



# How Vector databases and embeddings can change how the web is created

**Mike Curtis**

7DOTS

 @hello7dots

 [7dots.com/brighton-seo](https://7dots.com/brighton-seo)



THE WEB AS WE KNOW IT ONLY

HAS 5 YEARS LEFT TO LIVE

**DEAL WITH IT**

# HEY, HAVE YOU HEARD OF THIS THING CALLED CHATGPT?



**You**

come up with a joke about vector databases and embeddings.

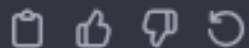


**ChatGPT**

Sure, here's a light-hearted joke about vector databases and embeddings:

Why did the vector database go to therapy?

Because every time it tried to get close to its embeddings, it ended up with too much cosine similarity and lost its sense of direction!



**“ WE MADE OUR  
DOCUMENTATION  
AVAILABLE VIA A CHAT BOT”**

*What on  
earth does  
that mean!?*



# YOU AIN'T FITTING MUCH OF A DOCUMENT WITHIN THESE LIMITS!

<code>gpt-3.5-turbo</code>	Currently points to <code>gpt-3.5-turbo-0613</code> .	4,096 tokens
<code>gpt-3.5-turbo-16k</code>	Currently points to <code>gpt-3.5-turbo-0613</code> .	16,385 tokens

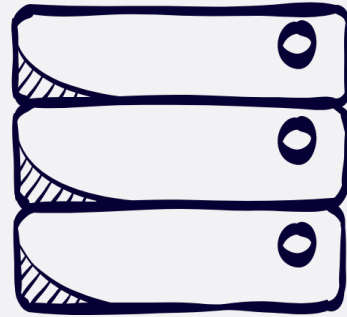
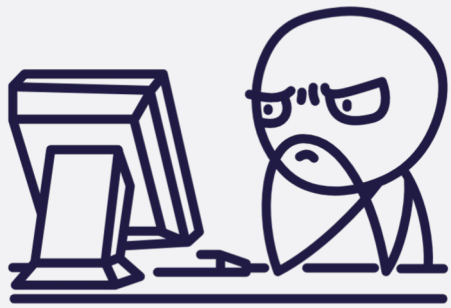
*~2,000 words*



**SO, WHAT ARE  
THEY DOING?**

# THE INTERMEDIATE STEP

## RETRIEVAL AUGMENTED GENERATION (RAG)



# SQL IS AN OLD, LIMITED LANGUAGE

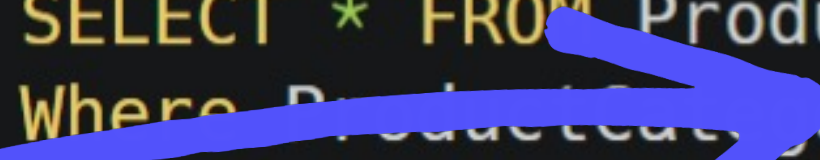
## Department

### Books

Antiquarian, Rare &  
Collectable  
Arts & Photography  
Biography  
Business, Finance & Law  
Calendars, Diaries &  
Annuals  
Children's Books  
Comics & Graphic Novels



```
SELECT * FROM ProductDatabase  
Where ProductCategory = 'biography'
```



