Generative AI

AI/ML/LLM Overview

Steinn Örvar Bjarnarson 2023

Who is this guy?

- Steinn Bjarnarson (Steinzi)
- Lead product development engineer at Advania
- Network engineer turned Automation specialist/developer
- Cisco CCIE #60715
- Fortinet NSE7
- Juniper Specialist
- BSc Computer Science

Modern network

Advania approach to modern networking

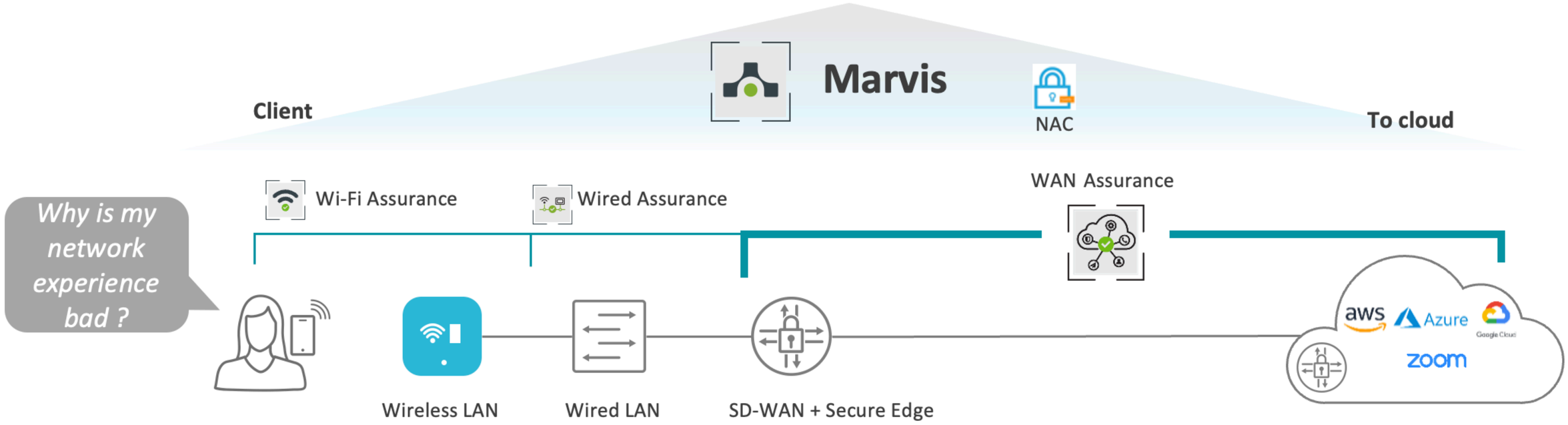
Client-to-Cloud – AI Driven driving

Mist
A Juniper Company

JUNIPER NETWORKS

advania

Black box



CONNECTED SECURITY – CLIENT TO CLOUD

Walmart amazon CVS pharmacy BEST BUY PETSMART T-Mobile servicenow

One-Stop-Shop

for **all** your **Network** needs

Overview

- What is ChatGPT and AI?
- **Four** Examples
  - **1. Text & data processing**
  - **2. Sound & Music**
  - **3. Vision**
  - **4. Education**
- Tools / What’s next / News!
- Summary

COMPUTER SCALE

40 years

1977
$6,532
3510 transistors
@ 50Khz

46 years

52 billion
Times
“better”

2023
$6,499
67 Billion transistors
@3.5Ghz

Apple 2 (3510 transistors) $1295 USD in 1977
$6,532 USD in 2023 (with inflation)

