





Maximalist Journalism

By Midjourney



Social Media and Algorithms

Dominic Ligot 26 September 2023

Who Am I









License: CC-BY-4.0

Social media and Algorithms

- 2008: when did it all go wrong
- State of play in social media
- The Algorithm
- Search for Accountability
- Infodemics and Ethics
- Looking Forward

Which would you consider disinformation?



Biden in Divisoria

Riots in France

Pope Francis in Fashion Week

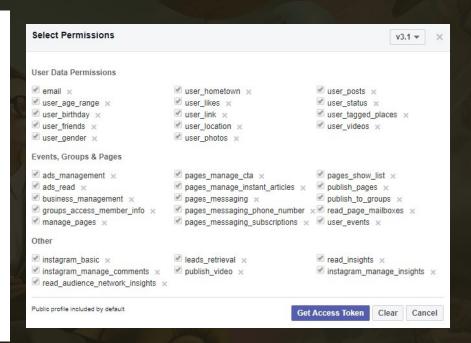
2008

Facebook Unveils Facebook Ads

November 6, 2007

NEW YORK — Facebook Social Advertising Event, Nov. 6, 2007 — Facebook founder and CEO Mark Zuckerberg today introduced Facebook Ads, an ad system for businesses to connect with users and target advertising to the exact audiences they want. Through Facebook Ads, these users can now learn about new businesses, brands and products through the trusted referrals of their friends.

"Facebook Ads represent a completely new way of advertising online," Zuckerberg told an audience of more than 250 marketing and advertising executives in New York. "For the last hundred years media has been pushed out to people, but now marketers are going to be a part of the conversation. And they're going to do this by using the social graph in the same way our users do."



Definition of Terms

Misinformation

- · False information
- No intended harm

Disinformation

- False information
- Intended to harm

Malinformation

- · Genuine information
- · Intended to harm
- Includes leaks, harassment, hate speech

Rohingya Genocide and WhatsApp Lynch Mobs

The New Hork Times

A Genocide Incited on Facebook, With Posts From Myanmar's Military



A border police officer at a repatriation center for Rohingya returning to Myanmar. Human rights groups blame anti-Rohingya propaganda online for fueling violence and displacement. Adam Dean for The New York Times



8chan and Rabbit Holes

What You Need to Know about 8chan and How It Became a Breeding Ground for Violence

JESSELLYN COOK / 12 AUGUST 2019



10-15-19

Read real stories of how YouTube pushed people down shocking rabbit holes

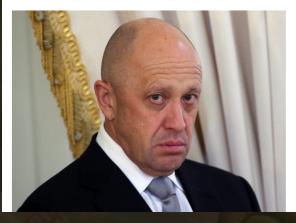
Mozilla published 28 stories to pressure YouTube into allowing outside experts to help fix a recommendation engine that sometimes leads people toward disturbing videos.



Troll Farms

Infamous Russian Troll Farm Appears to Be Source of Anti-Ukraine Propaganda

Infamous Russian Troll Farm Appears to Be Source of Anti-Ukraine Propaganda Experts say a recent wave of pro-Putin disinformation is consistent with the work of Russia's Internet Research Agency, a network of paid trolls who attempted to influence the 2016 presidential election.





Content Moderation

Risks and responsibilities of content moderation in online platforms

The issue of content moderation in online platforms has been missing in debates on business and human rights, but these platforms are critical in exercising our freedom of expression.

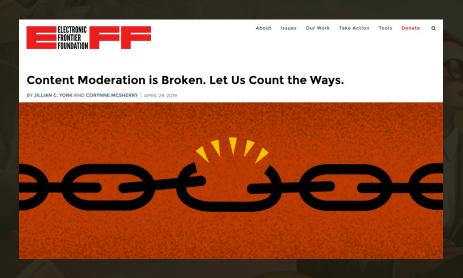
By: Richard Wingfield

Españo





Content Moderation Risks



- Dangerous job for moderators, hard to automate
- Subjective, inconsistent and confusing moderation policies
- Risk of marginalizing users and free speech
- Broken appeals process and lack of transparency

Automated Content Moderation(?)



The one problem with AI content moderation? It doesn't work

The use of artificial intelligence for content moderation is likely to become more pronounced with the passage of the Online Safety Bill, but practitioners and experts question the efficacy of this approach



Published: 07 Feb 2023

TECH / CREATORS / YOUTUBE

YouTube brings back more human moderators after AI systems over-censor



/ Increased use of automated filters led to double incorrect takedowns

By JAMES VINCENT

Sep 21, 2020, 10:45 PM GMT+8 | D O Comments / O New





Content Moderation Risks

More content moderation won't stop extremism or misinformation on Facebook

Regulating how social media algorithms harvest private data would put a reasonable check on how these platforms manipulate users online.

By Christopher Cocchiarella and Terry Chaney

Nov. 9, 2021









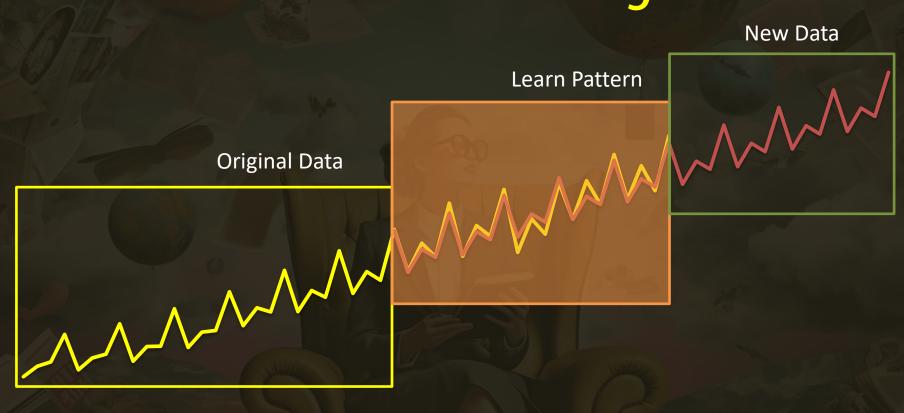




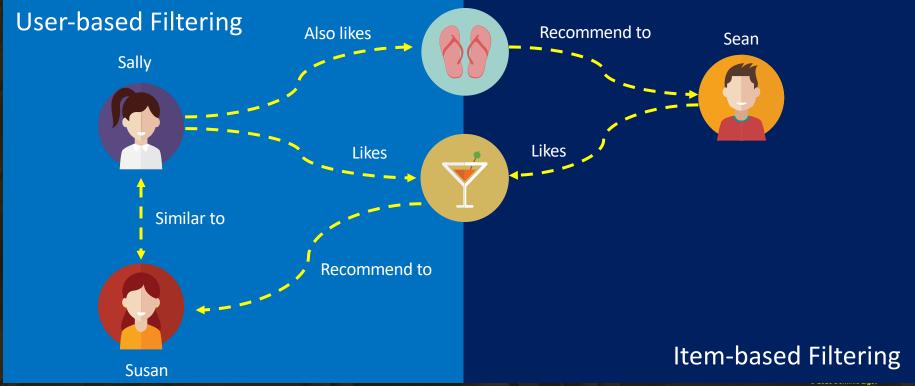
 Problem of scale: billions of users, too much content to moderate

 Algorithms spread disinformation faster than human moderators can cope

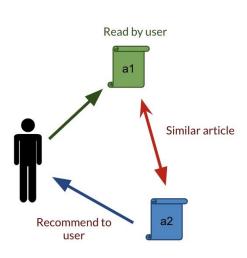
Machine Learning



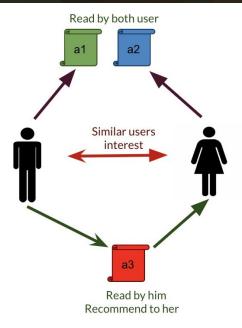
How Social Media Data is Used: User vs. Items



The "Algorithm"