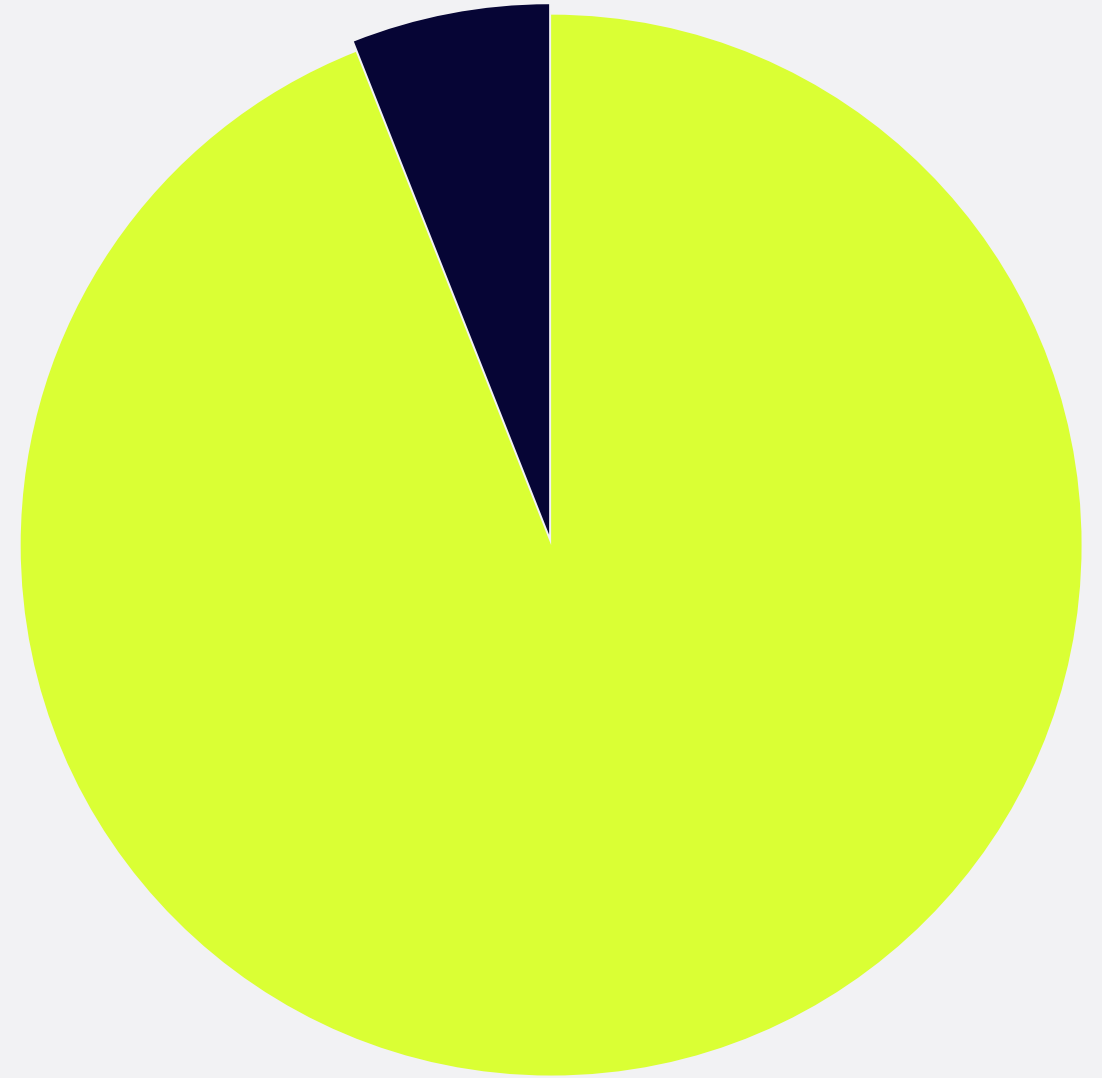


```
SELECT * FROM ProductDatabase
Where ProductCategory = 'biography'
& type = 'sport'
& game = 'football'
& league = 'epl'
& date = [2010,2011,2012,2013,2014]
....
```

# THIS IS NOT SUITED TO THIS SORT OF LONG FORM QUERY

**“Show me some books that are biographies about footballers in