M2 Max (67 billion transistors) 3.49GHz
$6500 USD in 2023

What is

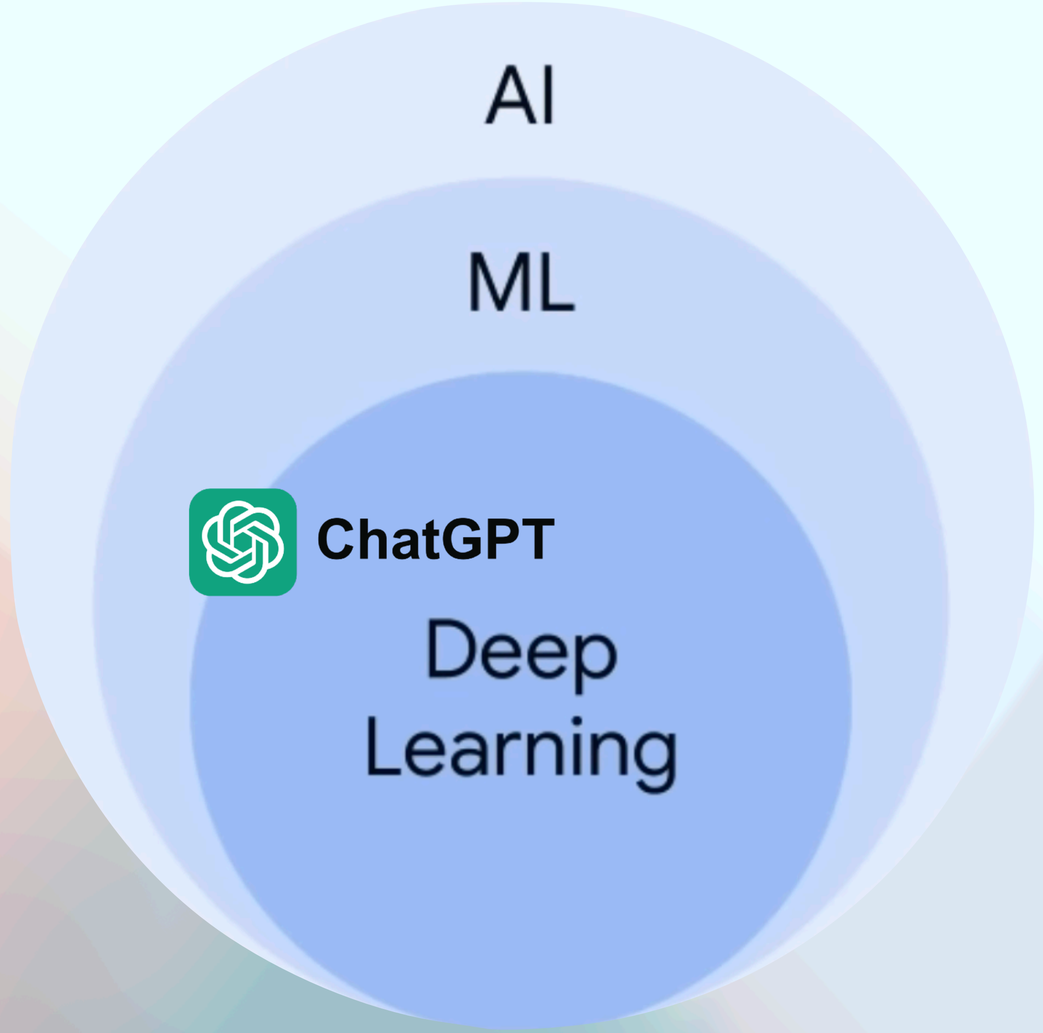
AI / ML

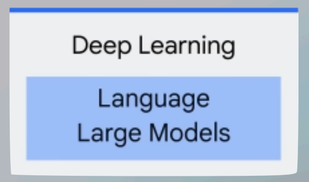
GenAI ?

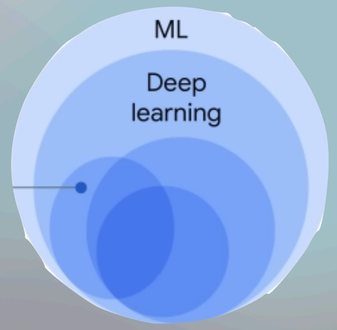
AI and ML?

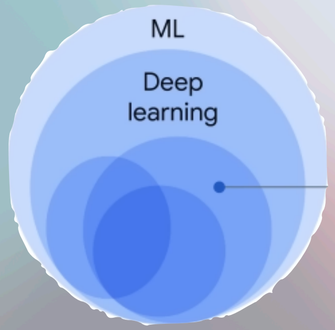
What is it?