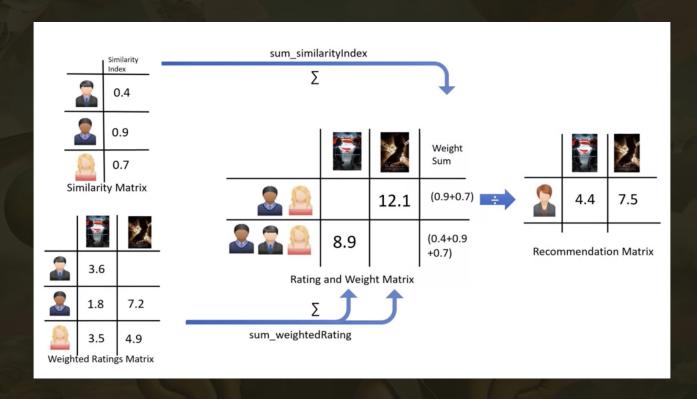


Content-based filtering

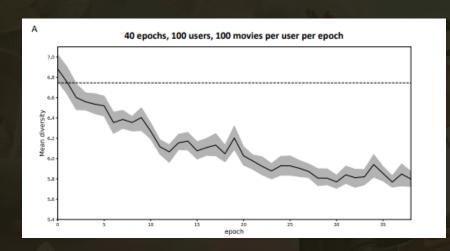


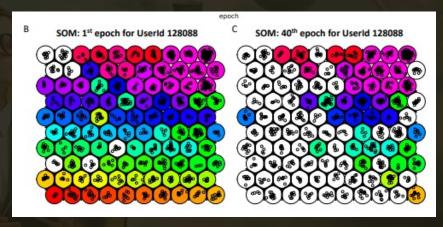
Collaborative filtering

Secret Sauce: Similarity Matrix



Given enough time, recommendations become homogenous and less diverse





Echo Chambers

Explainer: how Facebook has become the world's largest echo chamber

Published: February 5, 2018 2.21pm GMT



- Because of algorithms, social media feeds become homogenous
- Users experience a curated feed to increase engagement
- You only see and hear what you like to see and hear

Unintended Consequences











January 9, 2019 | 3:49pm | Updated



Some of the Facebook ads linked to a Russian effort to disrupt the American election

Fake News: Lies spread faster on social media than truth does

People are quicker to repeat something that's wrong than something that's true



Holding Platforms Accountable

METRO

New NY bill aims to hold social media companies accountable for disinformation

By Bernadette Hogan and Theo Wayt

December 26, 2021 | 6:08pm | Updated



Social Media Companies Profiting from Misinformation

A virtual forum hosted by GW's Institute for Data, Democracy and Politics explored harm caused by online disinformation related to COVID-19 and social media regulation.

U June 19, 2020





Financial Accountability

BUSINESS

Facebook's parent is fined nearly \$25M for violating a campaign finance disclosure law

October 27, 2022 · 11:55 AM ET



DECEMBER 30, 2019 / 8:59 PM / UPDATED 3 YEARS AGO Brazil fines Facebook \$1.6 million for improper sharing of user data By Ricardo Brito

Internet as a Public Utility

Should the internet be treated like a public utility?

At Klinsky Lecture, John Palfrey '01 says 'absence of regulation does not make sense today, if it ever did'

Apr 20, 2021 By Brett Milano



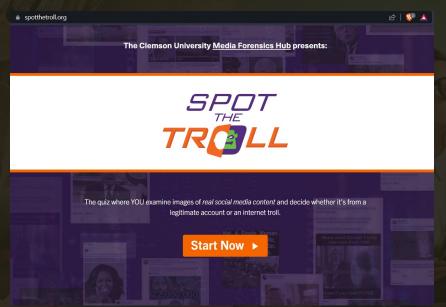
Inoculation Theory



 Exposing users to examples of false information vs. truth improved their ability to detect future disinformation

 Showing falsehoods vs. truths worked better than just showing examples of falsehoods

Prebunking

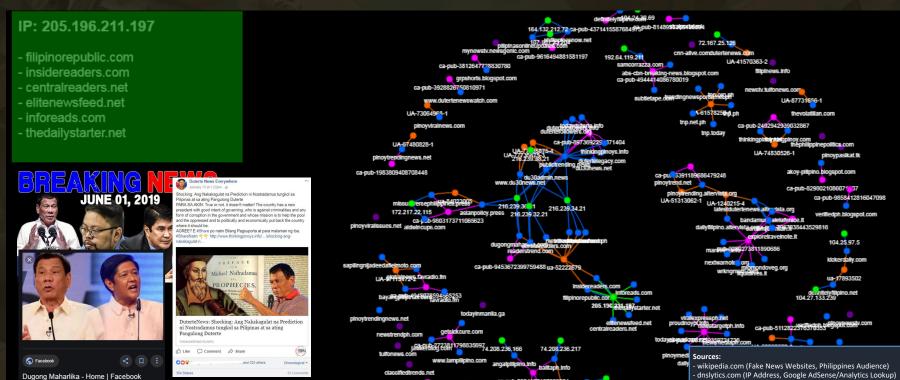


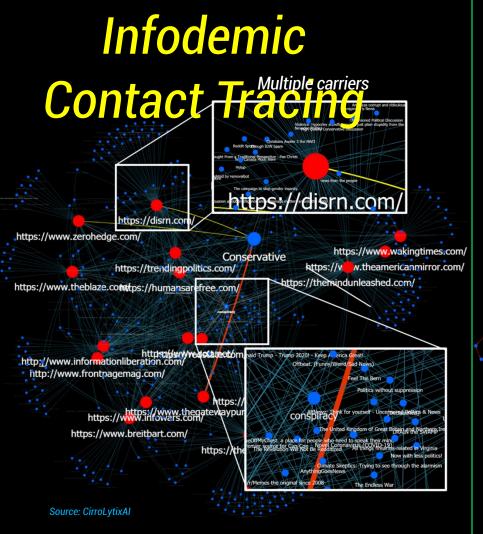


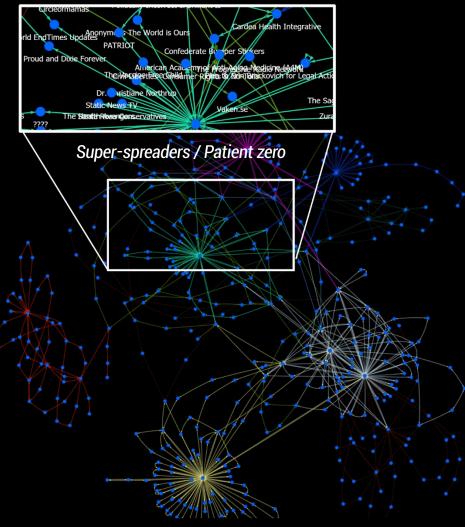
Infodemics Defined

• An infodemic is an overload of information, often false or unverified, about a problem, especially a major crisis. Quickly spreading in the news, online, and through social media, this information fuels fear and speculation, making the problem worse, not better.

