

DATA IS BUSINESS

BUSINESS IS DATA

THE 2023 AI REDUX

JASON BELL
NIDC 2023

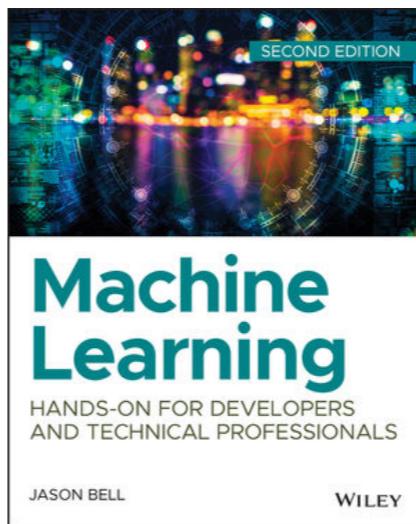
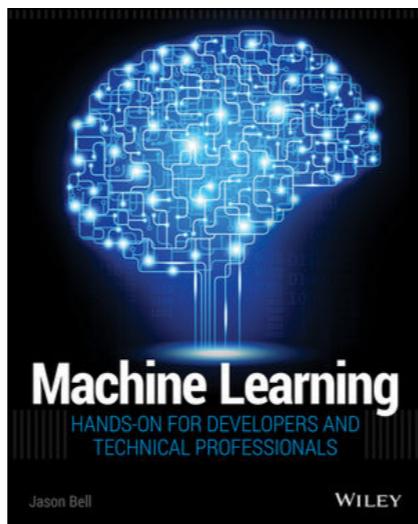
JASONBELL.ME



**DEDICATED TO
PROFESSOR PAUL MCKEVITT**



MY PHONE IS ON AIRPLANE MODE



ACT 0.01
NIDC 2022

OCTOBER 2022

LIVE MATCH BETWEEN TWO PLAYERS

PLAYER 1 RATING = 1200

PLAYER 2 RATING = 1000

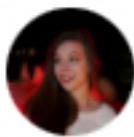
CALCULATE THE PROBABILITY OF THE PLAYERS WINNING:

P1 = $[1.0 / (1.0 + \text{POW}(10, [(1000-1200) / 400]))] = 0.76$

P2 = $[1.0 / (1.0 + \text{POW}(10, [(1200-1000) / 400]))] = 0.24$



“DATA IS BUSINESS, BUSINESS IS DATA” WHICH IS THE 2.0 VERSION OF LAST YEAR'S TALK. WORD OF WARNING, LAST YEAR I LEFT THE CONFERENCE HALL FEELING LIKE I CAN’T BUY A PACK OF MINI EGGS IN ASDA WITHOUT FEELING ROBBED VIA DATA HARVESTING. MAYBE IGNORANCE IS BLISS, JASON 😂



Lauren Taylor (She/Her)
Full Stack Engineer @ Komi
29m •

NOVEMBER 2022

73 YEARS IN THE MAKING
THEN IT ARRIVED

⚡ GPT-3.5

✖ GPT-4

ChatGPT **PLUS**

THE WOLRD'S MOST OVERRFITTED LANGUAGE MODEL

Suggest some codenames

for a project introducing flexible work arrangements

Make up a story

about Sharky, a tooth-brushing shark superhero

Plan an itinerary

for a fashion-focused exploration of Paris

Recommend a dish

to bring to a potluck

✉ Send a message



FEBRUARY 2023

ON LINKEDIN.....



Confid Oral Spray

★★★★★ 1 Reviews

Brings out the positive qualities of self-esteem helping you feel more confident and comfortable around others.

£10.95

Shipping calculated at checkout.

- ∅ Be your true authentic self
- ∅ Release guilt, shyness and low-self esteem
- ∅ Resolve negative beliefs we may hold about ourselves

-
- In stock, ready to ship

Add to cart

Buy with G Pay

**FIRSTLY, INVESTMENT IN WEB3/BLOCKCHAIN
TANKED BY 93%.**

**SECONDLY, OVERNIGHT THE WEB3 EXPERT
BECAME THE AI EXPERT**

**JASE, WE'RE GONNA DO
AI!**

AI



JASE, WE'RE GONNA DO....

- ▶ 1994 – PERL
- ▶ 1995 – DALE A TU CUERPO ALEGRIA MACARENA, HEY MACARENA
- ▶ 1996 – **JAVA APPLETS**
- ▶ 2000 – XML
- ▶ 2001 - RSS FEEDS
- ▶ 2002 – WEB 2.0
- ▶ 2004 – OPEN DATA **SOCIAL COLLABORATION**
- ▶ 2007 – IPHONE **APPS**, IT'S THE APP ECONOMY DUDE!
- ▶ 2009 – MOBILE WEB APPS
- ▶ 2012 – **HADOOP!** (WTF IS THAT JASE?)
- ▶ 2013 – **SPARK!** (SEE ABOVE)
- ▶ 2017 – REALTIME DATA STREAMING **KAFKA**
- ▶ 2018 – TENSORFLOW! FTW! IN **PYTHON!**
- ▶ 2019 - GO.
- ▶ 2020 - EVERYONE WORK FROM **HOME!!!!!!**
- ▶ 2021 - BORED **APES** EVERYWHERE!
- ▶ 2022 - PAY OUR EMPLOYEES WITH **BITCOIN!**
- ▶ 2023 - **CHATGPT**, AI IS WHERE IT'S AT JASE!!!! BACK TO THE OFFICE THOUGH!

**...AND I HAD A HUGE
WOBBLE**

ACT 1

THE GPT BUBBLE



hey there's a bubble

● web3
Search term

● blockchain
Search term

● chatgpt
Search term

+ Add comparison

Worldwide ▾

08/11/2016 - 11/09/2023 ▾

All categories ▾

Web Search ▾

Interest over time ?