the 2010s who played in the English Premier League....”**

**95% OF SEARCHES  
ARE LONG TAIL TERMS  
WITH A VOLUME <10**



Source: **Search Vol Distribution**, *Ahref's*

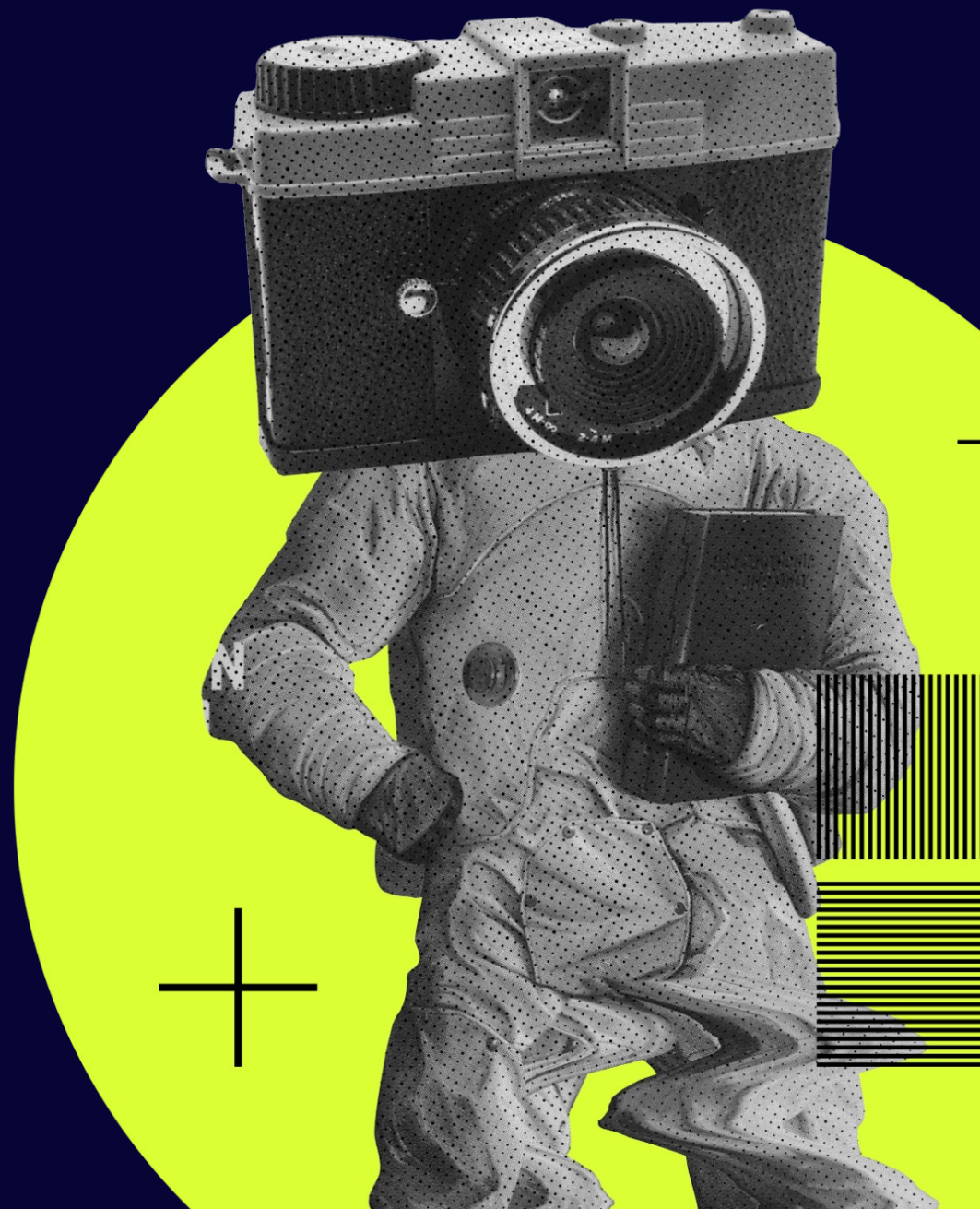
**A NEW CLASS OF 'VECTOR  
DATABASES' ARE EMERGING,  
TO FACILITATE THIS SORT OF  
NATURAL LANGUAGE DATA RECALL**