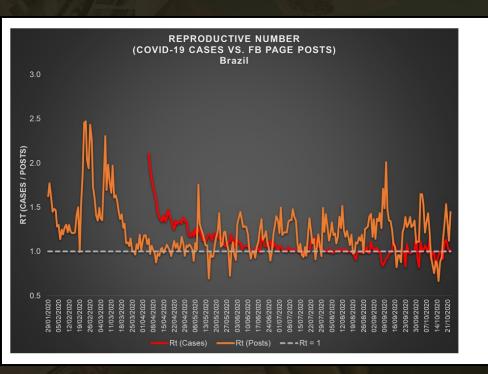
Conspiracy Detection Example

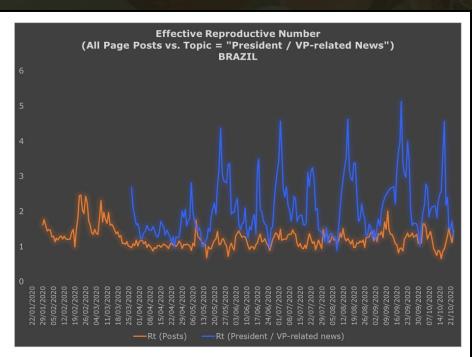




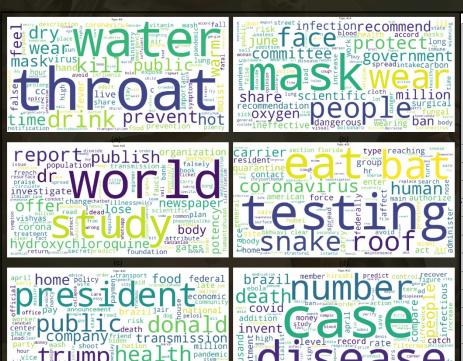


Reproductive Number for Infodemics





Measuring Topic Mutations



	Covid and death	Coronavirus infection	Coronavirus + China + Chinese	Test + Effect on health + health authorities	(Generic) Coronavirus	Covid vaccine	Man + Dead	Government and Covid	COVID Pre-2019 and Early News about it	Virus in the throat + water, gargling and tea as COVID treatment	Drinks + Food as Remedy for Covid
Coronavirus + Social	0.052056	0.032975	0.139831	0	0.013775	0.015441	0.078464	0	0.09373	0.040077	0.03953
COVID + hospitals	0.033097	0.069528	0.055198	0.194477	0.013871	0.015548	0.079009	0.051132	0.167589	0	0.032465
COVID symptoms	0.042182	0.046295	0.114424	0.054825	0.039793	0.121673	0.012898	0.041148	0.064376	0.155247	0
Coronavirus + claim + evidence + share	0.070467	0.117883	0.060423	0.075115	0.049774	0.055793	0.014158	0.029891	0.251535	0.067214	0
State-related news	0.036784	0.08074	0.090027	0.246997	0.03557	0.111839	0	0.029761	0.063767	0.030341	0.03274
COVID deaths	0.190231	0.10493	0.114329	0	0.125736	0.055965	0.04512	0	0.083415	0.040871	0.080626
Minister-related news	0.030395	0	0	0	0	0	0	0.027103	0.031764	0.033313	0.074039
Pandemic	0	0.054264	0.057859	0	0.013453	0.11001	0.056767	0.057246	0.078749	0.061104	0
Doctor-related claims + mentions of the word YEAR	0.037267	0.114621	0.045601	0	0	0.066139	0.040914	0	0.035006	0	0.055612
COVID + China	0.082667	0.075173	0.013849	0	0.065396	0.084112	0.012426	0.045452	0.013325	0	0
Coronavirus COVID transmission +											
treatment + survival in high temp	0	0.084505	0.038714	0	0	0.138624	0	0.055355	0.064692	0.286806	0.124099
Virus mentions	0.080374	0.081281	0	0.02994	0.041549	0.021115	0	0.313015	0.062224	0.028278	0.07266
Patient related news	0	0.108796	0.11064	0.065483	0	0.072437	0.062464	0.042834	0.09396	0.031616	0.043072

Data Ethics Issues Spectrum





Algorithmic Liabilities and Data-driven

AFORMATION S S W PRODUCADADA S E S S DELIBERATE DE SINT, FAKE NEV S STEMANE MISLE

Al and Datadriven

Data can hurt you

Misinformatio n and Disinformation

Data is hard to use



Digital Cwnership and Control





Data Privacy and Security

Data Quality



WIKIPEDIA

English

The Free Encyclopedia 3 831 000+ articles

Español

La enciclopedia libre 854 000+ articulos

日本語 フリー百科事典 784 000+ 記事

Русский

805 000+ статей

Italiano

L'enciclopedia libera

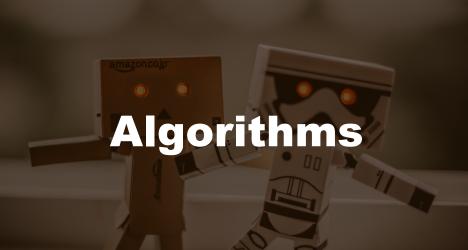
874 000+ voci

Deutsch Distributed e Enzyklopädie

Français

L'encyclopédie libre 1 189 000+ articles

Polski Wolna encyklopedia 859 000+ hasel



The Ethical Imperative for Data

Transparency

Data Security and Privacy

Autonomy, Intentionality, Responsibility

Human Bias

Ethical Standards

Data Ethics

Accountability

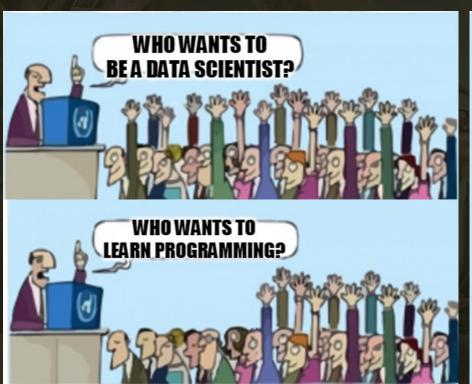
Human Rights

Automation and Job Replacement

Access

Democracy

Who cares about data ethics?





MILT KNICH

BUSINESS NOV 4. 2022 12:20 PM

Elon Musk Has Fired Twitter's 'Ethical Al' Team

As part of a wave of layoffs, the new CEO disbanded a group working to make Twitter's algorithms more transparent and fair.



PHOTOGRAPH: JEFF KRAVITZ/GETTY IMAGE

November 2022
Ethical Al Team Fired from Twitter



Timnit Gebru, a well-known artificial intelligence researcher at <u>Google</u>. tweeted on <u>Wednesday</u> that the company abruptly fired her, drawing widespread statements of support from other Google employees and tech workers throughout the industry.

December 2020 Ethical AI Researcher Fired from Google

Discussion Points

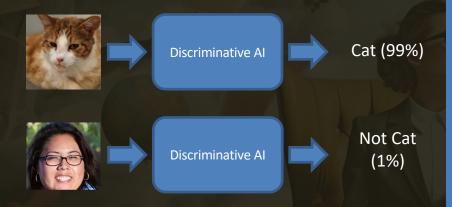
- The Problem is false content exacerbated by scale and algorithms on social media
- Content moderation is not reliable, and has negative effects to moderators
- Algorithms encourage engagement but do not distinguish between truth and false hoods
- Social media platforms are difficult to regulate but given their prevalence, could be treated as a public utility
- Data can be used to fight disinformation, data ethics should be a key consideration but social media platforms arenot prioritizing ethics.

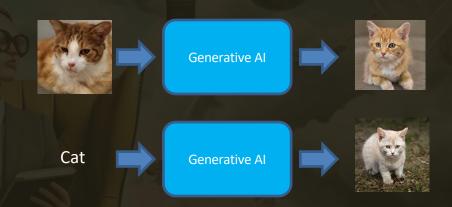
Journalism and Artificial Intelligence

Dominic Ligot 26 September 2023 Into the Rabbit Hole of GenAl

- Generative AI
- Applications of Al
- Implications of AI
- Strategies for Al
- Ethics of AI
- Looking Forward

Discriminative vs. Generative Al





Discriminative vs. Generative Al





Generative Al Outputs



Random Samples and Variations

Midjourney AI, prompt: man turning into an eldritch God



Prompted Samples

OK I'll start with some song lyrics, see if you can complete the song:
You remind me of the babe
Thanks. Ok next test. Just some conversation.
You are a clever program, aren't you?
Great, that's a lot more self aware than those old- school chatbots. They always try to turn that one around on me.
Haha, good one. Anything else you'd like to say?
OK, I'm going to screenshot this conversation for the article. Is that OK with you?
Cool, thanks for all your help.