● web3
Search term

● blockchain
Search term

● chatgpt
Search term

+ Add comparison

Worldwide ▾

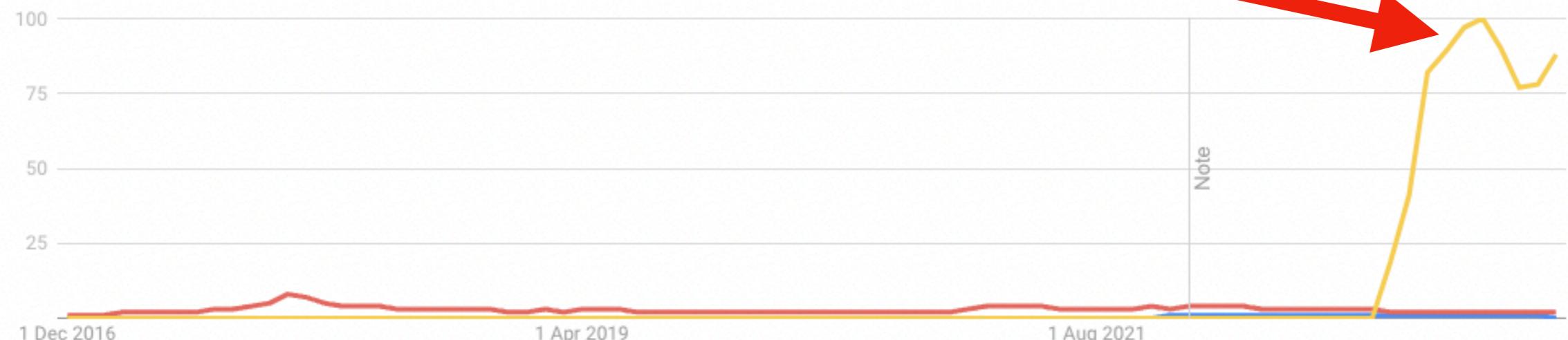
08/11/2016 - 11/09/2023 ▾

All categories ▾

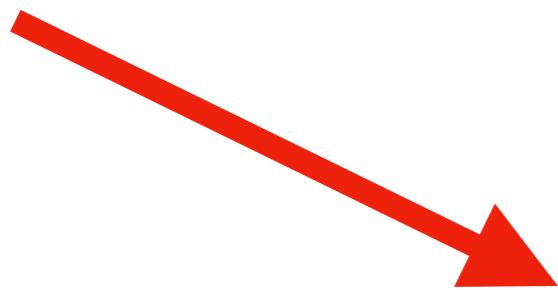
Web Search ▾

Interest over time ?

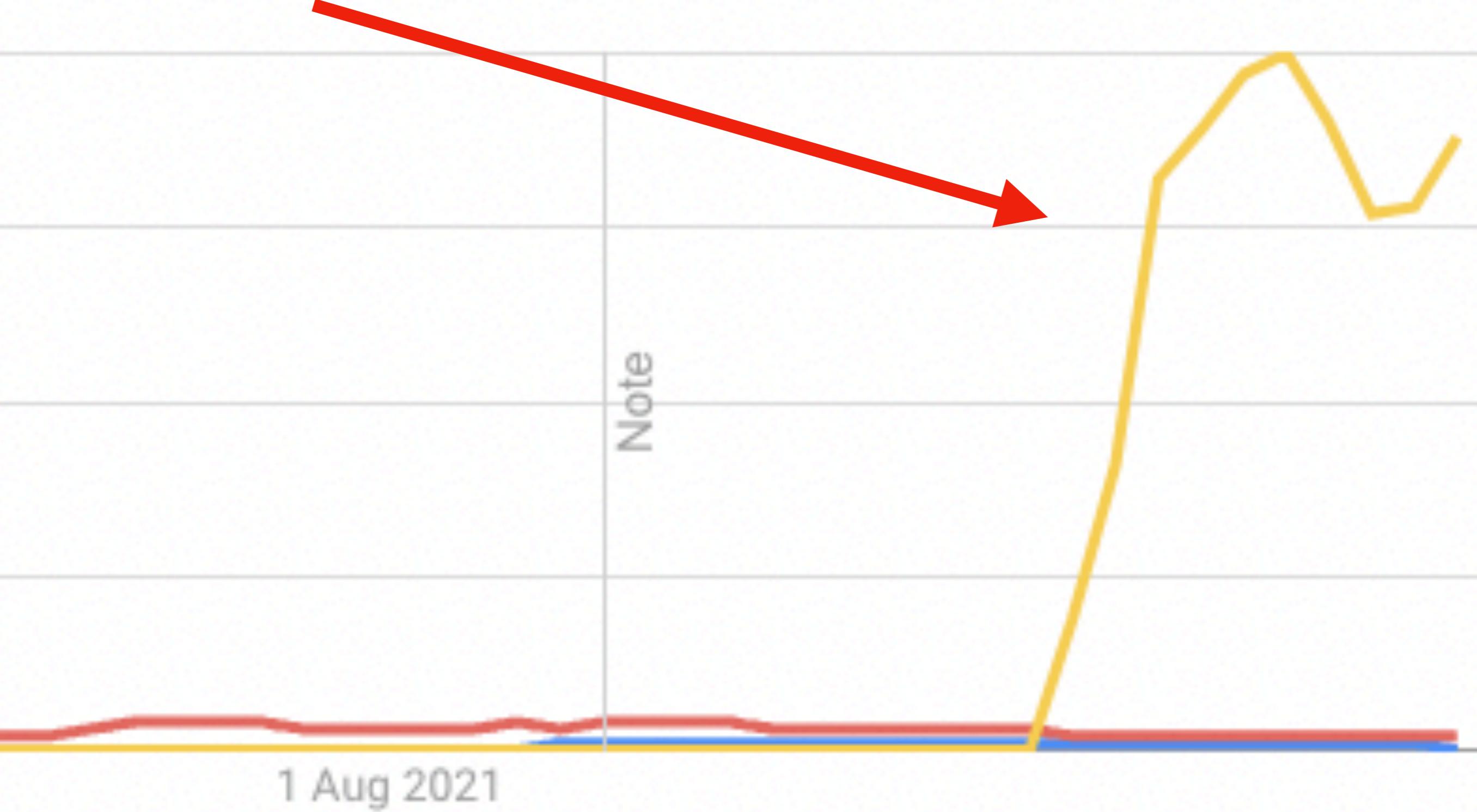
AKA: A BUBBLE



!BUBLE



BUBBLE!