**A VECTOR DATABASE STORES DATA AS A SERIES OF NUMERICAL EMBEDDINGS, PLOTS THEM ON A VECTOR GRAPH, AND RETURNS DATA BASED ON THE RELATIONSHIP BETWEEN THOSE EMBEDDINGS.**



**7DOTS**  
**THANK YOU**  
**BRIGHTON!**



**GROSS  
OVERSIMPLIFICATION  
INCOMING!**

# EMBEDDINGS TURNS WORDS INTO NUMBERS

	SHOW	BOOKS	BIOGRAPHIES	FOOTBALL
SHOW	1	0	0	0
BIOGRAPHIES	0	0	1	0
FOOTBALL	0	0	0	1
BOOKS	0	1	0	0

*BOOKS = 4,2*

A hand-drawn scatter plot on graph paper. The plot features a vertical y-axis and a horizontal x-axis, both drawn with black ink. The axes are marked with small black dots. Three blue labels are placed on the graph: 'BIOGRAPHY' is located in the lower-left quadrant, 'SPORT' is in the upper-right quadrant, and 'FOOTBALL' is in the lower-right quadrant. The background is a grid of small black dots.

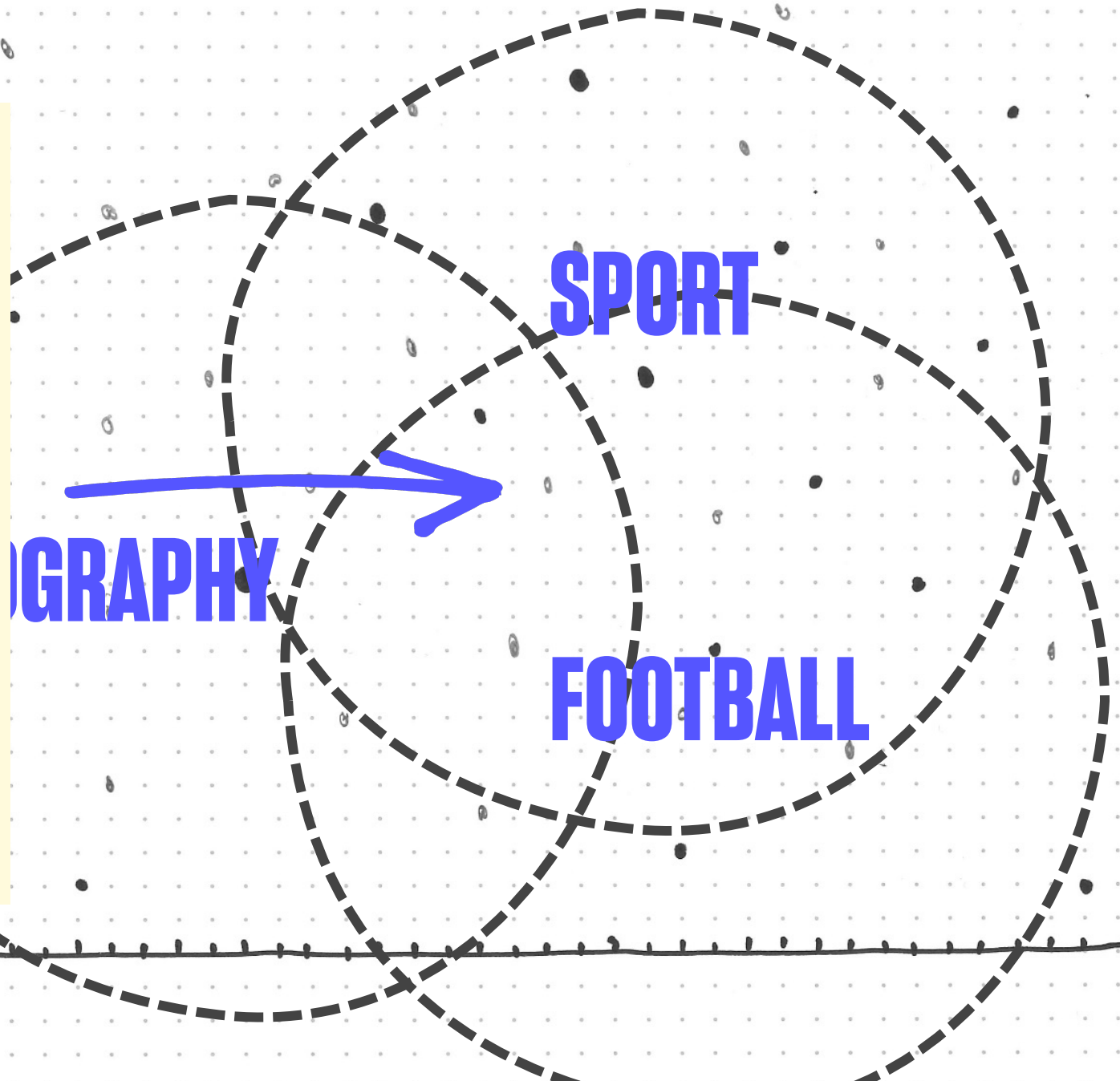
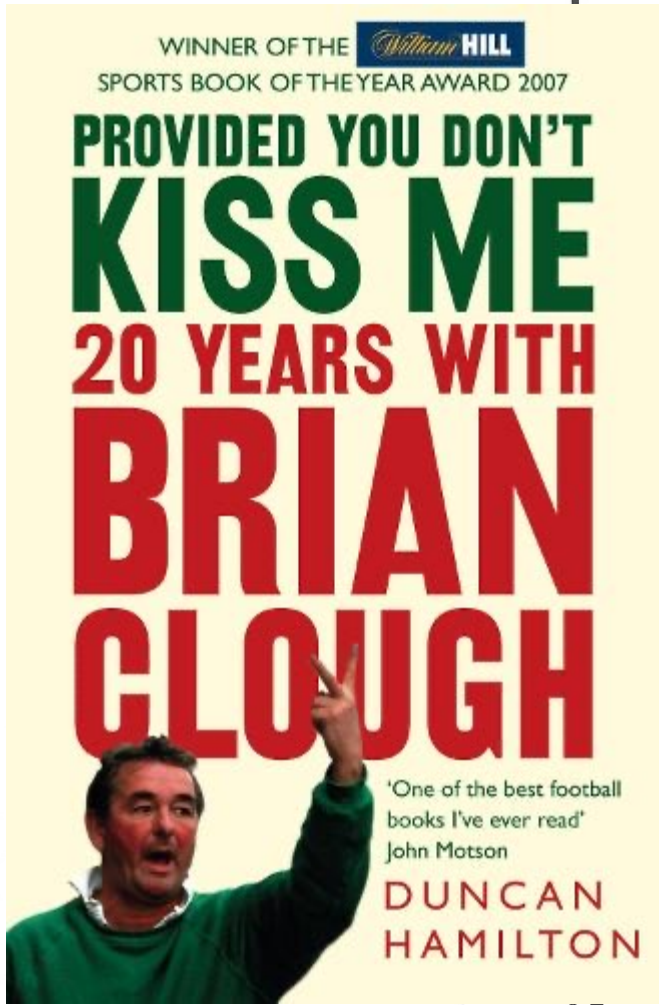
**BIOGRAPHY**

**SPORT**

**FOOTBALL**



**VECTOR DATABASES CAN  
RECALL DATA BASED ON IT  
BEING CLOSE TO THE  
LANGUAGE YOU ENTERED**




# THIS IS KINDA WHAT GPT BASICALLY IS...



**You**

please don't take my job and make me homeless and destitute



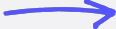
**IN REAL WORLD USAGE, THERE'S  
MULTI-SHOT MODELS ACROSS  
DOZENS OR HUNDREDS  
DIMENSIONS....**