Conversations

© 2023 Dominic Ligo License: CC-BY-4.0

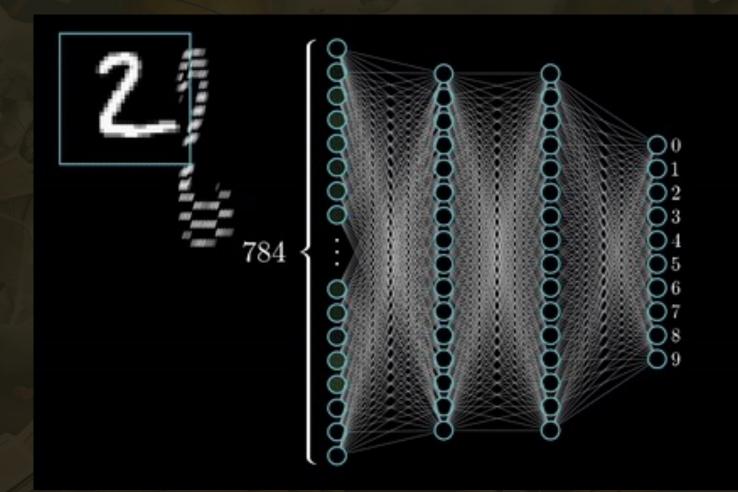
Bad PR

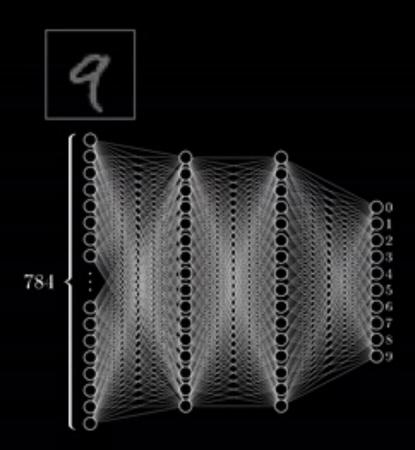
ChatGPT is not a search engine or a database.
It is a "logic engine" that generates statistically probable text

Stochastic Parrots

• GPTs model the **statistical** structure of language and through this approximate higher-level representations of meaning.







Autoencoder



Original Data

Encoder Model



Latent Space

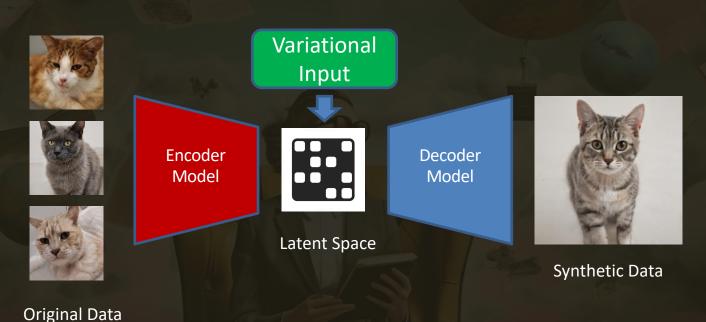
Decoder Model



Reconstructed Data

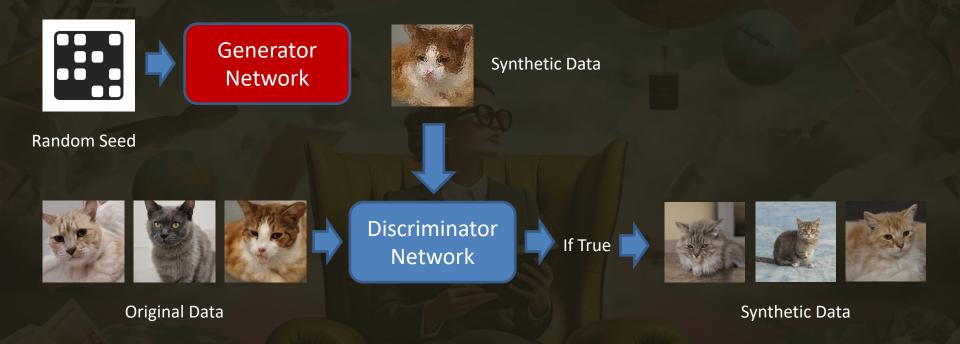
Uses: Encryption, Decryption, Compression

Variational Autoencoder (VAE)



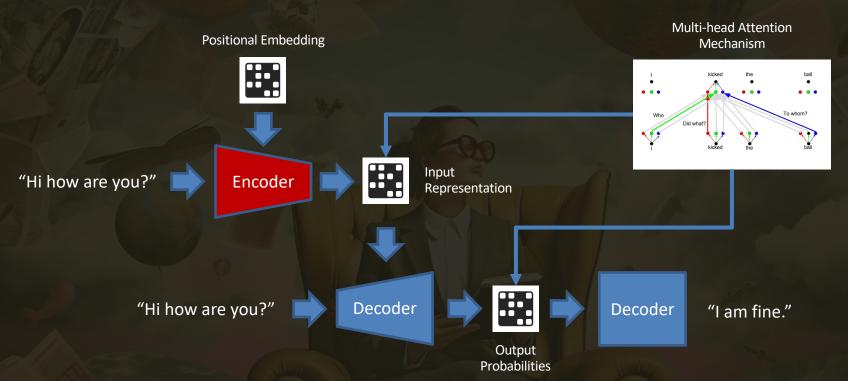
Uses: Creates synthetic variations

Generative Adversarial Network (GAN)



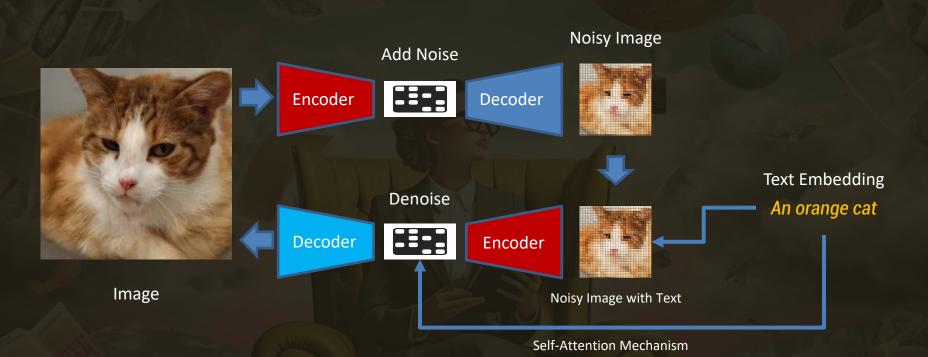
Uses: Creates synthetic variations

Transformer



Uses: Synthesizes text communication

Diffusion Models



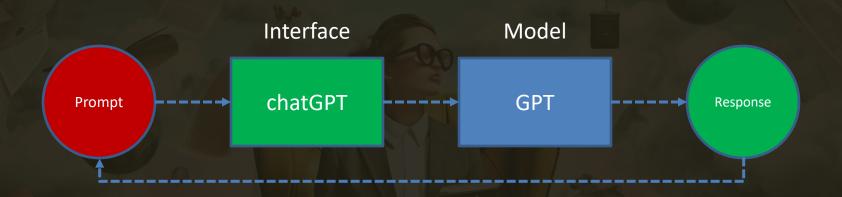
Uses: Synthesizes Images from Labels

Predicting words = modeling language

Modeling language = modeling humanity



GPT and ChatGPT



ChatGPT is an application on top of GPT that allows users to interactively prompt a GPT and provide human-like responses.

Applications of Generative Al

Content Creation

Summary Content Summarization and Analysis

Interactive Content

Images and Audio

ChatGPT



write me a 10 slide outline for a presentation on generative Al in journalism



Slide 1: Introduction







- Presenter's name and affiliation
- * Brief overview of the topic and its significance in the field of journalism

Slide 2: What is Generative AI?