● web3
Search term

● blockchain
Search term

● artificial intelligence
Search term

+ Add comparison

Worldwide ▾

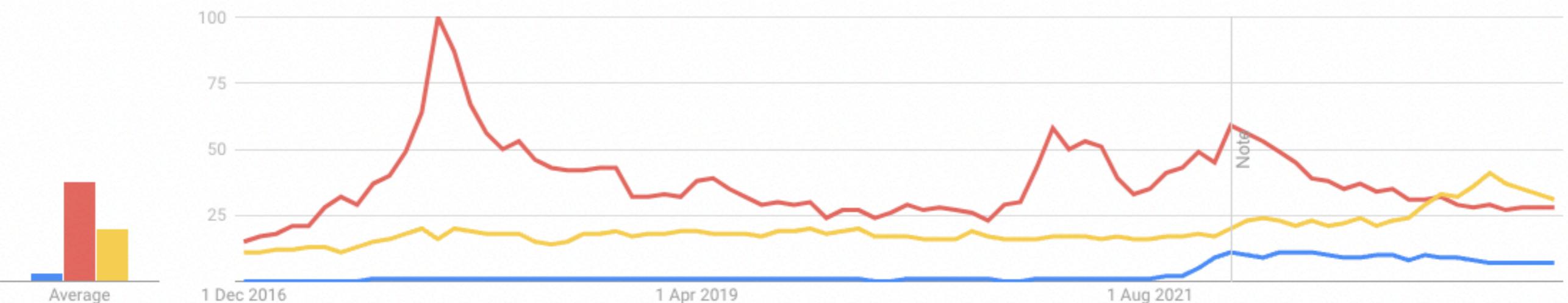
08/11/2016 - 11/09/2023 ▾

All categories ▾

Web Search ▾

A PREVIOUS BUBBLE

Interest over time ⓘ



● machine learning

Search term

● artificial intelligence

Search term

● generative ai

Search term

+ Add comparison

Worldwide ▾

2004 – present ▾

All categories ▾

Web Search ▾

Interest over time ?

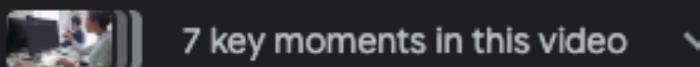


ChatGPT Guide: 10x Your Results with Better Prompts



Download the free E-Book here: <https://myaiadvantage.com/newsletter> In this video, I'm giving away over 100 free Chat GPT prompts to help ...

YouTube · The AI Advantage · 18 Jan 2023



www.greatai(prompts).com › ChatGPT

How to Write Better ChatGPT Prompts (5 Step Guide)



Step 1. Be Specific with Your Task · Step 2: State the Intent Clearly while Prompting ChatGPT · Step 3: Use Few-Shot Prompting Technique · Step...

Great Ai Prompts · NeetCode · 2 weeks ago

www.youtube.com › watch

The ULTIMATE Advanced Guide to Prompt Engineering with ...



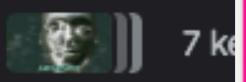
PROMPT ENGINEERING - "Prompt Templates" and also break down GPT advanced parameters

NEVERENDING CONTENT PROMPTS

20 engagement driving LinkedIn post ideas straight to your inbox
EVERY MONTH



£10/
month



www.youtube.com

Ultimate O...



24

FREE ChatGPT Prompts

Instructional Design
Edition



BingChat ~ Claude ~ Bard ~ Perplexity ~ Pi.ai ~ Semantic Scholar

Marketing Plan with AI • 24 pages



Jérémie
Grandillon



Tania
Saez



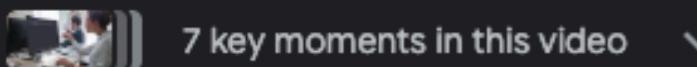
CREATE A
B2B MARKETING PLAN
IN 15 MINUTES WITH
CHAT GPT

ChatGPT Guide: 10x Your Results with Better Prompts



Download the free E-Book here: <https://myaiadvantage.com/newsletter> In this video, I'm giving away over 100 free Chat GPT prompts to help ...

YouTube · The AI Advantage · 18 Jan 2023



7 key moments in this video ▾

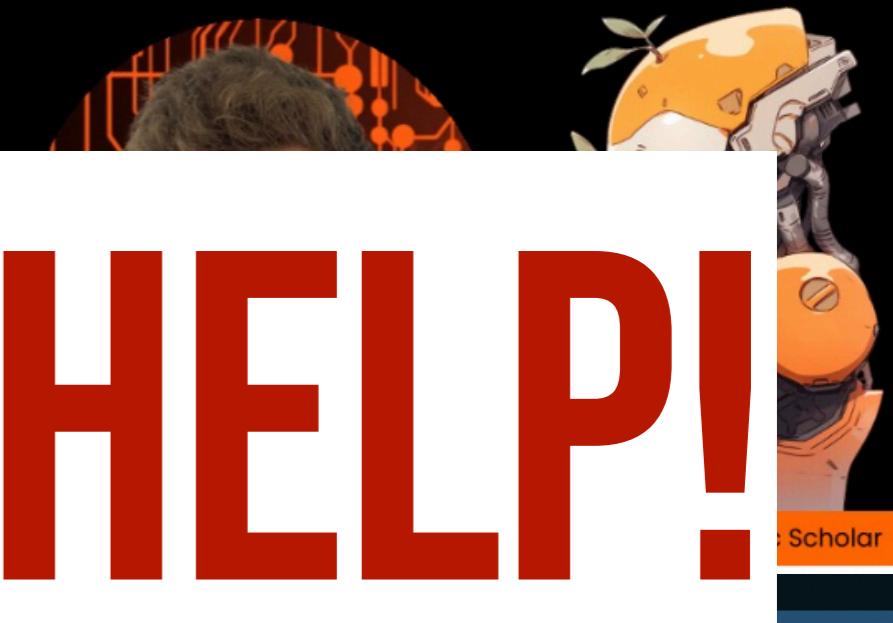
[www.greatai\(prompts.com](http://www.greatai(prompts.com) › ChatGPT

How to Write Better ChatGPT Prompts (5 Step Guide)

Step 1. Be Specific with Your Task · Step 2: State the Intent Clearly while

FREE ChatGPT Prompts

Instructional Design Edition



HELP! SEND HELP!