**1. TECHNOLOGY ENABLES TOOLS.**

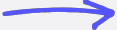
**2. TOOLS ENABLE CULTURE.**

# TODAY'S TOOLS ARE LARGELY SQL BASED

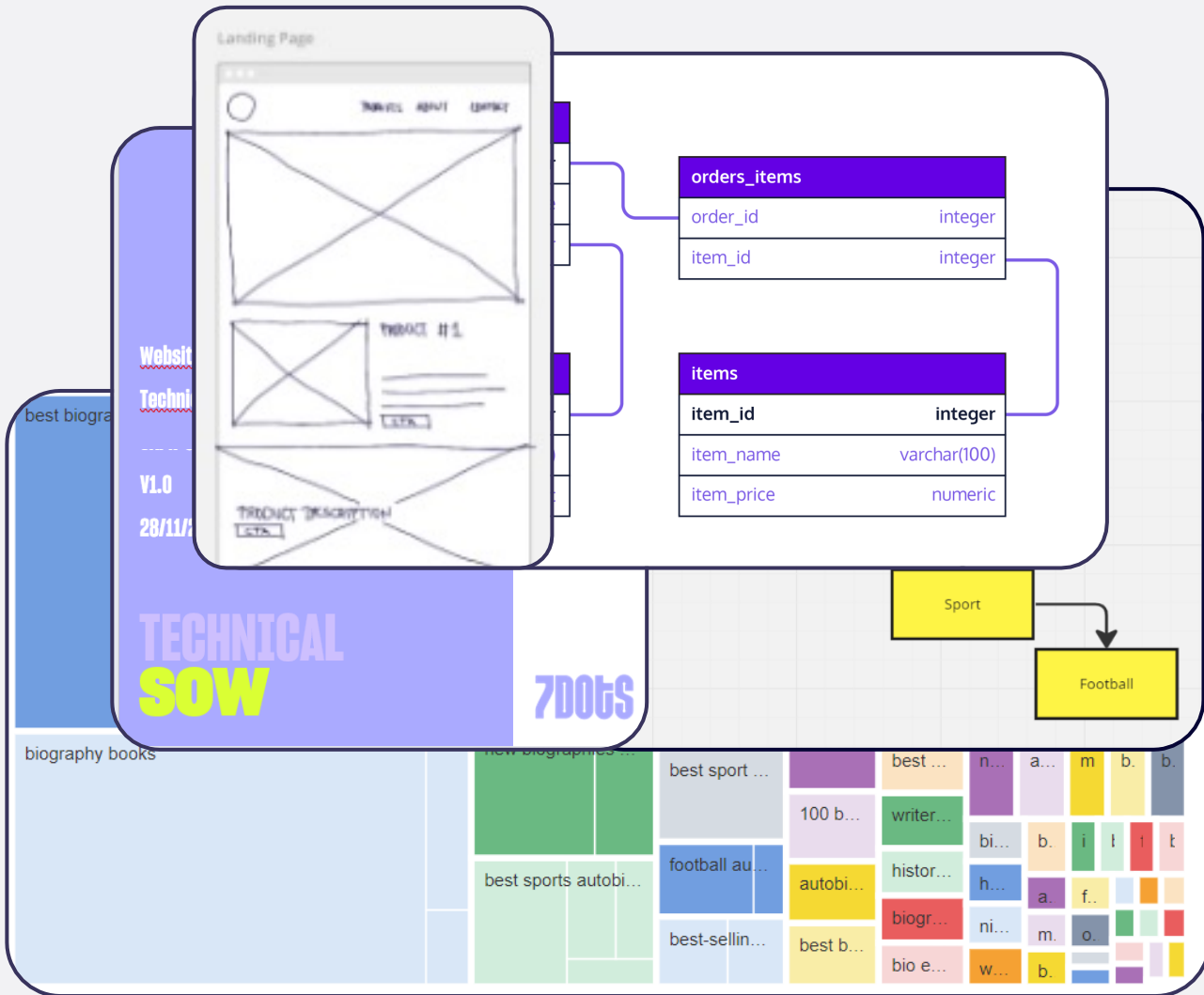
```
GET '/books/biographies',|  
to: 'templates#productListingPage'
```



```
SELECT * FROM ProductDatabase  
Where ProductCategory = 'biography'
```