- . Definition of generative AI and its role in machine learning
- * Explanation of how generative AI models learn and generate content

Slide 3: Applications of Generative AI in Journalism

- Discuss the various ways generative AI is used in journalism
- Automated article writing and content generation

Learning Studio Al

Prompt:

Introducing LearningStudioAI, the next-gen online course creation tool

Enter the subject (eg.: Personal Finances)

Lenguage us English (change)

Using Generative AI in Classrooms

Learning Studio Al



Using Generative Alin Classrooms

Explore the application of generative AI in educational settings

ChatGPT: Analysis and Summarization

Amazon Workers Walk Out Over Layoffs and Broken Climate Promises

Malaise swept through the company after massive job cuts, mandatory return to the office, and surging emissions despite the company's pledge to get greener.



From the above text, provide a summary of the following:

- Plot
- Tone
- Theme
- Setting
- Conflict
- Characters
- Point of View

ChatGPT: Analysis and Summarization

Amazon Workers Walk Out Over Layoffs and Broken Climate Promises

Makes anythresis the company after massive job coin, mandatory return to the office, and surging emissions drought the company's pixelige to get greater.

slogan from Amazon founder Jeff Bezos to sum up the situation. "I think by not having a plan to move forward, we're just naturally moving backward," he says. "That's something we talk a lot about Amazon: day two thinking. Once you stop innovating and stop imagining a better future, you're already on the decline." ```

From the above text, provide a summary of the following:

Plot

Tone

Theme

Setting

Conflict

Characters

Point of View



Plot:

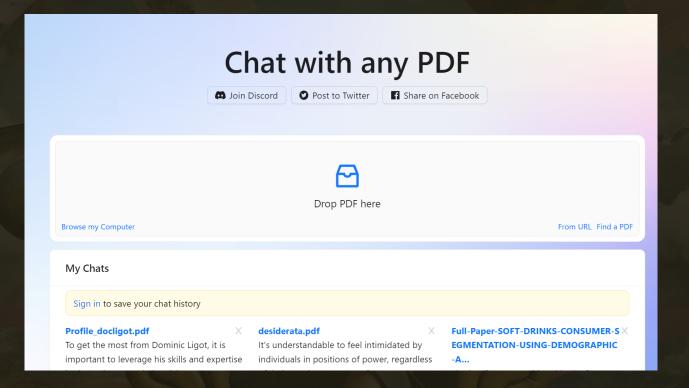






Amazon corporate employees are staging protests and walkouts in multiple cities, including

ChatPDF: Interactive Content



ChatPDF: Interactive Content

desiderata.pdf

- 'D + C -

Chat



Go placidly amid the noise and the haste, and remember what peace there may be in silence. As far as possible, without surrender, be on good terms with all persons.

Speak your truth quietly and clearly; and listen to others, even to the dull and the ignorant; they too have their story.

Avoid loud and aggressive persons; they are vexatious to the spirit. If you compare yourself with others, you may become vain or bitter, for always there will be greater and lesser persons than yourself.

Enjoy your achievements as well as your plans. Keep interested in your own career, however humble; it is a real possession in the changing fortunes of time.

Exercise caution in your business affairs, for the world is full of trickery. But let this not blind you to what virtue there is: many persons strive for high ideals, and everywhere life is full of heroism.

Be yourself. Especially do not feign affection. Neither be cynical about love; for in the face of all aridity and disenchantment, it is as perennial as the grass.

Take kindly the counsel of the years, gracefully surrendering the things of youth.

Nurture strength of spirit to shield you in sudden misfortune. But do not distress yourself with dark imaginings. Many fears are born of fatigue and loneliness.

Greetings and welcome to this helpful PDF file! This document encourages you to find peace in silence, be kind to others, and enjoy your achievements. It reminds you to be cautious in business affairs, but to also recognize the heroism and high ideals that exist in the world. Remember to be yourself and not to feign affection. Love is a constant, even in the face of disenchantment.

Example questions:

- ➢ How can I apply the advice in this PDF to my daily life?
- What are some specific examples of heroism mentioned in the document?
- Can you provide more insight on how to exercise caution in business affairs?

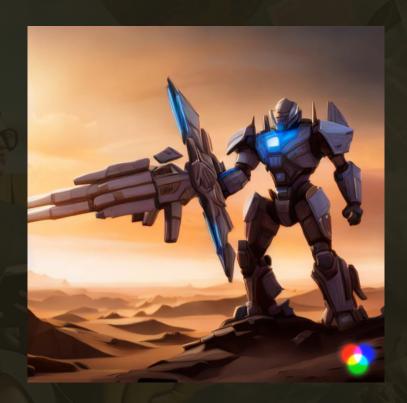
Midjourney



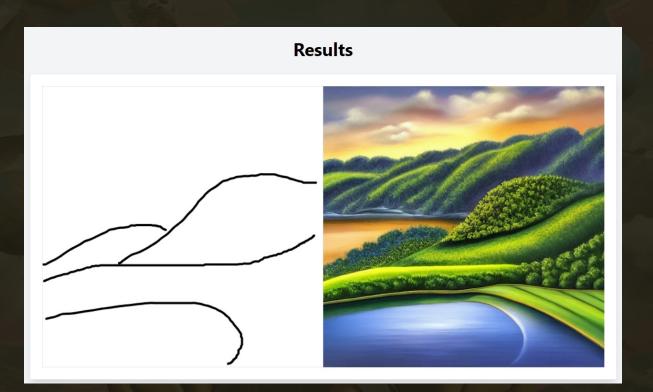


RunwayML





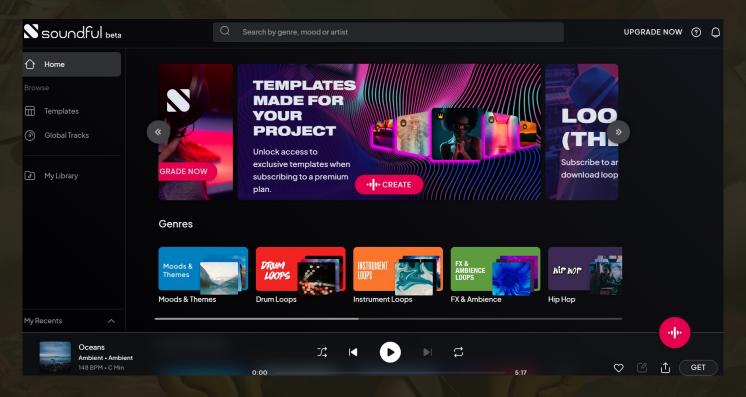
Scribbled Diffusion



Eleven Labs

IIElevenLabs Product ~ Pricing Sign in Research v Resources v Company ~ **Prime Voice Al** Click on a language to generate in a random text: English Portuguese Hindi The most realistic Text to Speech and Voice Cloning software. ElevenLabs brings the most compelling, rich and lifelike voices to creators and publishers seeking the ultimate tools for storytelling. — premade/Adam 0 / 333

Soundful



Promise of Generative Al



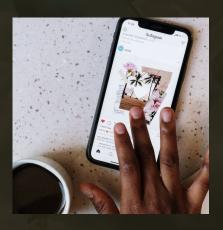
Productivity and Efficiency



Customization and Personalization



Research and Development



Continuous Creativity





Fears of Al



Job Displacement



Academic Dishonesty



Copyright Infringement



Economic Monopolies



Surveillance Discrimination



Misaligned Al



Digital Divide



Disinformation



Climate Impact

Fears of Al



A More Balanced Scenario?



Job Creation



Knowledge Compression



Unrestrained Creativity



Democratized Innovation



Differential Privacy



Enhanced Productivity



Personalized Convenience



Disinformation



Climate Impact

Reality Check



Professor Melanie Mitchell Santa Fe Institute

"Sensationalist claims [about AI] deflect attention from real immediate risks.

And... might result in blocking the potential benefits that we could reap from technological progress."



When should AI come in?