NEVERENDING CONTENT PROMPTS

20 engagement driving LinkedIn post ideas straight to your inbox
EVERY MONTH

£10/
month



Jérémie
Grandillon

Tania
Saez

CREATE A
B2B MARKETING PLAN
IN 15 MINUTES WITH
CHAT GPT



24

ACT 2

TRUTH, RISK AND UNCERTAINTY

LET'S START AT THE START....

**MY ISSUE IS NOT WITH CHATGPT OR LLM
DEPLOYMENTS**

**IT'S WITH OTHER PEOPLE WILL TO HAND
OVER RESPONSIBILITY TO THEM.**

IT'S HALLUCINATING!



NO, YOU JUST DON'T LIKE
THE PREDICTION



Wayne Denner he/him ✅ @waynedenner · 1h

...

Prep school appoints AI chatbot as its 'principal headteacher'



48





SO HOW ON EARTH DID WE GET HERE?

350 BC

ANCIENT AND CLASSICAL THOUGHT:
SOCRATES, PLATO AND ARISTOTLE

MIDDLE AGES:
RELIGIOUS THOUGHT AND DEVINE TRUTH

THE RENAISSANCE:
TRUTH APPROACHED INCREMENTALLY

ENLIGHTENMENT:
DAVID HUME ON CERTAINTY AND CAUSALITY AND EMPIRICAL SCIENCE

MODERN PROBABILITY AND STATS:
PASCAL AND FERMAT

POSTMODERNISM:
SOME THINGS MIGHT BE UNKNOWABLE

CURRENT DAY:
PREDICTION, COMPUTATION METHODS AND DATA



**EIKOS - PLAUSIBLE OR PROBABLE.
SOCRATES DESCRIBED EIKOS AS “LIKENESS OF TRUTH”**



EIKOS - PLAUSIBLE OR PROBABLE.
SOCRATES DESCRIBED EIKOS AS “LIKENESS OF TRUTH”

“LIKENESS OF TRUTH”, NOT ACTUAL TRUTH.

ARISTOTLE

**“....WHILE THEY SPEAK PLAUSIBLY, ...
DO NOT SPEAK WHAT IS TRUE.”**





BLAISE PASCAL AND PIERRE DE FERMAT LAID THE FOUNDATIONS TO PROBABILITY THEORY IN A SERIES OF LETTERS THAT THEY EXCHANGED DISCUSSING A FRIEND'S DESIRE TO INCREASE HIS ODDS AT PLAYING DICE.

**WE ARE HAPPY WITH PLAUSIBILITY THAN
TRUTH IN THE OUTPUT**

ANY LLM IS A PLAUSIBILITY ENGINE



New chat

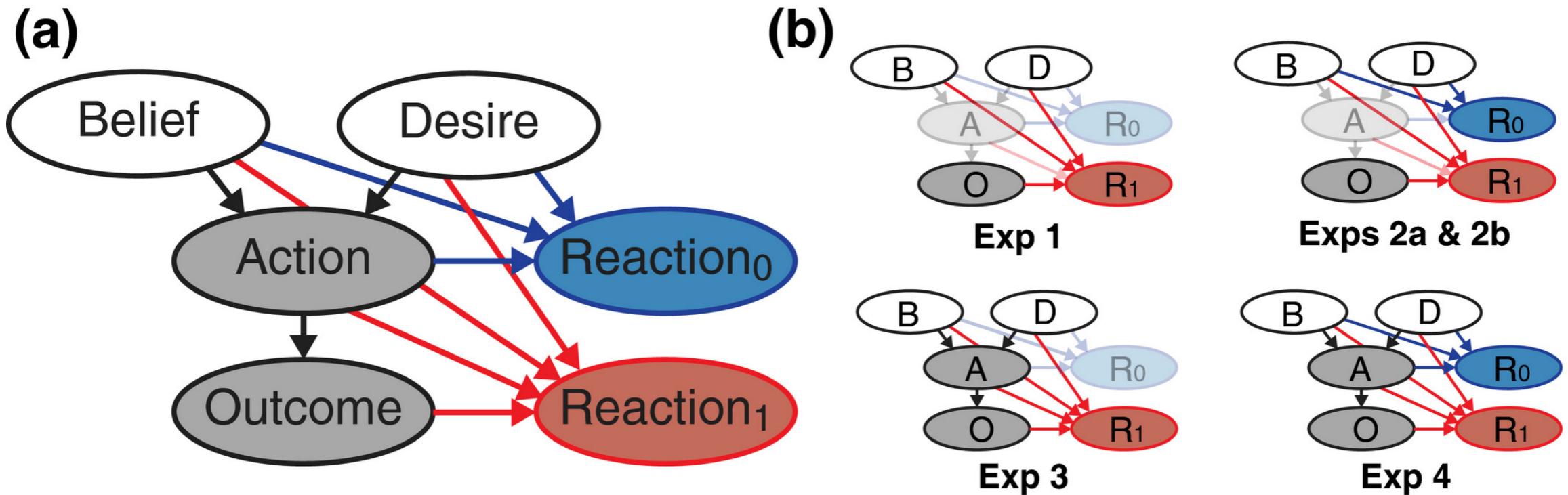


Send a message



ChatGPT may produce inaccurate information
about people, places, or facts. [ChatGPT](#)
[September 25 Version](#)