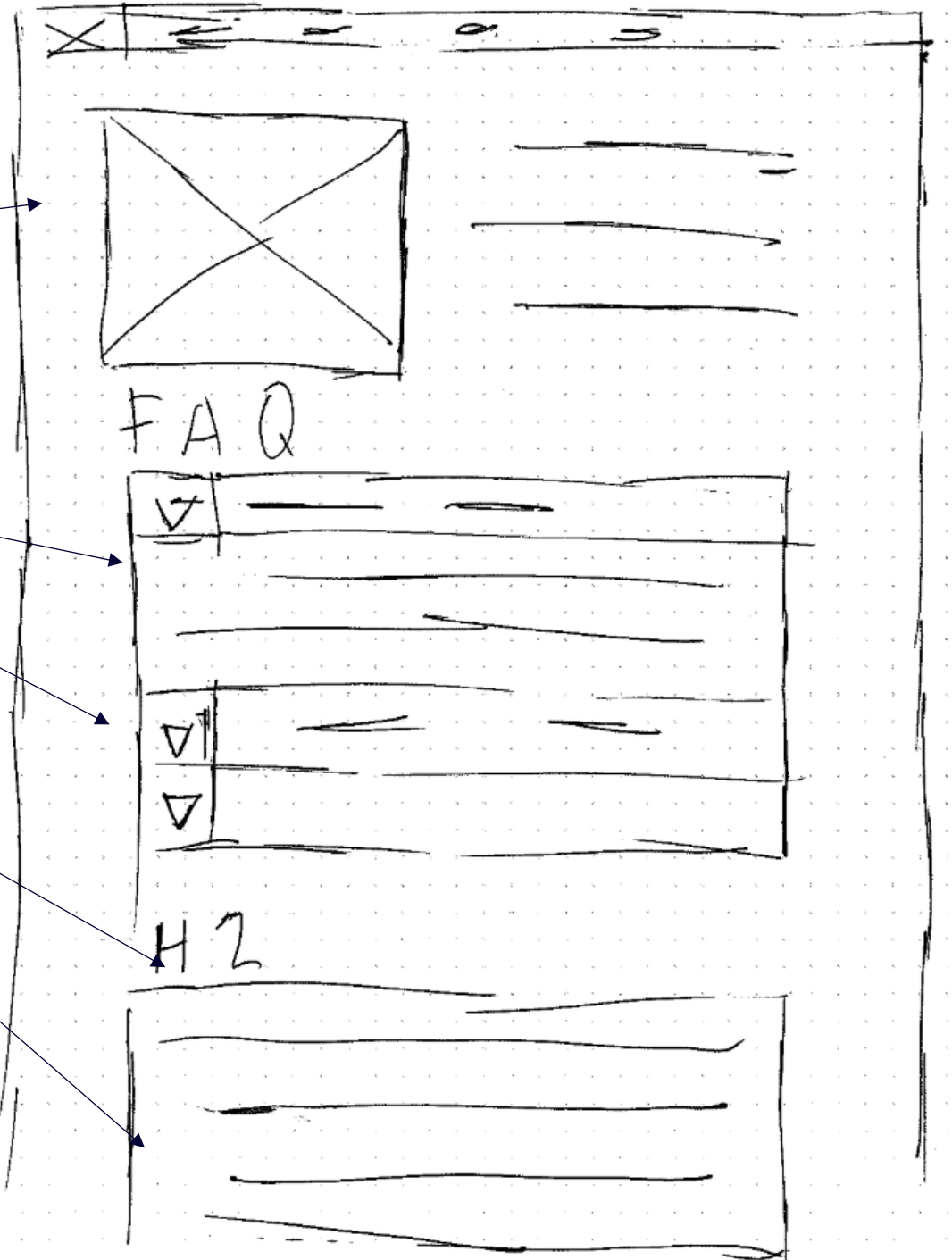
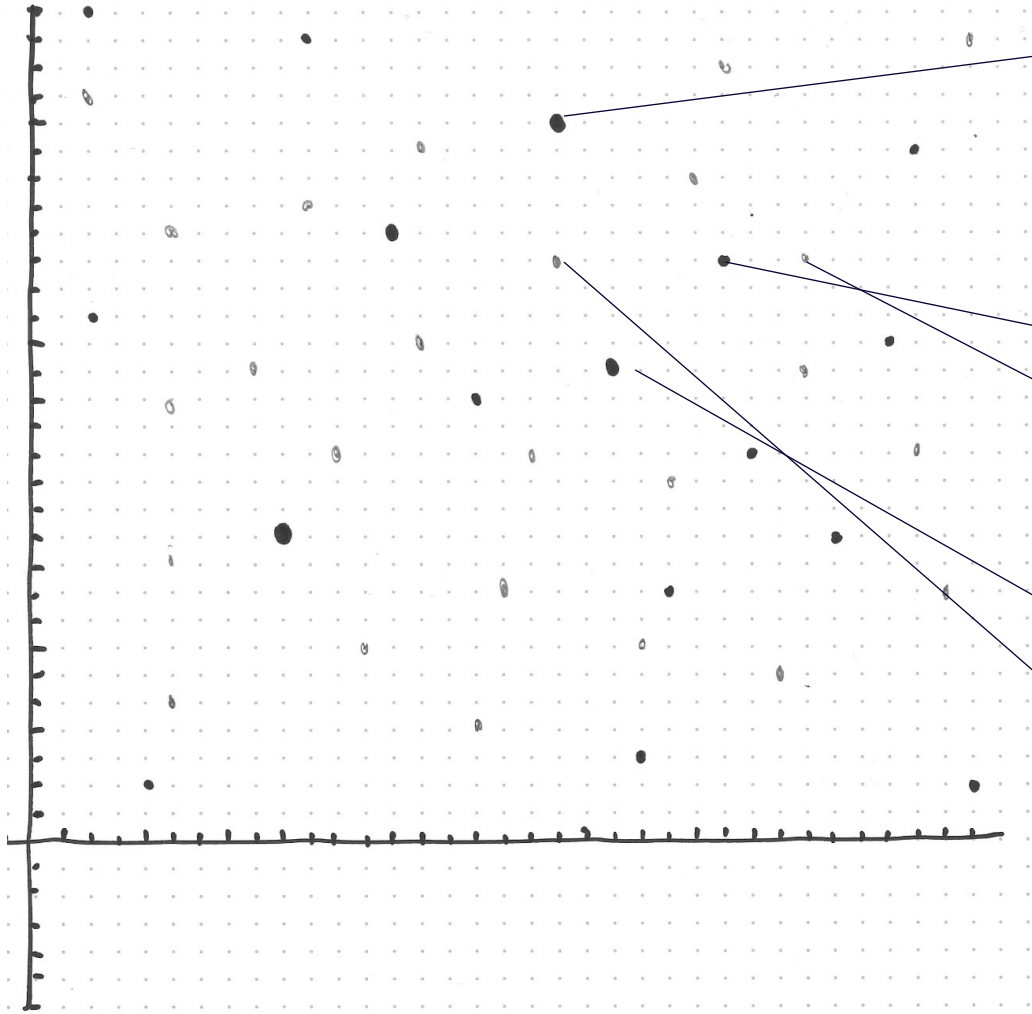


