of Journalism and at New York University's Studio 20 program. He has also given talks on data visualization and analysis at conferences around the world, including SXSW, Eyeo, and Strata.

Soma is the author of "The League of Moveable Type," a collection of open source fonts, and Frankenstein!, a guide to creating your own font. He has also written about data, technology, and education for publications such as The New Yorker, The Atlantic, and Wired.

Rule of fives

You don't know the predictable tendencies of the machine until you can compare and contrast its answers repeatedly.

By [Sharon Otterman](#), [Maria Cramer](#) and [Wesley Parnell](#)

Published Sept. 24, 2023 Updated Sept. 26, 2023

A few blocks away from where [a 1-year-old boy died](#), possibly from fentanyl poisoning at his Bronx day care this month, an open-air drug market persists along a trash-strewn underpass.

On Wednesday, a man on a moped arrived in the late afternoon and about a dozen users ambled over to purchase drugs with dollar bills in their hands. People tied off their arms, prepped needles or packed pipes. After a while, the ground was littered with syringes and bloodstained alcohol swabs.

No police officers were seen, although they are often nearby — sometimes steps away in the Kingsbridge Road subway station. But they rarely intervene, according to local residents, elected officials who have tried to clean up the area for several years, and nonprofit workers who distribute food and clean needles at the site.

Residents and community leaders say they are regularly told by city officials that they are doing their best to make the area safer. But the Kingsbridge Road underpass is just one of many locations in the city where drug use has become more open, even as lives

```
template = """
Write a headline for the article below:
```

```
{text}
"""


```

```
prompt = template.format(text=article)
response = llm.predict(prompt)
print(response)
```

Open-Air Drug Markets Persist in NYC Amid Rising Overdose Deaths and Softened Law Enforcement



```
template = """
```

Write five headlines for the story below. Limit each headline to around twelve words. Headlines should be elegant, and with the tone of the New York Times.:

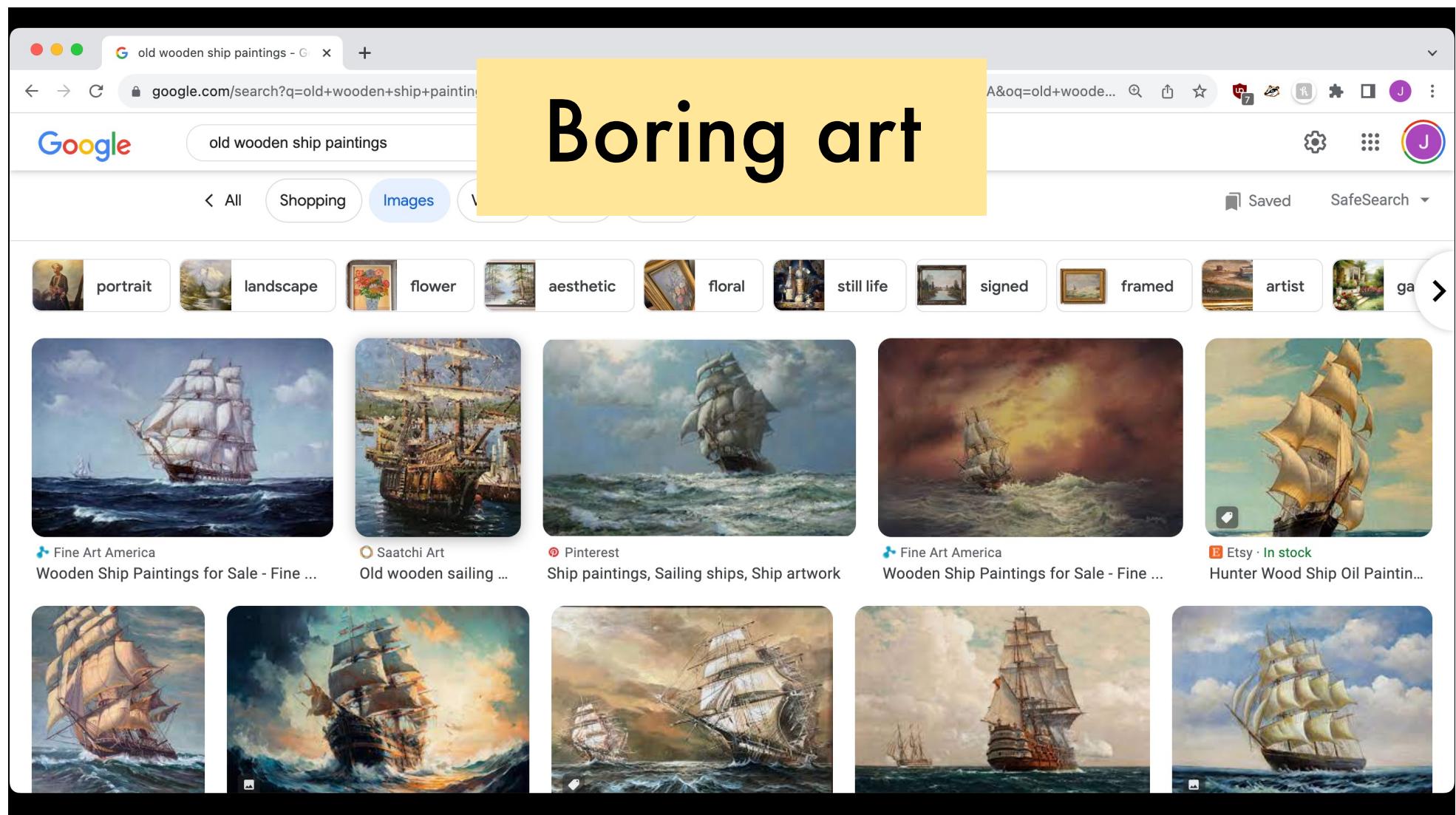
1. "Open-Air Drug Market Thrives Near Bronx Day Care Amid Fentanyl Crisis"
 2. "Bronx Residents Confront Open Drug Use Amidst Rising Overdose Deaths"
 3. "NYC Struggles to Curb Drug Epidemic as Overdose Deaths Hit Record High"
 4. "Bronx Community Grapples with Visible Drug Use and Fentanyl Threat"
 5. "Drug Market Flourishes in Bronx as City Battles Opioid Overdose Crisis"
- 

The New York Times

A Day Care Death and the Dilemma Over How to Crack Down on Drugs

Fentanyl use is increasingly out in the open, and increasingly fatal. But New Yorkers are divided over what to do: decriminalize it or make more arrests.

It won't replace journalists if
it doesn't know fact. But what
about the future?





Shad M. Brooks 
@shadmbrooks

My skills are increasing.

Making a video as we speak showing my process and how I get these results. The level of artistry needed, the technical knowledge required, and amount of work involved I think is going to shock people.