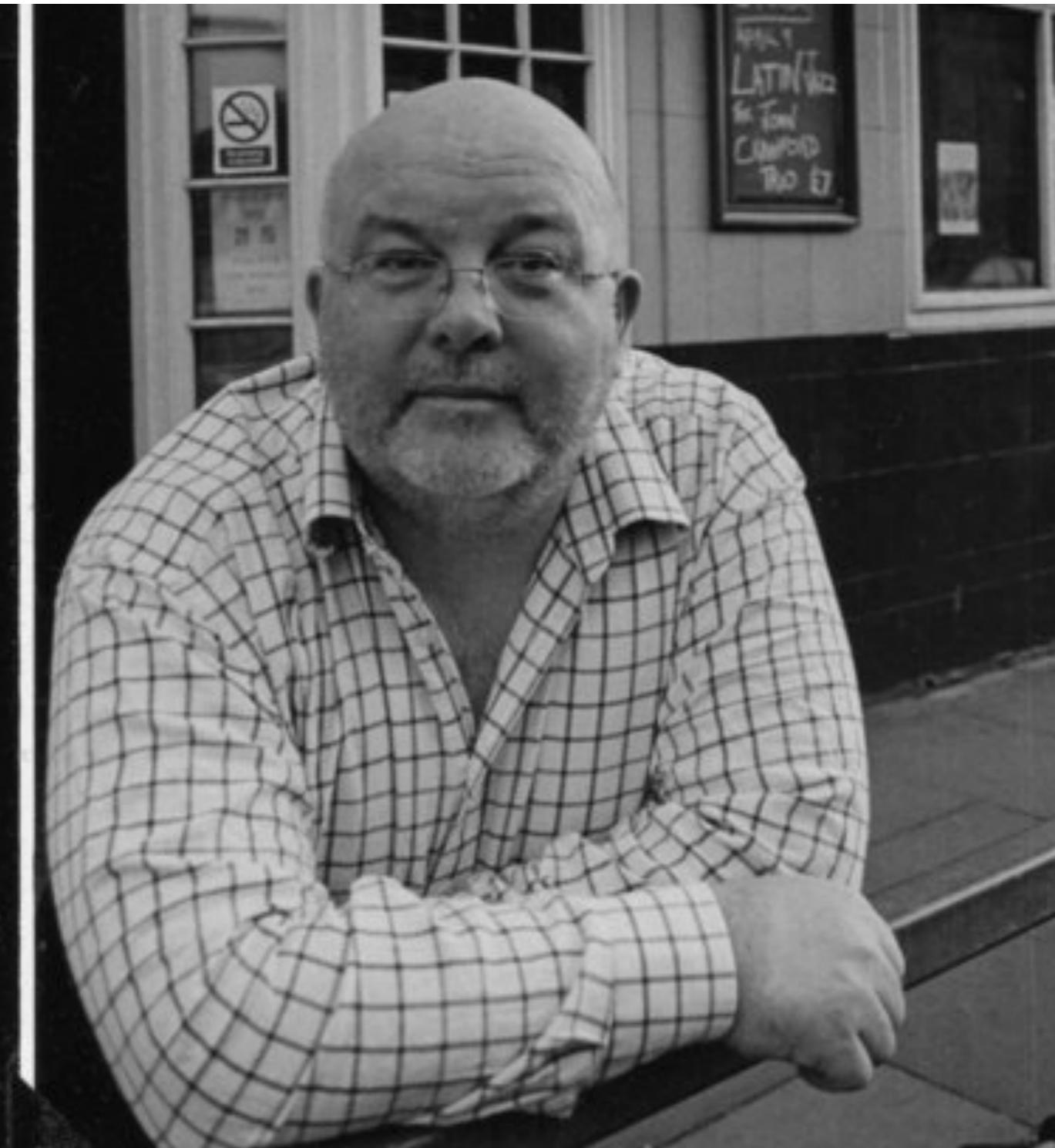
**IT'S YOUR EMOTIONAL EXPECTATION, NOT A
HALLUCINATION**



$$P(B, D | A, O, R_0, R_1) \propto P(R_1 | B, D, A, O) \times P(R_0 | B, D, A) \times P(A | B, D) \times P(B, D),$$

...PEOPLE DO NOT REACT TO OBSERVED EVENTS; THEY REACT TO A MENTAL REPRESENTATION OF THOSE EVENTS, A REPRESENTATION THAT IS AFFECTED JOINTLY BY THEIR BELIEFS AND DESIRES

RISK



IN 1991 THIS MAN TAUGHT
ME ABOUT **PREDICTION** WITH
DATA.

HOW DO YOU **PREDICT** THE
TOP TEN SINGLES TO ORDER
FOR NEXT WEEK? WITH ONLY
SOME BITS OF **CARDBOARD**
AND **NO COMPUTER?**