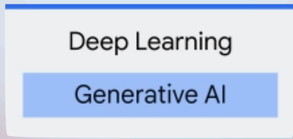
- AI is the theory and development of computer systems able to perform tasks normally requiring human intelligence.
- ML - Machine Learning, learn without me showing you how to learn exactly*
- Deep Learning (Cutting edge ML)
- Generative AI







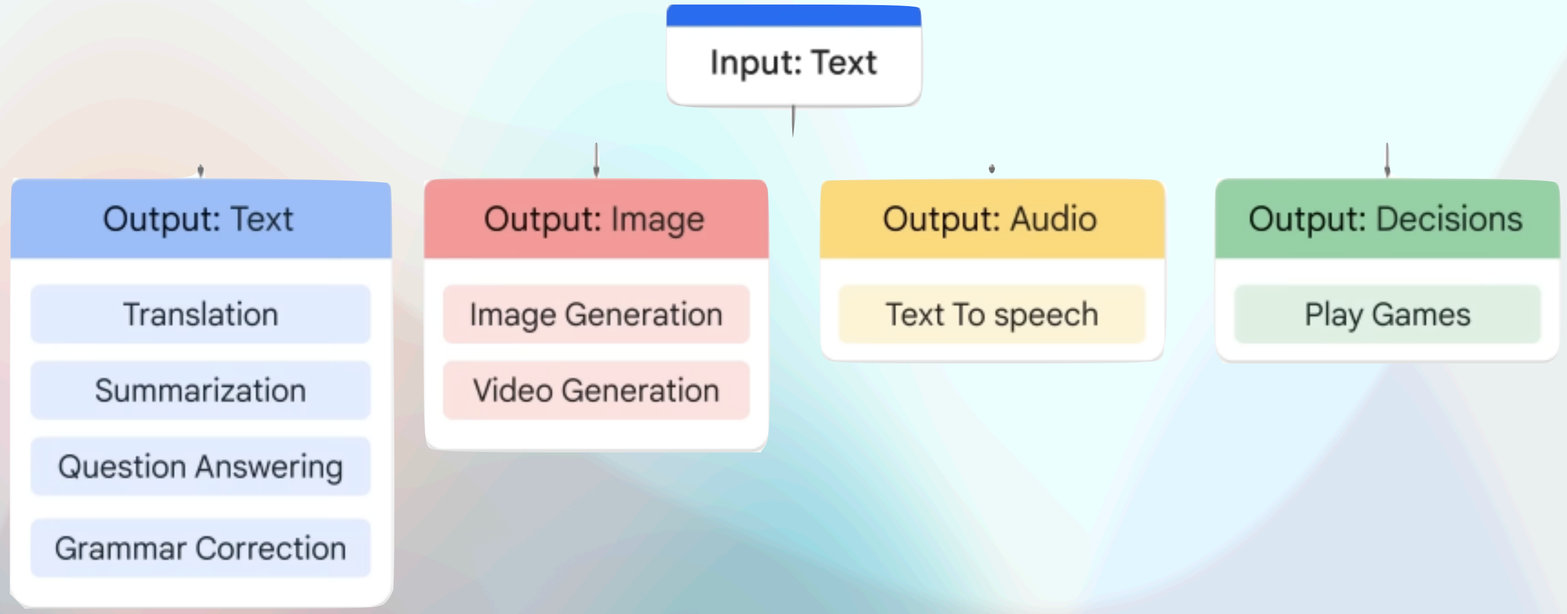




GenAI

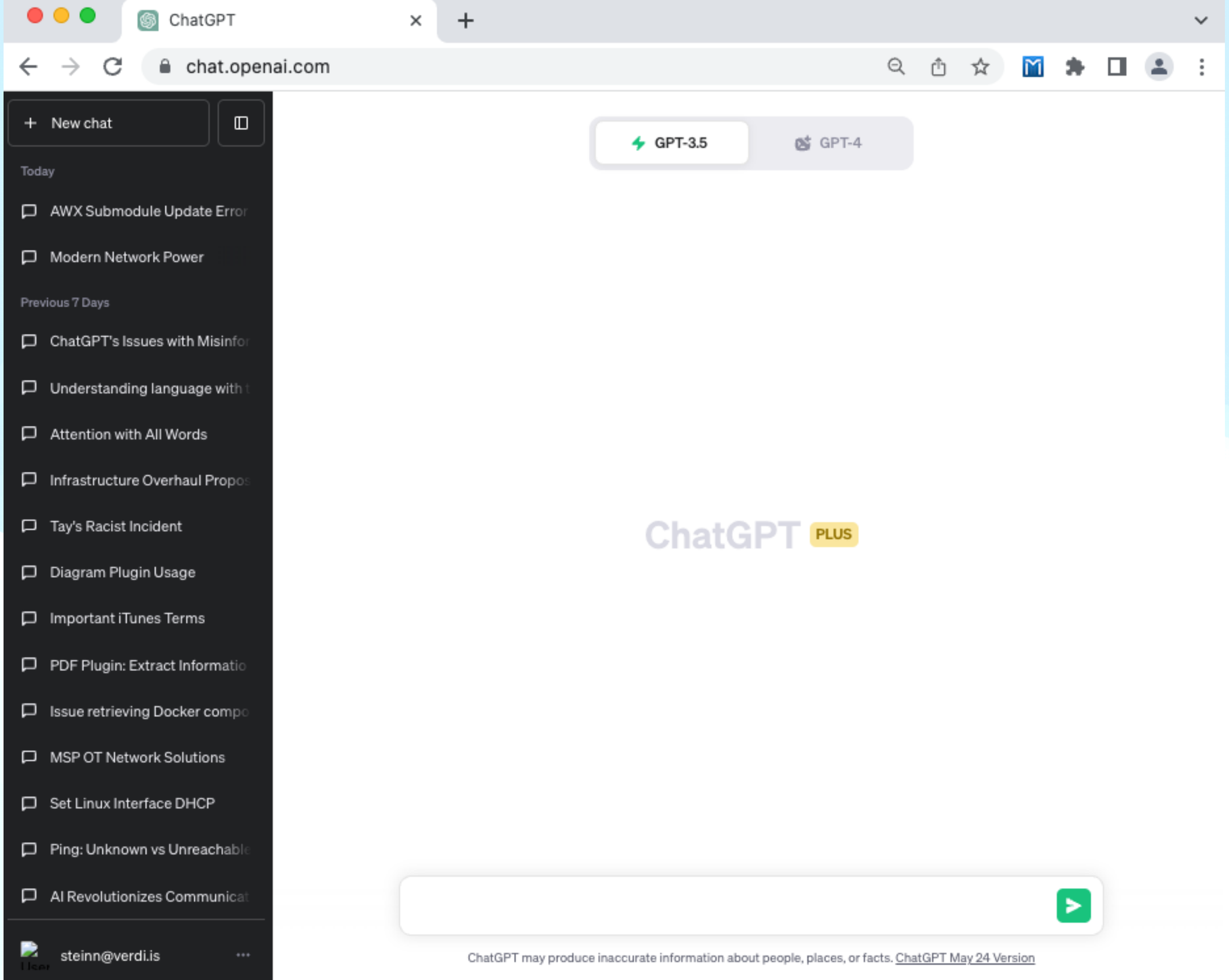
Types of GenAI based on Data

- ChatGPT = Text
- Midjourney = Image
- ElevenLabs = Audio



What is ChatGPT?