MY AI ASSISTED ART
SKILL 6 MONTHS AGO



DON'T LET PEOPLE TELL YOU AI ART REQUIRES NO SKILL AND YOU 'JUST' PRESS A BUTTON TO MAKE AI ASSISTED ART

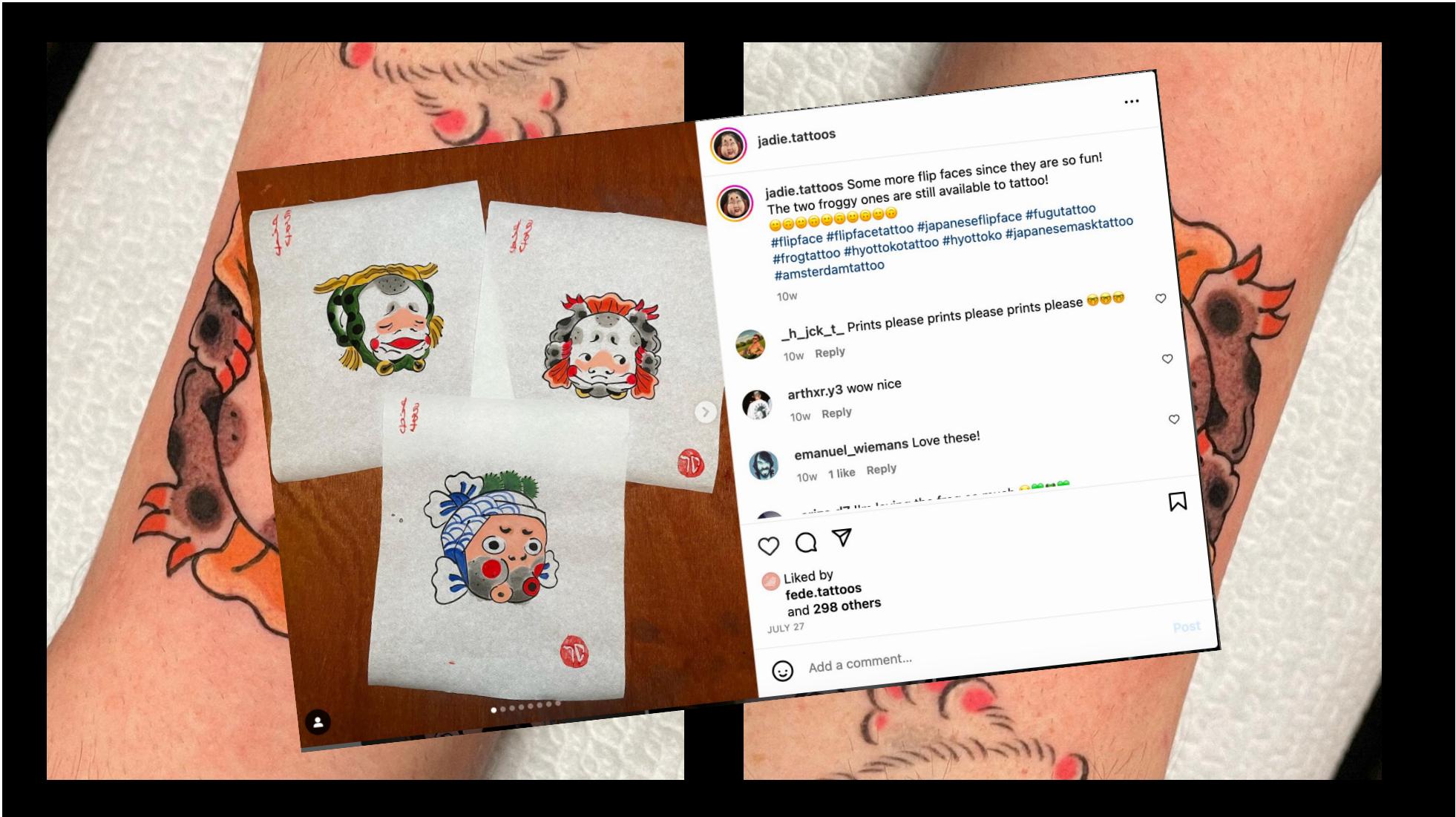
Last edited 6:00 PM · Oct 8, 2023 · 12.6K Views

 Readers added context they thought people might want to know

The above is simply AI-generated art. No "skills" are being improved on for the user, rather it's the technology that's improving.

He made nothing except type a prompt into a text box and clicked a button while it did the work for him.

SKILL NOW



robbanks
ofamerica

Following Message +9 ...

1,035 posts 27.3K followers 579 following

ROB BANKS

Artist

Five Points Tattoo @fivepointsnyc
fivepointstattoonyc@gmail.com

Followed by bigolebrat, cakpatrick, caue.gottardi + 35 more

POSTS REELS TAGGED

A collage of tattoo images and a green smoothie. The tattoo images include a large ship on a leg, a butterfly, a skull, a heart with cherries, and a grid of various designs. The green smoothie is in a clear cup with a straw.