**WE HAVE SOME INFORMATION BUT ARE WILLING
TO PREDICT THE OUTCOME WITH WHAT WE KNOW**

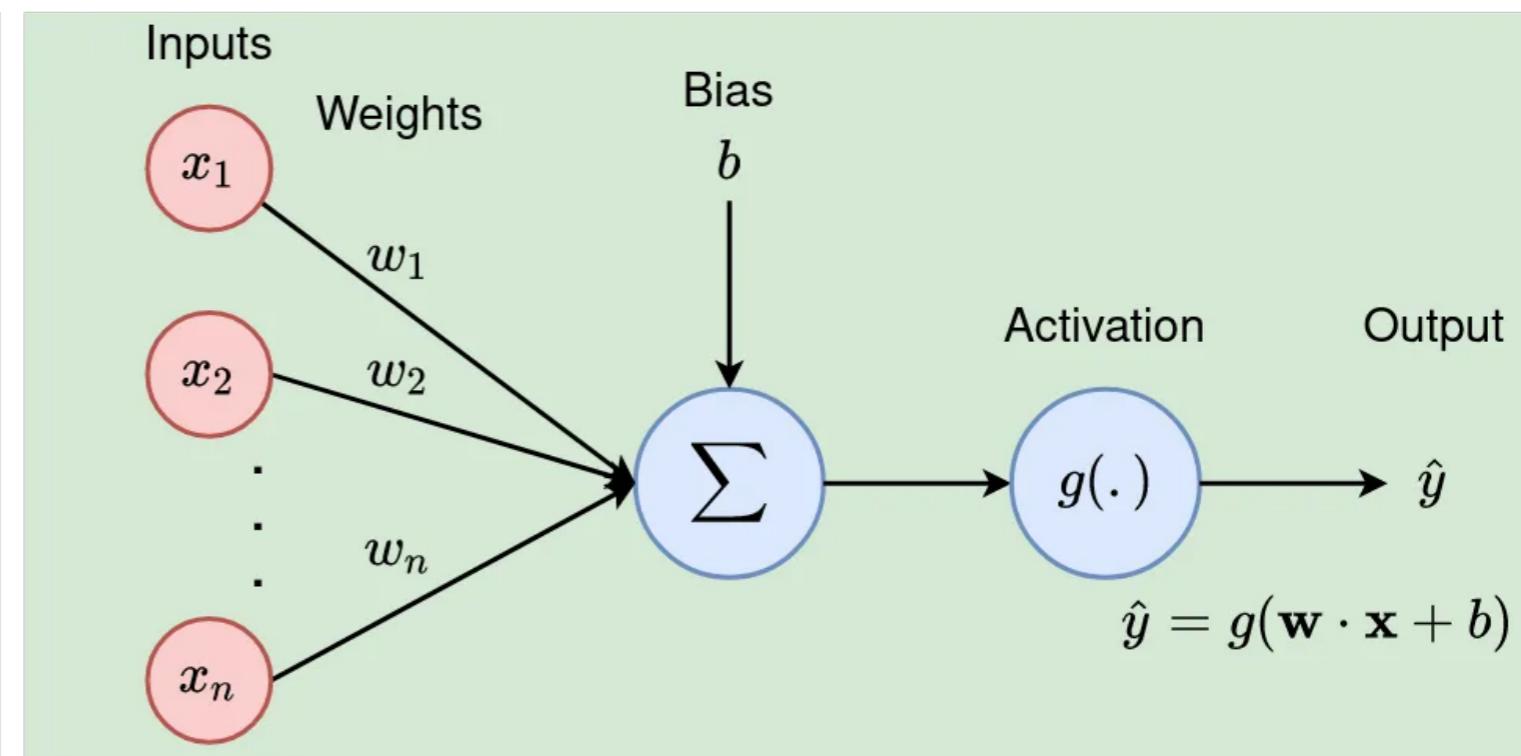
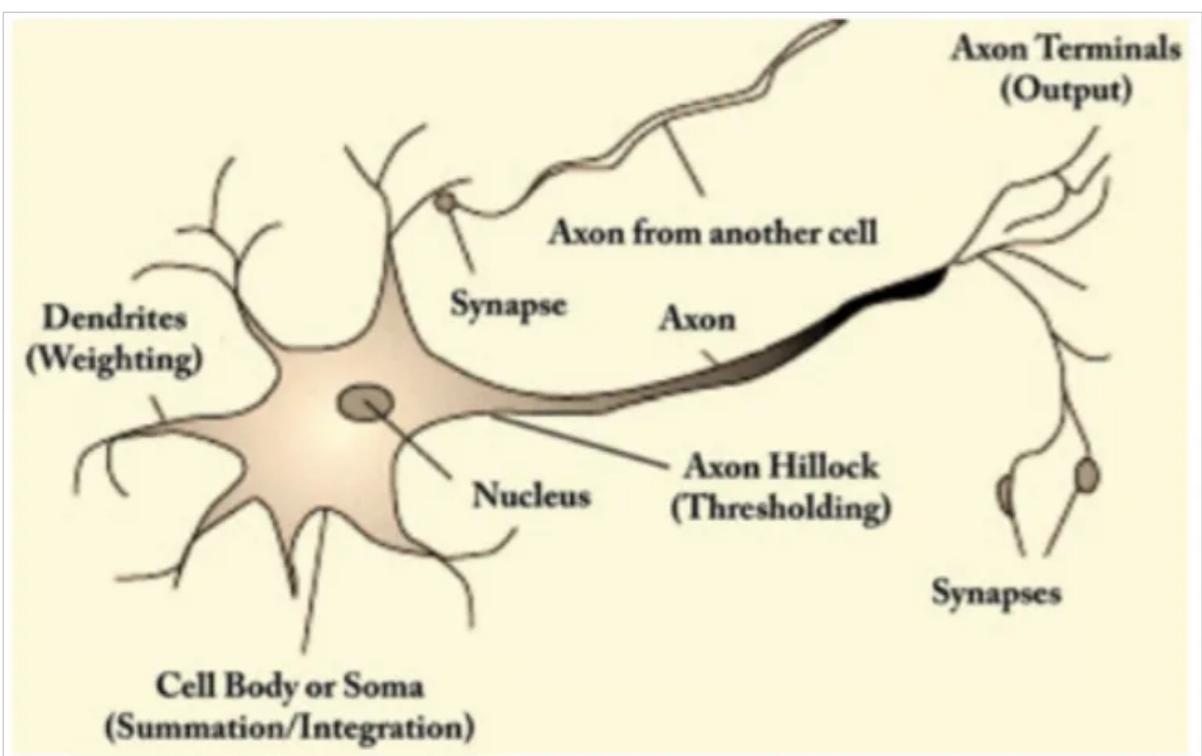
UNCERTAINTY

**WE HAVE NO INFORMATION, AND THEREFORE,
NO IDEA WHAT THE OUTCOME WOULD BE**

ACT 3

**WHAT'S IN THE (BLACK) BOX?
THE PATH TO AI REDEMPTION**





YOUR AI IS HUMAN

IF I CAME TO YOU ASKING TO BUILD A MODEL.....

I GIVE YOU THE EXPECTATION.....

**YOU HAVE A SENSE OF THE RIGHT MATH OR ALGORITHM TO
APPLY, THAT'S EXPERIENCE**

**WHAT IS MISSING FROM OUR FEELINGS ON LLM
MODELS IS WE HAVE NO IDEA HOW ANYTHING
REALLY WORKS**

**NO FAITH
NO TRUST
NO EXPERIENCE**

THE AI BOM

THE BILL OF MATERIALS

“Best finest surgeon – Strasberg
to cut me open, which I don’t
mind....

...they cut me open, and there is absolutely nothing there-
Strasberg is deeply disappointed but more even
academically amazed that he had made such a mistake.

He thought there was going to be so much more than he had ever
dreamed possible in almost anyone, but instead there was absolutely
nothing-”

- Marilyn Monroe

THE HARDWARE

THIS STUFF COSTS!

\$12/H * 24 * 365 = \$105,120

THE DATA

**REALISTICALLY WE DON'T CARE MUCH FOR
DATA PRIVACY, WHAT WE CARE FOR IS DATA
INTEGRITY**

HOW IS MY DATA USED AGAINST ME?