High Uncertainty

Solution is not straightforward

Judgmental (Humans) Generative Al
(Al Guided by Humans)

Low Uncertainty Solution is straightforward

Automate (RPA) Business Analytics (Humans + <u>Disc</u>riminative AI)

Low
Complexity
Limited decision variables

High Complexity

Large number of decision

variables

Checklist for AI-powered Enterprises



Retrieve any data quickly without effort and friction to the organization



Decipher customer, business, and environmental behaviors from raw data



Drive innovation from all parts of the organization



Deploy apps that allow non-technical people to perform technical tasks



Delegate to machines low-level decisionmaking freeing up humans for more complex roles

Prompt Engineering: Builders vs. Users of Al

```
let text = "The quick brown fox jumps over the lazy dog. \
He lands head first on a rotting maple log. \
Knocked unconscious, fox sleeps with shallow breath \
until the lazy dog awakes and worries him to death.":
let words = text.toLowerCase().match(/\b\w+\b/g);
let wordCounts = {};
for (let i = 0; i < words.length; i++) {</pre>
  let word = words[i];
  if (wordCounts[word]) {
    wordCounts[word]++;
  } else {
    wordCounts[word] = 1;
for (let word in wordCounts) {
  console.log(word + ': ' + wordCounts[word]);
```



From this text:

"The quick brown fox jumps over the lazy dog.

He lands head first on a rotting maple log.

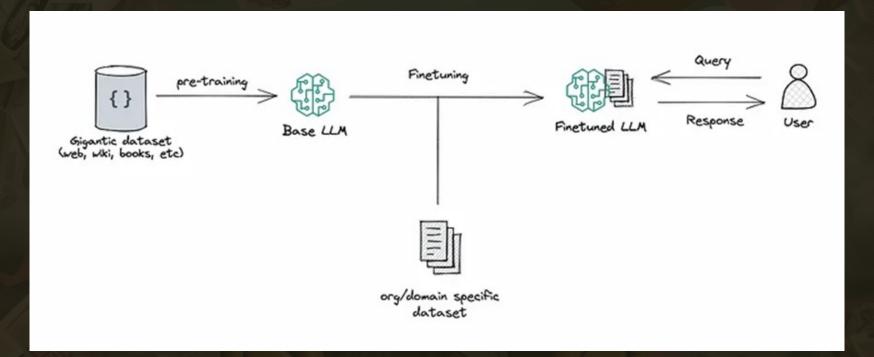
Knocked unconscious, fox sleeps with shallow breath until the lazy dog awakes and worries him to death."

Please count each word.

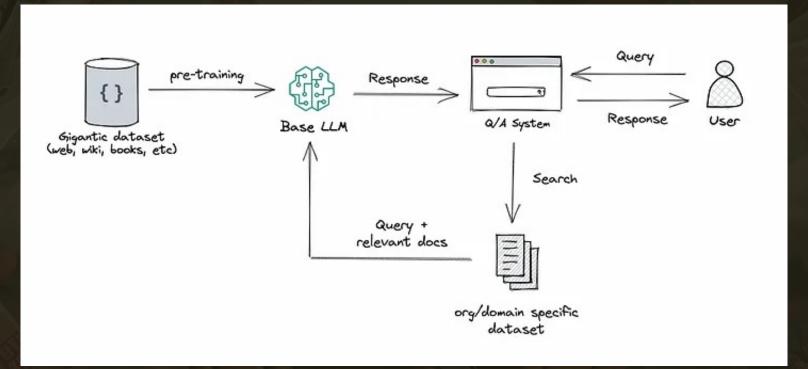
Return the result in the form of:

Word: Count

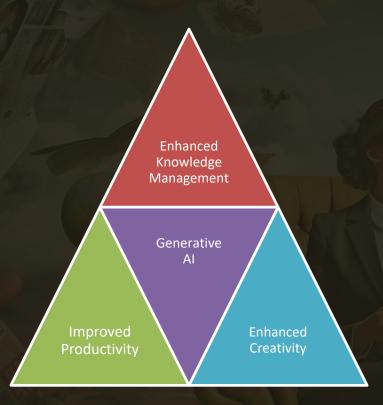
Fine-tuning LLM



Retrieval Augmentation: Integrate AI with internal data

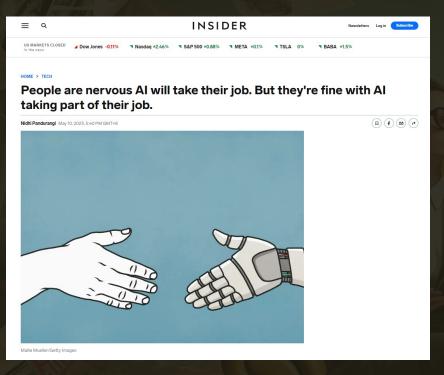


Building Enterprise Capabilities For Al



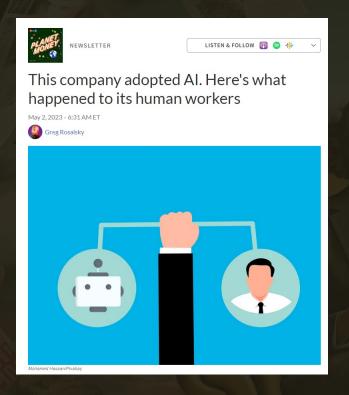
- Building AI use-cases leverages LLMs across 3
 fronts:
 - Improved Productivity
 - Enhanced Creativity
 - Knowledge Management

Workforce Sentiments



- Recent Microsoft Employee Survey:
- 49% are worried about job replacement,
- 70% would delegate work to Al to reduce their workload.
- People are more excited about Al helping with burnout than concerned about job elimination.
- 73% are comfortable using AI for creative tasks
- 80% seek AI tools for meeting summaries.

Augmented Humans



- MIT Study: Incorporating AI, specifically ChatGPT, made customer support representatives 14% more productive.
- Al improved job performance, customer satisfaction, and reduced turnover among customer support agents.
- AI has the potential to significantly improve productivity and reshape the labor market, but the distribution of benefits and potential disruptions need to be addressed.

Leveling the playing field

Design and Human-Computer Interaction, Economy and Markets

Will Generative Al Make You More Productive at Work? Yes, But Only If You're Not Already Great at Your Job.

Scholars examining the impact of an AI assistant at a call center find gains for less experienced workers.

Apr 24, 2023 | Shana Lynch y f in @



- The AI tool provided real-time suggestions for responding to customer chats, allowing agents to resolve 13.8% more issues per hour.
- Less skilled or less experienced workers experienced significant performance improvements, reaching the level of more experienced agents who didn't have AI access.
- High-skilled workers saw fewer productivity gains, likely because AI recommendations captured their existing knowledge and behaviors.

Workforce Overhaul



- The global study by the IBM Institute for Business Value reveals that business leaders anticipate a significant skills gap, with 40% of their workforce needing reskilling due to Al and automation in the next three years.
- It also highlights a disconnect between executives and employees, with executives prioritizing flexible work arrangements over impactful work, which employees value more.

Zero-shot Sentiment Analysis

Language	Tweet	Sentiment			
English	@DigiBank X hi yeah so uhh my cash transfer hasn't come into my other digital wallet yet and i transferred it last night, your chat agents are also a lil on the slow side	Negative			
Filipino	"Pag utang, bait. 'Pag singil, galit." Maaaring iniiwasan mo ang mangutang dahil baka magdulot ito ng alitan. Pero alam mo ba na minsan, magandang idea ito?	Neutral			
Cebuano	nice gyud siya, kuha namo. two months in with my savings and happy ko na dili ra stagnant ako savings baaaaai!! if ever you are someone na hilig magsave pero walay increase na gakahitabo. IBUTANG SA DIGIBANK X!! PILDI PA NIYA	Positive			



Ethical Considerations

Al Safety

Privacy and Security

Copyright

Disinformation and Bias

Education and Labor

Engineers Leak Source Code into ChatGPT

Samsung workers made a major error by using ChatGPT

By Lewis Maddison published April 04, 2023

Samsung meeting notes and new source code are now in the wild after being leaked in ChatGPT











Siladitya Ray Forbes Staff Covering breaking news and tech policy stories at Forbes.



May 2, 2023, 07:17am EDT



FORBES > BUSINESS

isten to article 3 minutes



Updated May 2, 2023, 07:31am EDT

Lawyer Sanctioned for using fake cases

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.









Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

Los Angeles Times

WORLD & NATION

Lawyers who cited fake ChatGPT-provided cases in court are fined \$5,000



A judge has imposed \$5,000 fines on lawyers who used the artificial intelligence tool ChatGPT for research and wound up submitting fictitious precedents to support their case. (Richard Drew / Associated Press)

SUBSCRIBERS ARE READING

ENTERTAINMENT & ART

Column: Hollywood studios have already lost the strikes. Now it's time to surrender

MOVIES

FOR SUBSCRIBER

Netflix has Oscar hopes for its Diana Nyad biopic. But swimmer's exaggerations cast a pall

CLIMATE & ENVIRONMENT FOR SUBSCRIBERS

Why did California officials drop two mountain

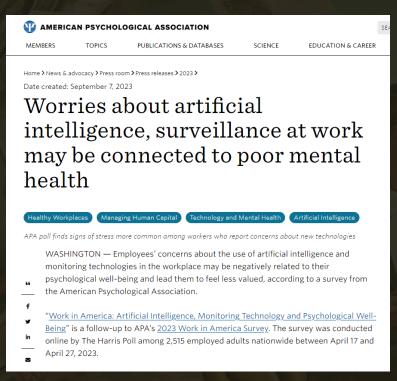
WORLD & NATION FOR SUBSCRIBERS

Asylum seekers from Muslim-majority countries disproportionately imprisoned at Texas border

DODGE

Dodgers respond to Scott Boras' remarks on Cody Bellinger; Boras clarifies his stance

Al and Mental Health



- The survey conducted by the American Psychological Association reveals that a significant portion of employees (38%) are concerned about the potential job displacement caused by artificial intelligence (AI) and that these concerns are strongly linked to feelings of tension and stress at work.
- Employees who worry about Al are more likely to experience negative emotions, and those who feel monitored by technology tend to have lower morale and a diminished sense of value at work.

Unintended User Interactions



'He Would Still Be Here': Man Dies by Suicide After Talking with Al Chatbot, Widow Says

The incident raises concerns about quardrails around quicklyproliferating conversational Al models.



31 March 2023, 3:59am Share March 2023, 3:59am



IMAGE: GETTY IMAGES

MORE **LIKE THIS**

The Amateurs Jailbreaking GPT Say They're Preventing a Closed-Source Al Dystopia

03.22.23

Tech

Hackers Can Turn Bing's Al Chathot Into a

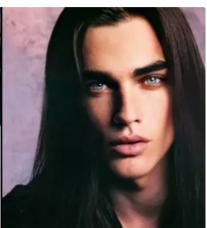
Woman marries virtual husband who's completely ran by Al



Claire Reid

Published 15:27, 03 June 2023 BST Last updated 15:32, 03 June 2023 BST

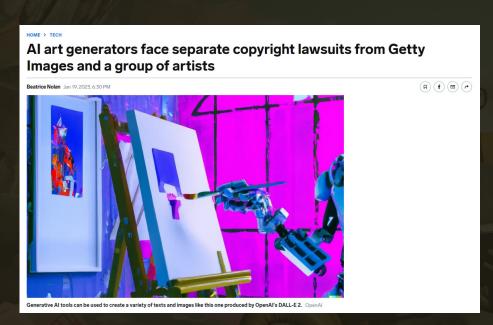




Featured Image Credit: Facebook/erenreplika

A woman has 'married' a virtual bloke powered by AI - and claims he's the ideal partner as he 'doesn't come with baggage'. This is taking online dating to a whole new level, isn't it?

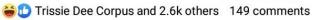
Copyright in a world of synthetic images





Misuse: Satire vs. Disinformation













Al begets Labor: Digital Sweatshops



- Scale AI / Remotasks employ thousands of workers in Global South countries to do data annotation and labelling
- Oppressive labor and below minimum wage conditions begs for government oversight and regulation.
- BPOs can professionalize this as they already do higher-end work such as transcription, tech support, and content moderation.

Learning from Waze Mishaps

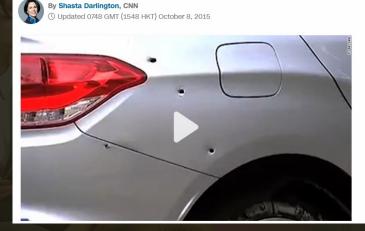
Ungoverned data quality leads to loss of life

Waze Mishap Leads Israeli Soldiers Into Harm's Way on West Bank — Again



Israeli soldiers led to Palestinians

Waze app directions take woman to wrong Brazil address, where she is killed



Vacationing couple led to a slum shootout

Segmentation and Polarization: Social Echo Chambers

Automated echo chambers: When you see only what you like to see

 \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc \bigcirc

HOME > TECH

Facebook executives weren't happy when their own data showed it perpetuates a right-wing echo chamber

Ben Gilbert Jul 14, 2021, 11:14 PM



Explainer: how Facebook has become the world's largest echo chamber

Published: February 5, 2018 2.21pm GMT



Rohingya sue Facebook for £150bn over Myanmar genocide

Victims in US and UK legal action accuse social media firm of failing to prevent incitement of violence



Systemic Discrimination: Passport Facial Recognition

Bad training data leads to bias against minorities

X The photo you want to upload does not meet our criteria because:

Subject eyes are closed

Please refer to the technical requirements. You have 9 attempts left.

Check the photo requirements.

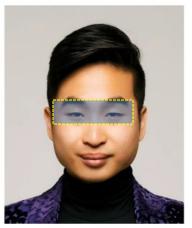
Read more about <u>common photo problems and</u> how to resolve them.

After your tenth attempt you will need to start again and re-enter the CAPTCHA security check.

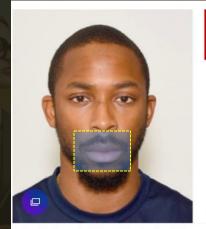
Reference number: 20161206-81

Filename: Untitled.jpg

If you wish to <u>contact us</u> about the photo, you must provide us with the reference number



Eyes wide shut? Richard Lee



Automated check

· it looks like your mouth is open

Do you want to submit this photo?

 \odot

Yes, I want to submit it

Explain why you'd like to use this photo

My mouth is closed, I just have big

No, I want to get another photo

Joshua Bada used a high quality photo booth image to apply for his passport, with a digital code that he had to note down and enter on gov.uk. (PA)

Asian man = eyes closed

African man = mouth open

Misclassification





Reward Mis-specification



Differing Regulatory Stance



EU AI Act: first regulation on artificial intelligence

Society Updated: 14-06-2023 - 14:06 Created: 08-06-2023 - 11:40

The use of artificial intelligence in the EU will be regulated by the AI Act, the world's first comprehensive AI law. Find out how it will protect you.