A large tattoo on a person's arm featuring several stacked barrels. The word "TATTOO" is written vertically along the side of the barrels, and "AMERICA" is written across the top barrel. The barrels are shaded with dark ink.

Are you being creative?
Are you designing it?

Not a unique relationship!

A client to a tattoo artist

An editor to an intern

(Or a reporter to the data/graphics rep)

A patron to an artist

But who gets the credit?

= WIRED

SUBSCRIBE

Why This Award-Winning Piece of AI Art Can't Be Copyrighted

Matthew Allen's AI art won first prize at the Colorado State Fair. But the US government has ruled it can't be copyrighted because it's too much "machine" and not enough "human."



What is the role of execution in the birth of a creative work?

And what counts as execution?

Journalism isn't just a list of facts

The story being written in the first place

The angle the story is being written from

Who is cited as a source

Why you're writing the story

Who you are writing the story for

Word choice, "...XXX died when..." vs "...XXX were killed when..."

What pieces might an AI-generated story take away?

ARTIFICIAL INTELLIGENCE / TECH

CNET found errors in more than half of its AI-written stories



"CNET issued corrections on **41 of the 77 stories** the outlet published that were written using an AI tool"



Illustration by Alex Castro / The Verge

The tech news site came under scrutiny this month after it was reported the outlet was using AI to write articles. Last week, executives said the use of the AI tool would be paused 'for now.'



Futurism



UNCUT GEMS | UPDATED MAR 30 by NOOR AL-SIBAI AND JON CHRISTIAN

BuzzFeed Is Quietly Publishing Whole AI-Generated Articles, Not Just Quizzes

These read like a proof of concept for replacing human writers.

/ Artificial Intelligence / Artificial Intelligence / BuzzFeed / Media



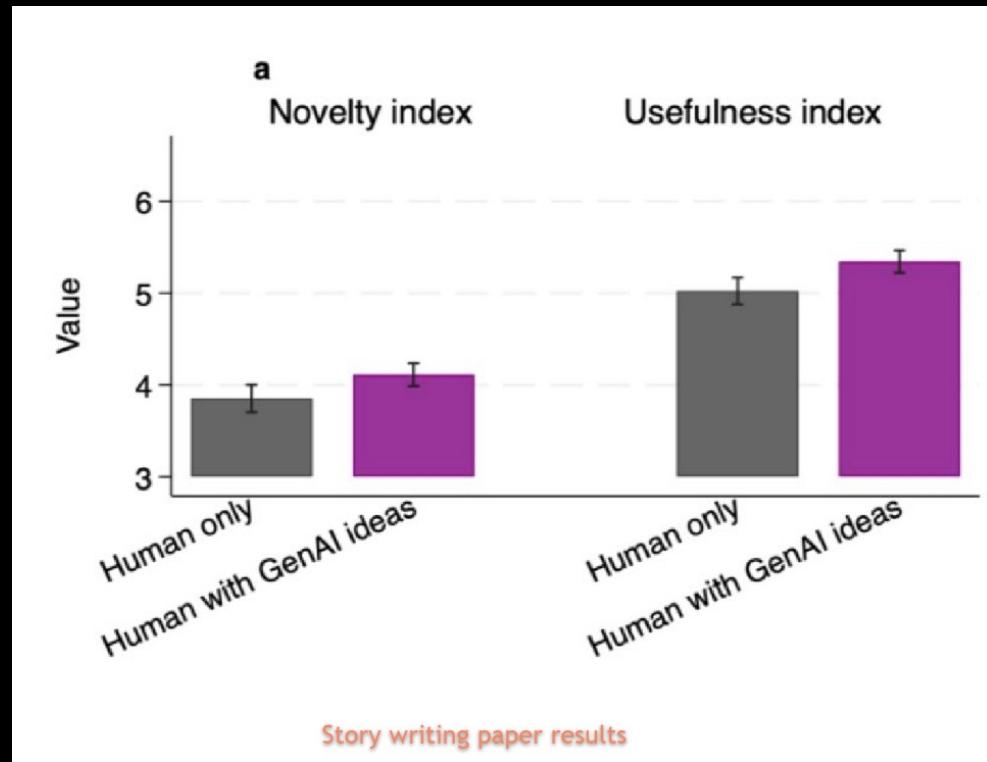
That's not the bot's only lazy trope. On review, almost everything the bot has published contains at least one line about a "hidden gem."