# HOW AND WHAT WE CREATE IS DICTATED BY THOSE TOOLS



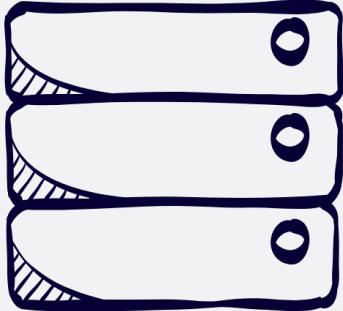
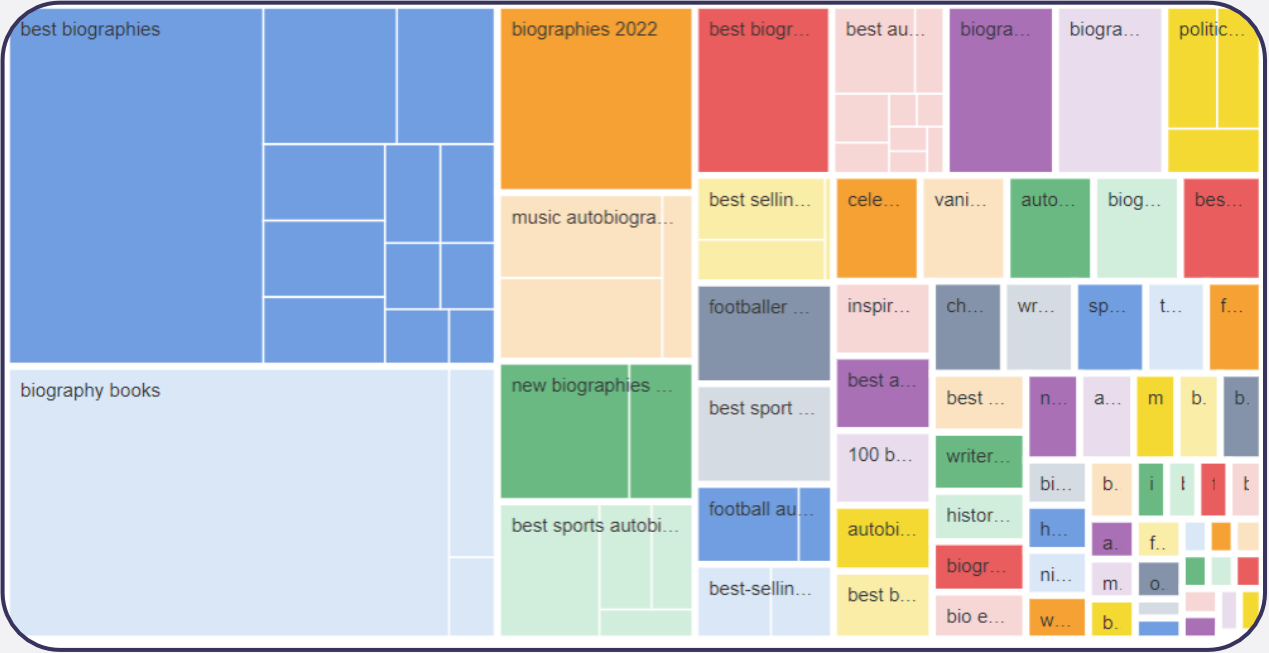
**WHAT HAPPENS IF  
WE SWAP OUT SQL?**







# OUR JOB BECOMES TO IDENTIFY USER 'STATES' AND POPULATE THE WAREHOUSE



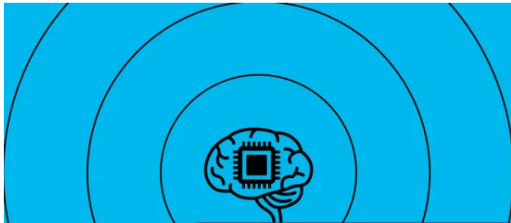
[OBLIGATORY  
BRIGHTONSEO  
CROCODILE CONTENT  
GOBLIN PARTY SLIDE]



# THERE IS A REASON SO MANY COMPANIES ARE BLOCKING CRAWLERS!

ARTIFICIAL INTELLIGENCE / TECH / POLICY

## The New York Times blocks OpenAI's web crawler

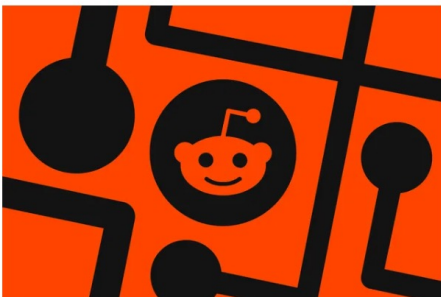


/ The NYT's robot.txt page that controls how it appears to automated bots built to index the internet now specifically disallows OpenAI's GPTBot.

REDDIT / TECH / ARTIFICIAL INTELLIGENCE

## 'Reddit can survive without search': company reportedly threatens to block Google

Illustration by Alex Cas



/ Reddit initially denied a report from The Washington Post that it might force users to log in to see content – but the Post says it may still block search crawlers.

By Jay Peters, a news editor who writes about technology, video games, and virtual worlds. He's submitted several accepted emoji proposals to the Unicode Consortium.