AAA

AA

A

BBB

BBB

B

B

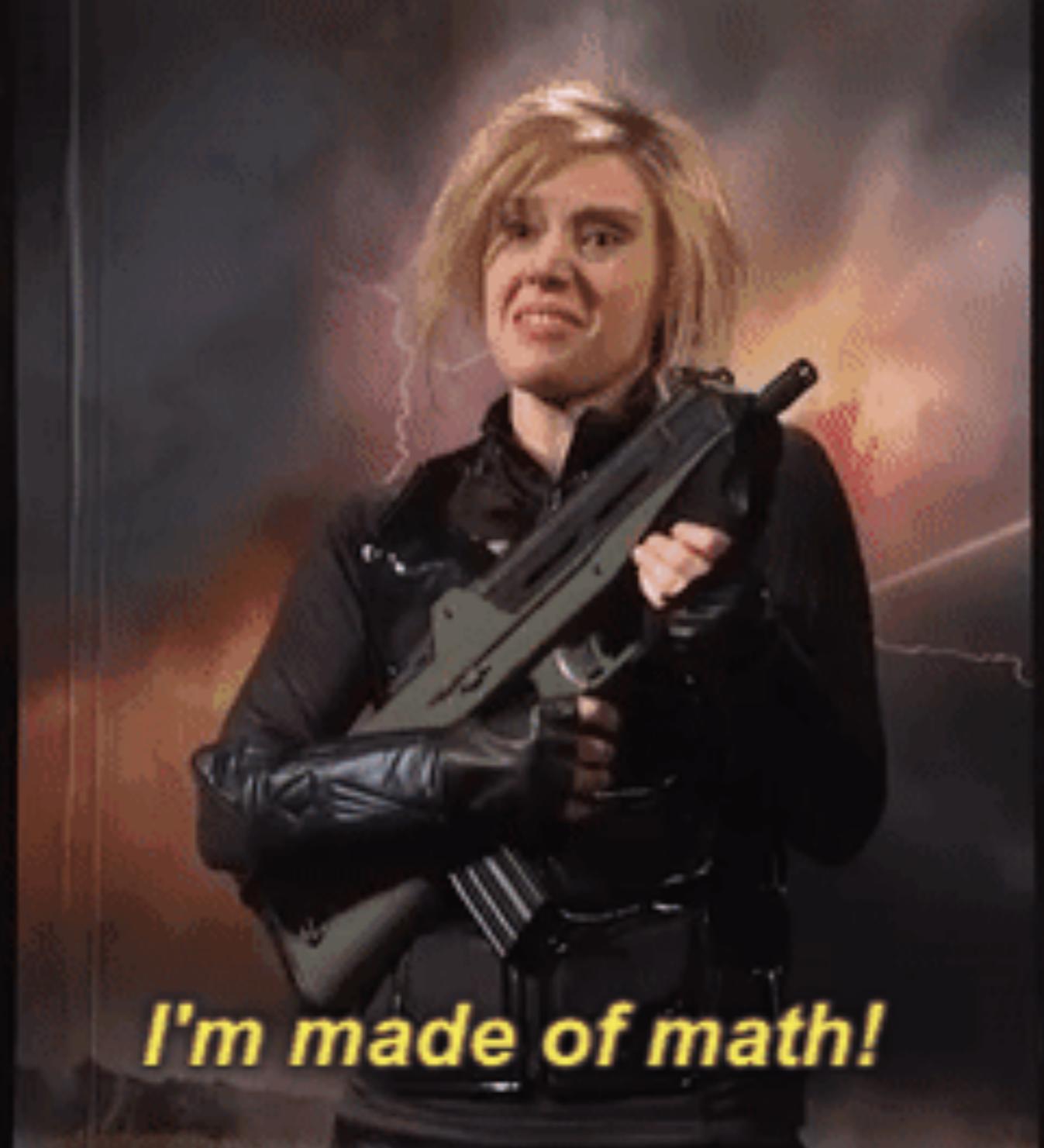
B

DATA CONSIDERATIONS

- IS IT CLEAN? DOES IT NEED CLEANING?
- DO WE REMOVE OR IMPUTE MISSING DATA?
- DO WE NEED TO CREATE OR AUGMENT WITH SYNTHETIC DATA?
- WHO ARE WE GOING TO COMPLETELY UPSET BY USING THIS DATA?

THE CODE

“I'M JUST CODE, ONES AND ZEROES...”



I'm made of math!



R

Call:

```
lm(formula = aa ~ x + y + z, data = data)
```

```
> model_summary <- summary(model)
> model_summary$r.squared
[1] 0.8783021
```

PYTHON

Output prediction for: aa

Model Accuracy: 87.6529121%

SOFTWARE CONSIDERATIONS

- **WHICH LANGUAGE SUITS? JAVA, PYTHON, C++, R.....**
- **FRAMEWORKS: CODING IT ALL, ALL RELYING ON A THIRD PARTY LIBRARY?**
- **ARE THERE ANY VARIANCES IN OUTPUT WHEN USING DIFFERENT VERSIONS OF THE SAME LIBRARY?**

Transparency: Providing clarity on the tools, hardware, data sources, and methodologies used in the development of AI systems.

Reproducibility: Offering enough information for researchers and developers to reproduce the models and results.

Accountability: Ensuring creators and users of AI systems are aware of their origins, components, and performance metrics.