Amelia Island, Florida is a "hidden gem of beaches," Carmel-By-The-Sea, California is a "hidden gem of California's coast," West Virginia is a "hidden gem of a state," Saugerties, New York is a "hidden gem where small town charm meets big city cool," Stanley, Idaho is a "hidden gem nestled right in the heart of the Sawtooth Mountains," Brewster, Massachusetts is "Cape Cod's hidden gem," Calistoga, California is a "hidden gem," June Lake, California is a "hidden gem," Mammoth Mountain, California is a "hidden gem," Providence, Rhode Island is a "hidden gem," Charleston, South Carolina is a "hidden gem," Connecticut is a "hidden gem," Aruba is a "hidden gem" and "truly a gem," Prague is a "hidden gem," the Cook Islands are a "hidden gem," Saint Maarten is a "hidden gem," South Dakota is both a "glorious gem of a state" and a "criminally underrated travel gem," Cape May, New Jersey is a "gem worth visiting," Arizona is a "desert gem," Ecuador is an "absolute gem," Bruges, Belgium is a "secret gem," and a "Belgium gem," Montreal, Canada is a "travel gem," and Alberta, Canada is a "Canadian gem," Sevilla, Spain is just a "gem."

Use AI for resilient tasks.

Art is resilient against “errors.” What else is?

What if there’s a human editor? Will it pass the self-driving car test?



AI-assisted ideas are **better than human ideas** (on average)

I haven't eaten
since this morning.
I'm **hungry**

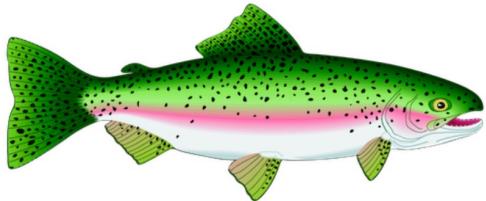
But they're also **similar to each other** (on average)

Explore shortcomings and gaps

Experiment, experiment, experiment

Don't just ask for chocolate chip cookie recipes as poems

Sadly: Twitter



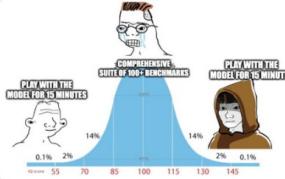
aifaq.wtf

"How do you know about all this AI stuff?"
I just read tweets, buddy.

Want suggestions? Try [bias](#), [labor](#), [hallucinations](#), or [tragedy](#)

Jason Wei  Follow ...

Large language models are notoriously hard to evaluate because (1) they are highly multi-task, (2) they generate long completions, and (3) grading is subjective. After spending ~5 months rigorously working on how to do language model evals, this is my verdict:



9:39 PM · Sep 27, 2023 · 150.3K Views
111 Reps 21 Quotes 1,223 Likes 187 Bookmarks



imagine a tiny little walrus. they're about the size of a loaf of bread you call them a 'snowbank' and put them in the fridge so they chill

David Shadoff  @Shadoff_d

Like "pasta is all 'spaghetti'" ?

2:19 PM · Sep 24, 2023 · 3,904 Views
2 Likes



Joseph Cox  Follow ...

New: a TikTok account is using facial recognition to do random people simply for clout with its millions of viewers. Takes requests from the comments
- spoke to multiple victims, "violated"
- TikTok refuses to remove because says doesn't violate policies



Armand Domalewski  Follow ...