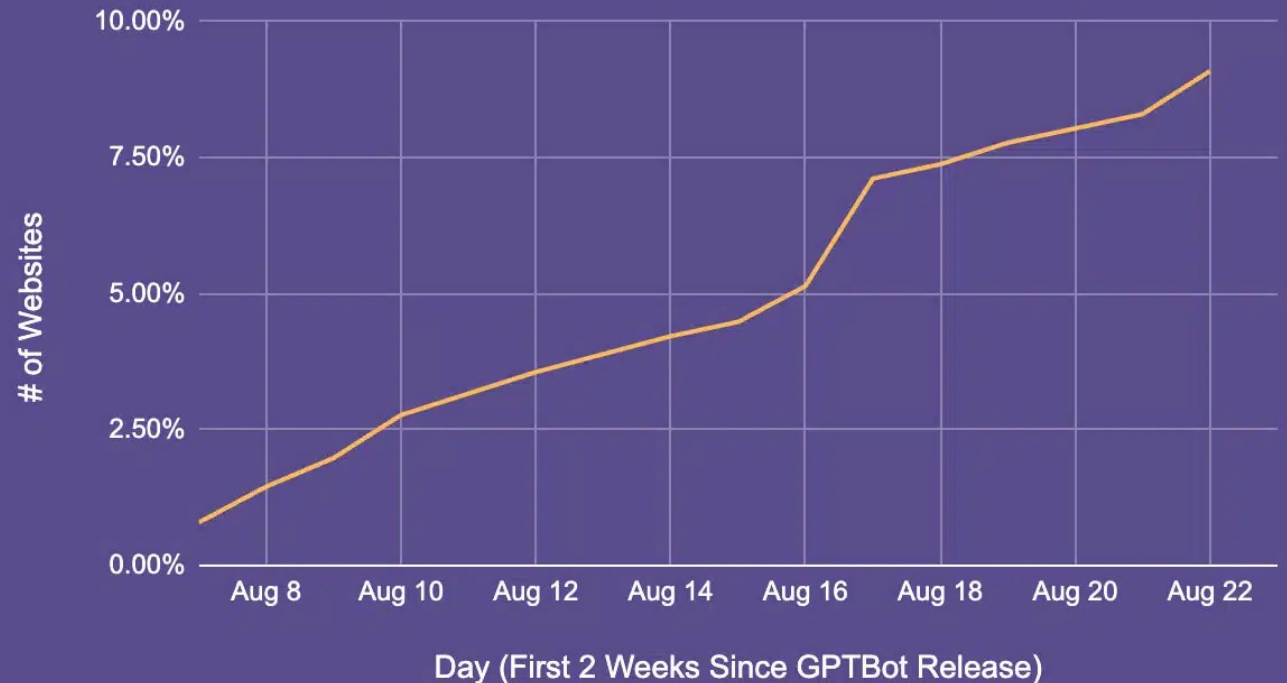
Oct 20, 2023, 6:55 PM GMT-1 | [67 Comments](#) / [67 New](#)



Illustration by Alex Castro / The Verge

Source: **The Verge** (*yeah, yeah I know*)

## % of Top 1000 Websites that Have Blocked GPTBot



Source: **Search Engine Land**

# DOES OUR CONTENT JUST BECOME A FEED INTO GENERATIVE AI?

The screenshot shows a Google search result for ebikes. At the top, there is a search bar with the text '< Search' and the Google logo. Below the search bar, a blue banner reads 'Generative AI is experimental'. The main text of the search result states: 'Ebikes can be a good choice for hilly commutes as they come with motors that provide extra power for cycling up inclines. Compared to standard bicycles, ebikes can reduce your commute time by almost half.' Below this, it says: 'Here are some of the best ebikes for five-mile commutes with hills that are available in red:'. There are two product listings. The first is the 'Aventon Adventure Step-Through Ebike', described as 'Good for hill climbing'. It features a step-through design, 4" fat tires, front suspension, an 8-speed drivetrain, and a 750W motor. It has a 4.8 star rating from 2,200 reviews and is priced at \$1,499. The second listing is the 'VIVI H7 High-Step Commuter Electric...', described as 'This 21-speed ebike has 27.5" tires to handle difficult terrain and is designed for mountain biking.' It has a 4.7 star rating from 46 reviews and is priced at \$549.



EXPERIMENTS  
WOT WE GONE  
AND DONE

# M.E.O.W

M.E.O.W



Keywords

Analyse

Jobs

Please Select

Plot

Keyword

Group by

Topic

Reset Z

100

80

60

40

20

0

-20

-40

-60

-80

-100

Term: google's sge

Volume: 1000

sgc generative ai in search: (-14.12, 72.375, 3)

cookie ch

website gd

dv360 view

shl case st

dyslexia sh

dwwindson

you need a

digital evol

linkedin co

sgc in seo

add cookie

hampsteac



# THE VECTOR REDIRECTOR

**7Dots** Migrator

- Dashboard
- Projects**
- Jobs

< Client: The Dyslexia Shop Project: SEN Pro Decommission

Name	nURLs
Destinations	3949
Sources	1535

Redirect Maps

Name
SEN Pro Decomission

7DOTS

**THANK YOU!**

[www.7dots.com](http://www.7dots.com)

[www.searchtoolbox.net](http://www.searchtoolbox.net)

<https://www.linkedin.com/in/michael-curtis-0571882b/>