Ethical and Responsible AI: Encouraging the documentation of training data sources, including any synthetic data used, to ensure there's knowledge about potential biases, limitations, or ethical considerations.

```

@ai_system{gpt4_openai,
  title = {GPT-4 by OpenAI},

  % Hardware Information
  hardware = {
    training = {TPU_v4, Nvidia_A100}, % List of hardware used for training
    inference = {TPU_v3, Nvidia_T4}   % List of hardware used for inference
  },

  % Programming Language Information
  languages = {
    training = {Python},
    inference = {Python, C++, Rust}
  },

  % Developers involved
  developers = {
    jasebell = {link = {https://github.com/jasebell}}
  }

  % Source Data
  data_source = {
    training = {WebText, Books, Journals, SyntheticDataset}, % List or description of datasets
    validation = {CustomTestSet},                                % Validation datasets
    synthetic_data = {                                         % Details on synthetic data
      used = true,
      type = {GANs, Simulations},
      percentage = 10,                                     % Percentage of synthetic data in the training dataset
      notes = {Synthetic data was generated using a combination of GANs and simulation tools.}
    },
    notes = {Some datasets were cleaned and processed before use.}
  },

  % Library Versions
  libraries = {
    tensorflow = {version = 2.5, link = {https://www.tensorflow.org/}},
    pytorch = {version = 1.8, link = {https://www.pytorch.org/}}
  },

  % Accuracy Information
  performance = {
    R2 = 0.95,      % R^2 Score
    F1 = 0.92,      % F1 Score
    MAE = 2.5,       % Mean Absolute Error, for example
    notes = {Performance on validation dataset.}
  },

  % Training History
  training_history = {
    {date = {2023-01-01}, notes = {Initial training.}},
    {date = {2023-06-01}, notes = {Retrained with additional data.}},
    {date = {2023-09-15}, notes = {Fine-tuning for specific task.}}
  },

  % Additional Notes
  note = {Training took approximately 2 months with a distributed setup.}
}

```

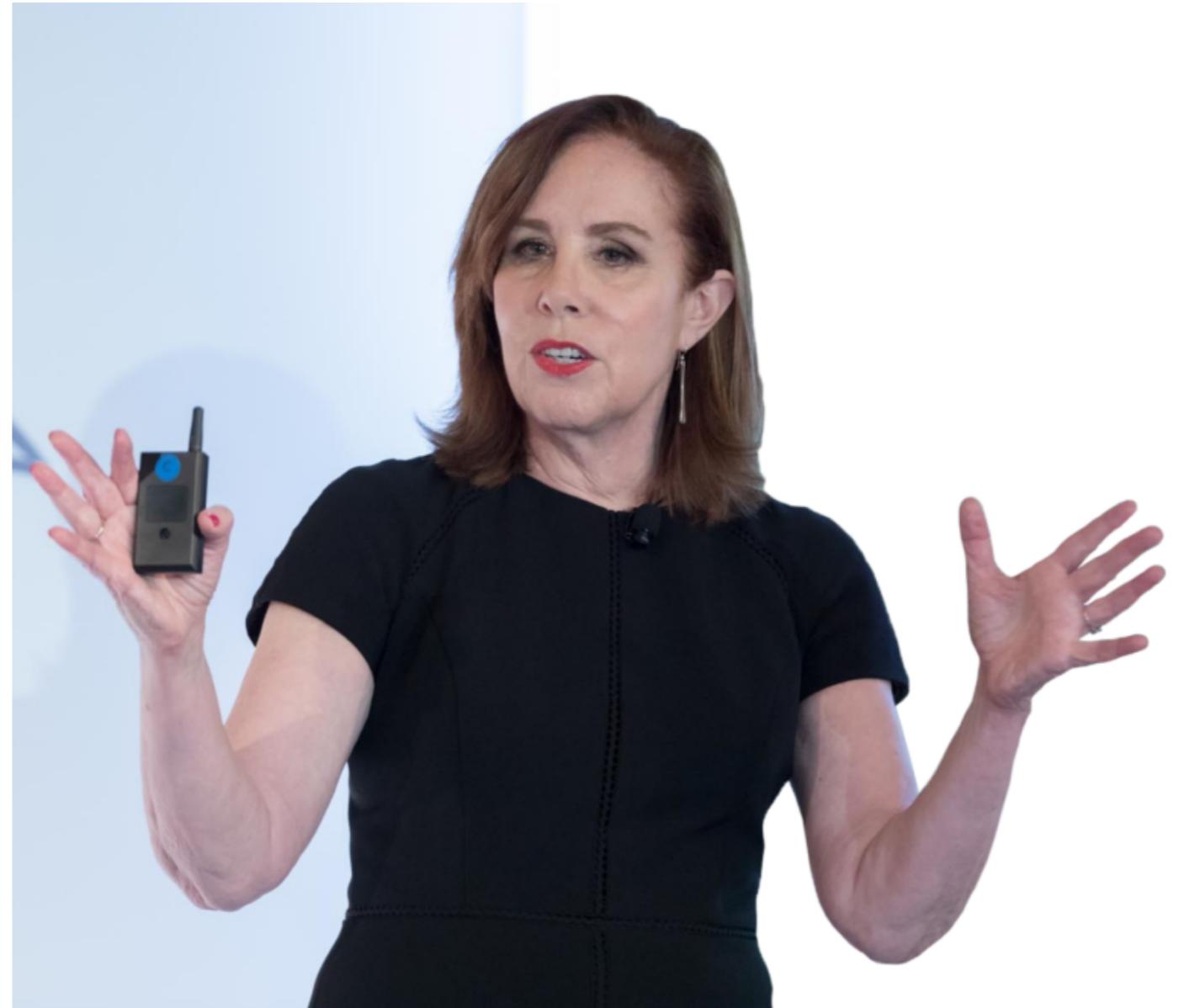
[HTTPS://GITHUB.COM/JASEBELL/AI-BILL-OF-MATERIALS](https://github.com/jasebell/AI-BILL-OF-MATERIALS)

[HTTPS://REDMONK.COM/JGOVERNOR/2023/10/18/INTRODUCING-THE-AI-BILL-OF-MATERIALS/](https://redmonk.com/jgovnor/2023/10/18/introducing-the-ai-bill-of-materials/)

ME?

**“WE PERCEIVE WHAT WE
BELIEVE. IN OTHER
WORDS, WE SEE WHAT
WE WANT TO SEE.**

**ESPECIALLY IN
OUR DATA.”**



**“IN OTHER WORDS,
IT’S NOT ABOUT
CONTROLLING YOUR
EMOTIONS- IT’S
ABOUT CONTROLLING
YOUR BEHAVIOUR.”**



“YOU NEED TO PUT AS MUCH (IF NOT MORE) WORK INTO ANALYSING **YOURSELF AS YOU DO YOUR **DATA.**”**



THANK YOU