Islamofuturism

"Robots can't replace senior clerics, but they can be a trusted assistant that can help them issue a fatwa in five hours instead of 50 days," said Mohammad Ghobbi, who heads a state-linked organisation in

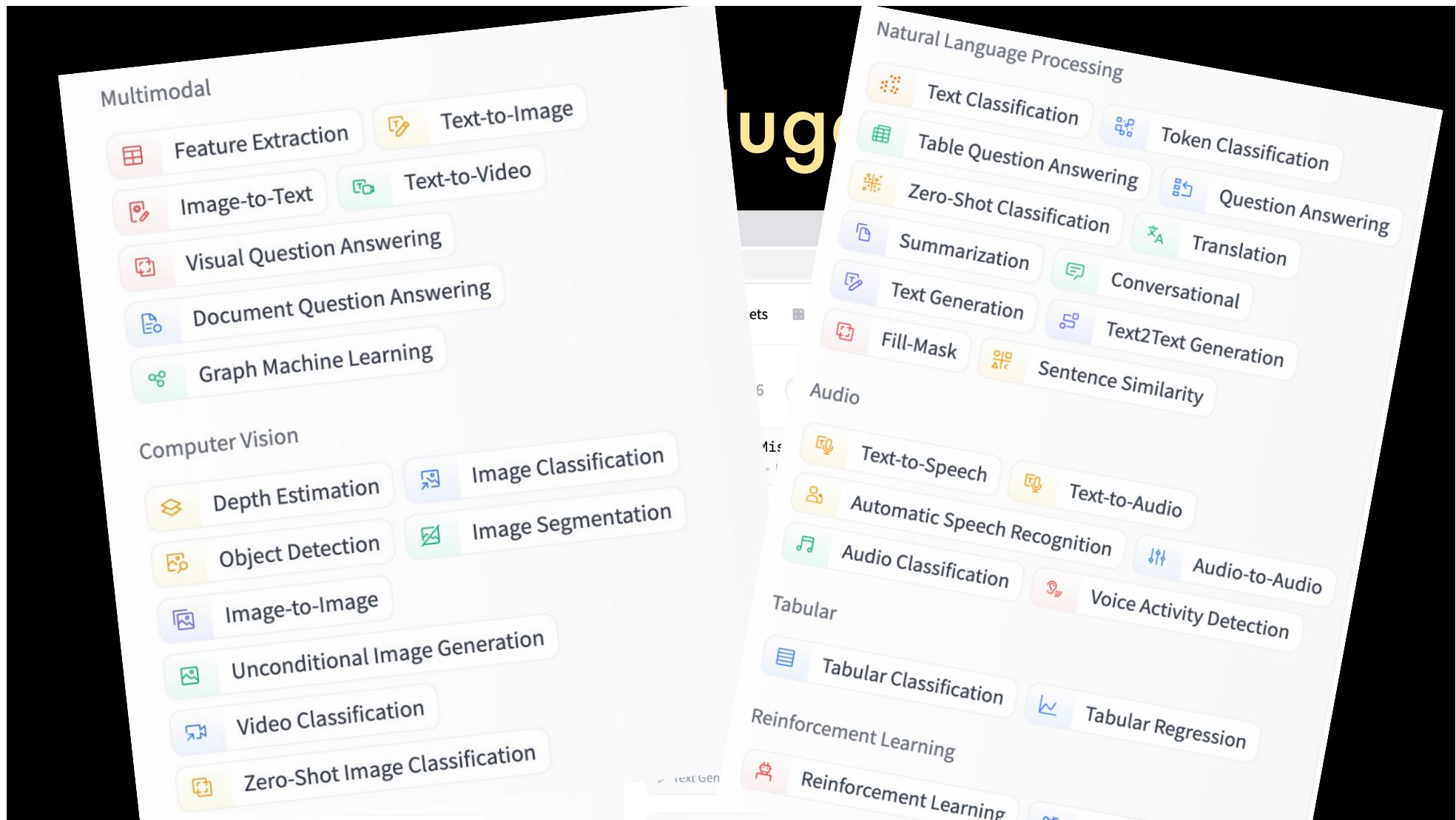
Ilya Sutskever  @ilyasut 21h

In the future, once the robustness of our some threshold, we will have "wildly effe AI therapy. Will lead to a radical improve experience of life. One of the applicatio awaiting. twitter.com/lilianweng/sta...