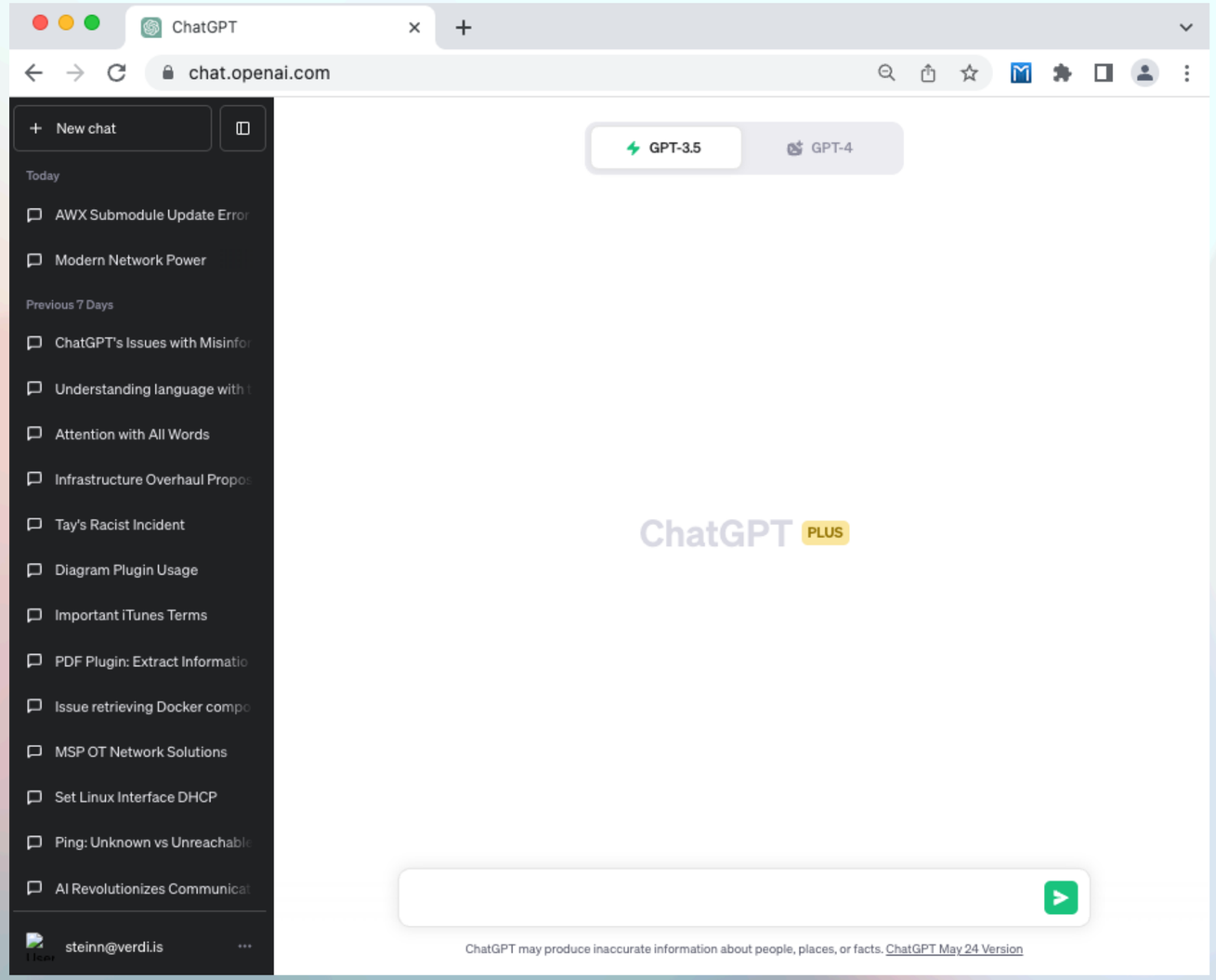
ChatGPT


ChatGPT
chat.openai.com
New chat
Today
AWX Submodule Update Error
Modern Network Power
Previous 7 Days
ChatGPT's Issues with Misinfor
Understanding language with t
Attention with All Words
Infrastructure Overhaul Propos
Tay's Racist Incident
Diagram Plugin Usage
Important iTunes Terms
PDF Plugin: Extract Informatio
Issue retrieving Docker compo
MSP OT Network Solutions
Set Linux Interface DHCP
Ping: Unknown vs Unreachable
AI Revolutionizes Communicat
steinn@verdi.is
GPT-3.5
GPT-4
ChatGPT PLUS
ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 24 Version

Chat Generative Pre-training *Transformer*

It’s a front end for an LLM

- LLM - Large Language Model
- GenAI - Generative AI
- Transformer
- Chatbot
- Text predictor
- Content generator