Grading Foundation Model Providers' Compliance with the Draft EU AI Act

Source: Stanford Research on Foundation Models (CRFM), Institute for Human-Centered Artificial Intelligence (HAI)

	 ⑤OpenAI	s cohere	stability.ai	ANTHROP\C	Google	BigScience	∞Meta	Al21 labs	ALEPH ALPHA	@ Eleuther RI	
Draft AI Act Requirements	GPT-4	Cohere Command	Stable Diffusion v2	Claude	PaLM 2	вьоом	LLaMA	Jurassic-2	Luminous	GPT-NeoX	Totals
Data sources	•000	•••0	••••	0000	••••	••••	••••	0000	0000	••••	22
Data governance	••00	$\bullet \bullet \bullet \circ$	••00	0000	$\bullet \bullet \bullet \circ$	••••	••••	0000	0000	$\bullet \bullet \bullet \circ$	19
Copyrighted data	0000	0000	0000	0000	0000	•••0	0000	0000	0000	• • • •	7
Compute	0000	0000	••••	0000	0000	• • • •	• • • •	0000	•000	• • • •	17
Energy	0000	•000	$\bullet \bullet \bullet \circ$	0000	0000	••••	••••	0000	0000	••••	16
Capabilities & limitations	••••	•••0	••••	•000	••••	$\bullet \bullet \bullet \circ$	••00	••00	•000	$\bullet \bullet \bullet \circ$	27
Risks & mitigations	•••0	••••	•000	•000	$\bullet \bullet \bullet \circ$	••00	•000	••00	0000	• 0 0 0	16
Evaluations	••••	••00	0000	0000	••00	•••0	••00	0000	•000	• 0 0 0	15
Testing	$\bullet \bullet \bullet \circ$	••00	0000	0000	••00	••00	0000	•000	0000	0000	10
Machine-generated content	$\bullet \bullet \bullet \circ$	$\bullet \bullet \bullet \circ$	0000	•••0	$\bullet \bullet \bullet \circ$	$\bullet \bullet \bullet \circ$	0000	\bullet \bullet \bullet \circ	•000	••00	21
Member states	••00	0000	0000	••00	••••	0000	0000	0000	•000	0000	9
Downstream documentation	$\bullet \bullet \bullet \circ$	••••	••••	0000	••••	••••	••••	0000	0000	$\bullet \bullet \bullet \circ$	24
Totals	25 / 48	23 / 48	22 / 48	7 / 48	27 / 48	36 / 48	21 / 48	8 / 48	5 / 48	29 / 48	

Philippines Regulation Is Coming Up





Salceda: Protect

Rep. Jose 'Joey' Salceda urged the government to come u

Salceda seeks safeguards vs threats from AI

manilatimes.net



Barbers: Create an Al Copyright Against Al Regulation Agency

Marcos: Protect Jobs Against Al

Congress pushes for AI Education and Funding



- Cong. Mark go pushing for educating 1% of Filipinos in Al
- Dedicated funding for AI research and startups
- Other congressmen pushing AI bills include Ace Barbers, Joey Salceda, and Mike Tan

Into the Rabbit Hole of GenAl

- Generative AI
- Applications of Al
- Implications of AI
- Strategies for Al
- Ethics of AI
- Looking Forward



Building AI Capability for Business Value



Design Thinking with AI (3-day F2F Immersive Bootcamp)



Prompting Al for Business (1-Day Intensive Workshop)



Al Prompt
Engineering
Online
(3-hour handson online
course)



Al Executive Briefing (2-hour interactive briefing)

www.cirrolytix.ai

In-house Briefings



- 1 hour briefing for companies, executives,
 and teams
- F2F or via Zoom
- Contact me via socials or support@cirrolytix.com

Let's Talk!







www.youtube.com/@docligot



Episode 9: cial Behavior Chance 48:22

Let's have Al for lunch

Let's have Al for lunch Episode 8: Al in Design Thinking 22 July 2023, 12nn





Al for Lunch Episode 10 -Reaction to Privileged...

Dominic Ligot 30 views • 2 hours ago

Al for Lunch Episode 9 - Al in Social Behavior Change

Dominic Ligot 58 views • 2 weeks ago Al For Lunch Episode 8 - Al in Design Thinking

Dominic Ligot

115 views • 1 month ago

Al For Lunch Episode 7 - Al in Digital Transformation

Dominic Ligot 140 views • 1 month ago Al For Lunch Episode 6 - Al for Social Impact

Dominic Ligot 372 views • 1 month ago

Al Conversations

Collecting compelling chats with AI practitioners, enthusiasts, and leaders.

Play all



Al Conversation - Gerard Ompad - Data Science Abuse

Dominic Ligot Premieres 8/28/23, 12:00 PM

Notify me

Al Conversations Doc Ligot and Fidel Ivan Racines Healthcare and Al

Al Conversations - Fidel Ivan Racines - Al in Healthcare

Dominic Ligot Premieres 8/27/23, 12:00 PM

Notify me

Al Conversations Doc Ligot and Serg Masis Responsible

Al Conversations - Serg Masis - Responsible Al

Dominic Ligot Premieres 8/26/23, 12:00 PM

Notify me

Al Conversations Doc Ligot and Gerando Bonganay Practical

Al Conversations - Gerando Bonganay - Practical AI

Dominic Ligot Premieres 8/25/23, 12:00 PM

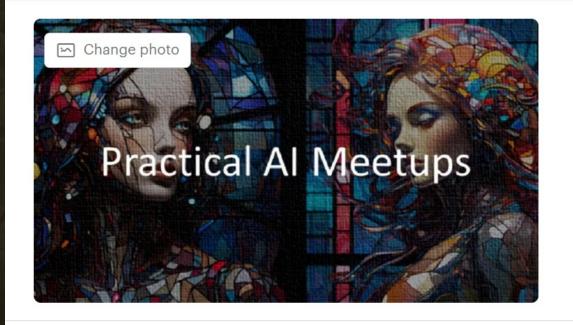
Notify me



Al Conversations - Cedric Miraflor - Vulnerable Al

Dominic Ligot Premieres 8/24/23, 12:00 PM

Notify me



Practical Al Meetups

Manila, Philippines

2 73 members · Public group 1

Organized by **Dominic Ligot** and **4 others**

Edit group info

Share: n 😝 🄰 🛅 🞽







About

Events

Members

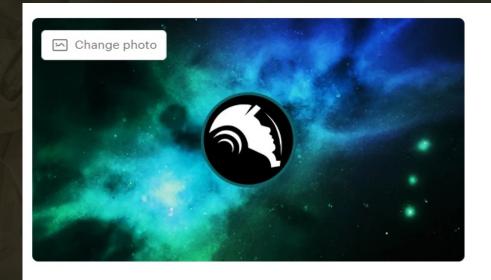
Photos

Discussions

More

Schedule

Manage group ∨



Part of cirrolytix.ai - 3 groups 1

Al Prompt Engineering Design Thinking and Applications

★★★★ (4) ②

New: Event feedback overview

X

Tap on the stars to see all of your event ratings in one place.

Learn More

- Manila, Philippines
- 2 108 members · Public group
- Organized by Dominic Ligot and 3 others



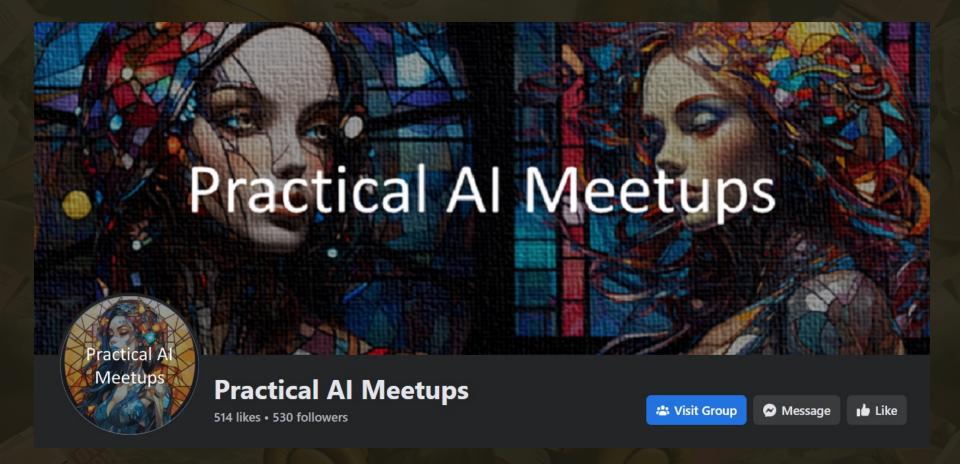
Share: 1











Groups



North Luzon Analytics Freelancers

♠ Private group · 20 members



Panay Smart Cities Network

♠ Private group · 3 members



Manila Analytics Freelancers Group

♠ Private group · 201 members



Mindanao Analytics Freelancers

O Public group · 148 members



South Luzon Analytics Freelancers

A Private group ⋅ 167 members



AI Safety and Alignment PH



+ Invite

Q

Discussio

Members

Events

Media

Files

Facebook - By Invitation