10:25 PM · Sep 27, 2023 · 56.9K Views
27 Retweets 4 Quotations 270 Likes 390 Retweets

models
"tools"





Object detection

img



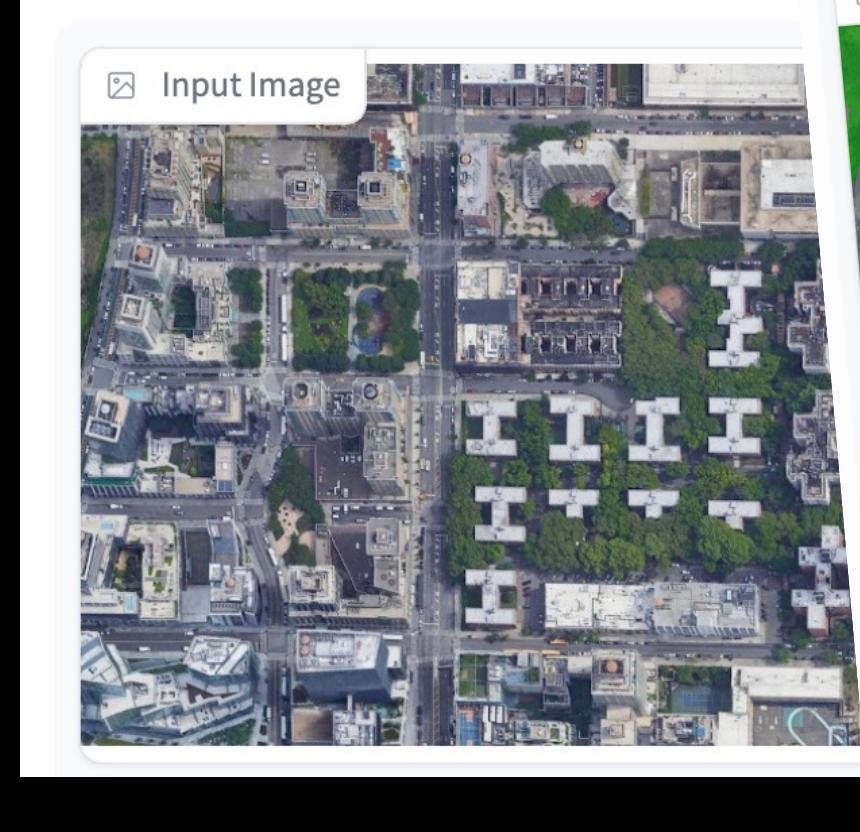
text_queries

human face, rocket, star-spangled banner, nasa badge

score_threshold



Semantic segmentation



Info		
Object Label	Pixel Percent	Square Length
vegetation	15.46 %	0.39 m



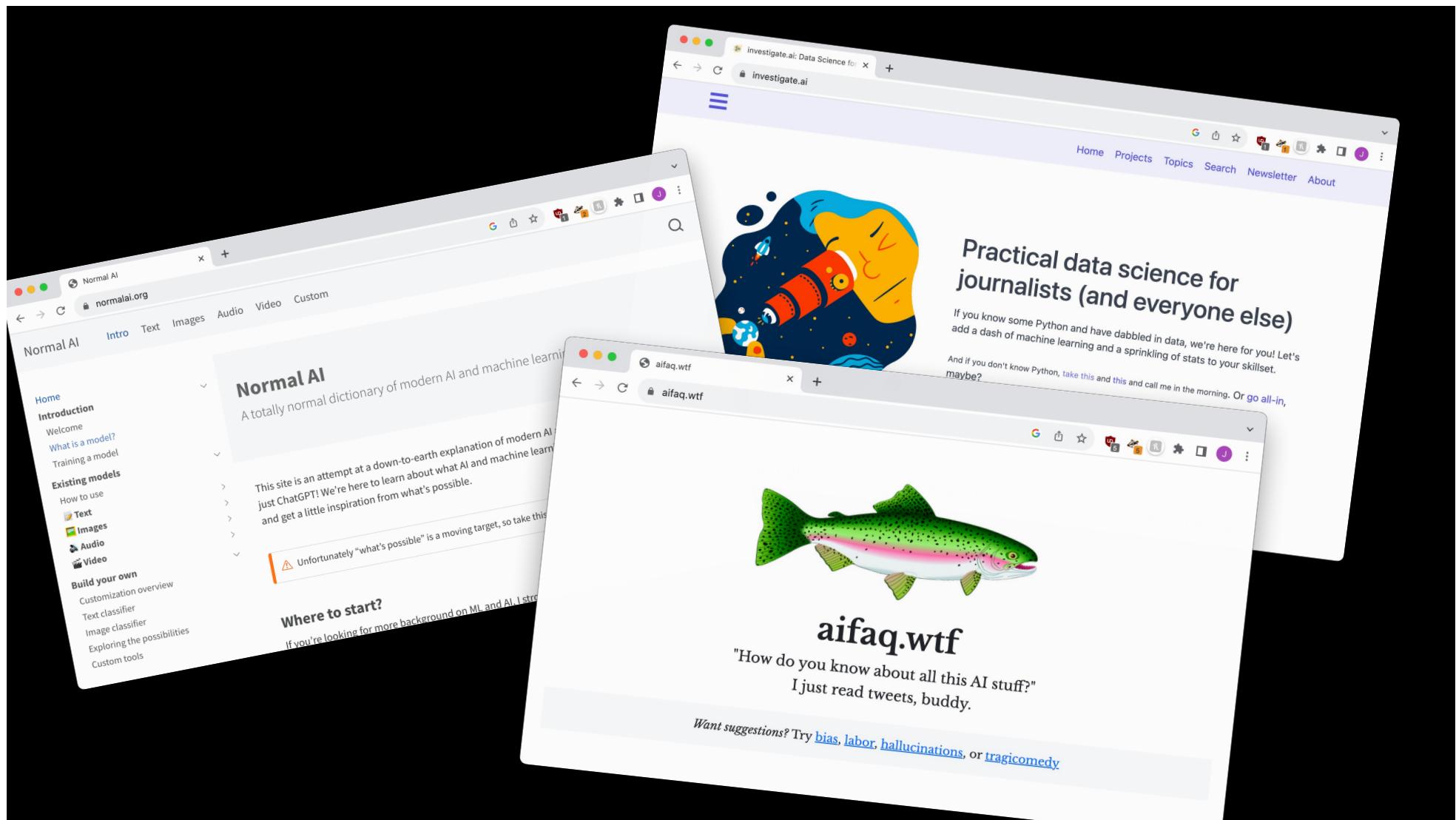
*Figure 2: Example images from the collected **Skull2Dog** dataset.*

Build tools, play with tools

It's how you learn, it's how you see what's possible

Where to learn more?

Especially when it's all changing so fast



Resources for advancement

Newsletters: Futurism, Ethan Mollick, Import AI, AlphaSignal

Orgs: JournalismAI at LSE, Associated Press

Twitter: ...it's too crazy right now, just find some lists

Hugging Face: huggingface.co/models or /spaces

Discord: Search topic you want + “discord”

Science Fiction: The Diamond Age!!! And... everything else.

In the end: Just go out there and try it all out!!!!

Alspiration

Reimagining the Humanities and Art with Generative AI

Jonathan Soma

Knight Chair in Data Journalism

Columbia University Graduate School of Journalism

js4571@columbia.edu

investigate.ai • normalai.org • aifaq.wtf