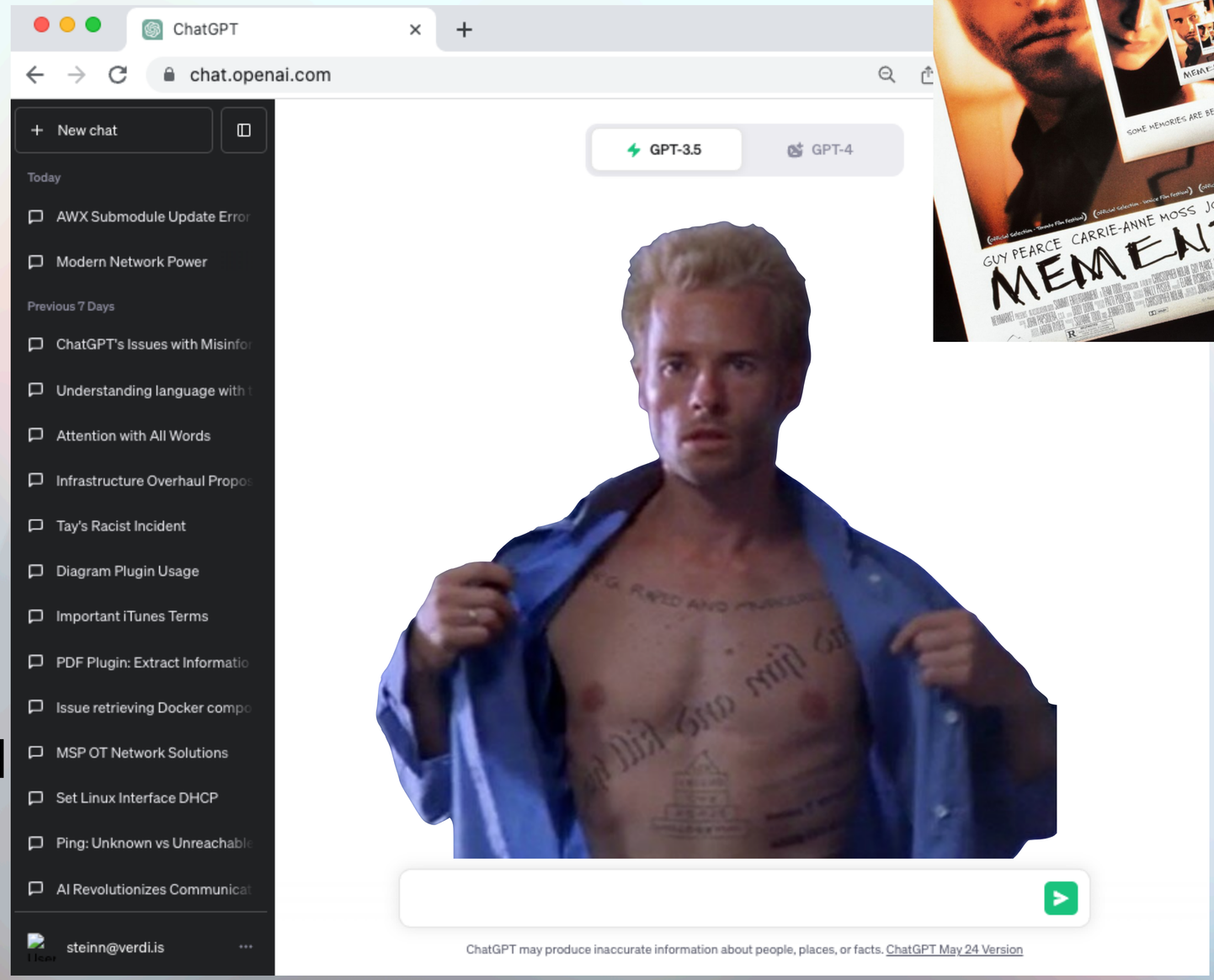
ChatGPT = Memento

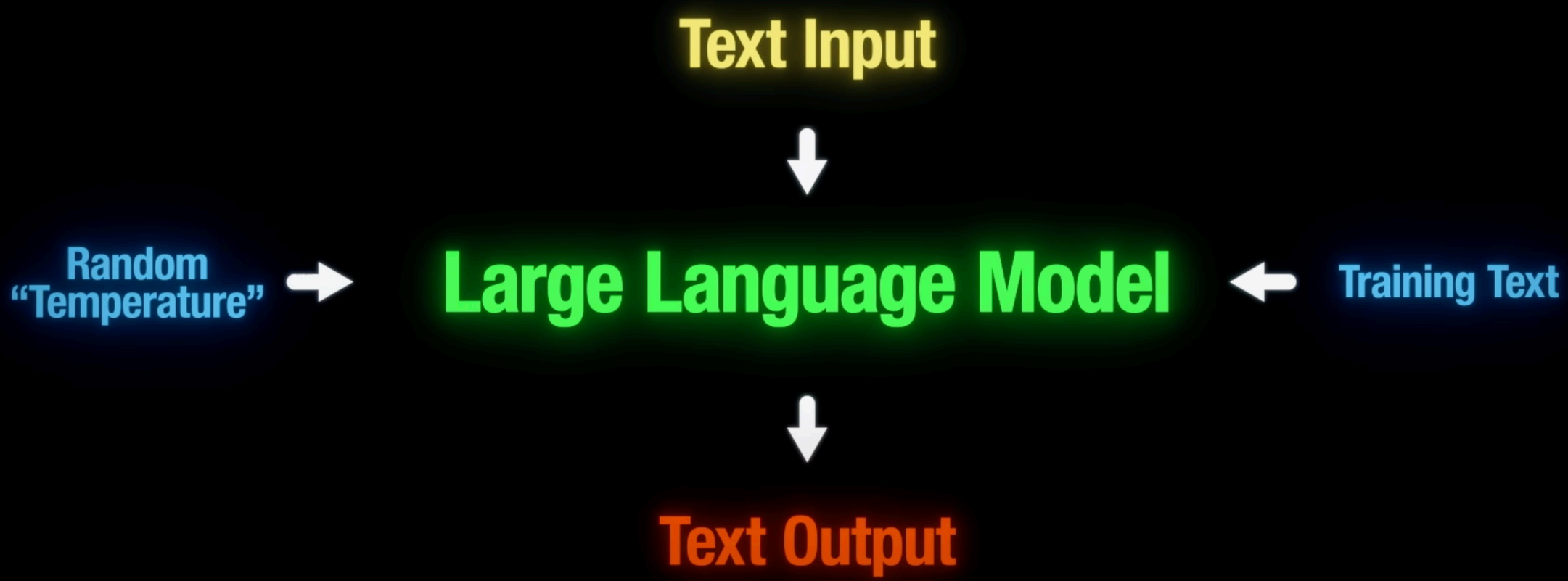
Memory is per session!

- No facts after **Sept 2021**

FRÉTTABLAÐIÐ

- Blank Canvas!
- No idea what you told him before!
- **Plugins** let it access limited new data!




Text Input
Random
“Temperature”
Large Language Model
Training Text
Text Output

Next-Token Prediction

Kyle is a _____

Kyle is a *gamer*
Kyle is a *clone*
Kyle is a *“scientist”*
Kyle is an *influencer*

Masked Language Modeling

Kyle _____ playing Sekiro

Kyle *loves* playing Sekiro
Kyle *fears* playing Sekiro
Kyle *hates* playing Sekiro
Kyle *avoids* playing Sekiro

Next-Token Prediction

Kyle is a _____

Kyle is a *gamer* (23)
Kyle is a *clone* (495)
Kyle is a *"scientist"* (12,008)
Kyle is an *influencer* (43,891)

Masked Language Modeling

Kyle _____ playing Sekiro

Kyle *loves* playing Sekiro (7)
Kyle *fears* playing Sekiro (34,622)
Kyle *hates* playing Sekiro (20,938)
Kyle *avoids* playing Sekiro (3,057)

Next-Token Prediction

Kyle is a _____

Kyle is a *gamer* (23)
Kyle is a *clone* (495)
Kyle is a *"scientist"* (12,008)
Kyle is an *influencer* (43,891)

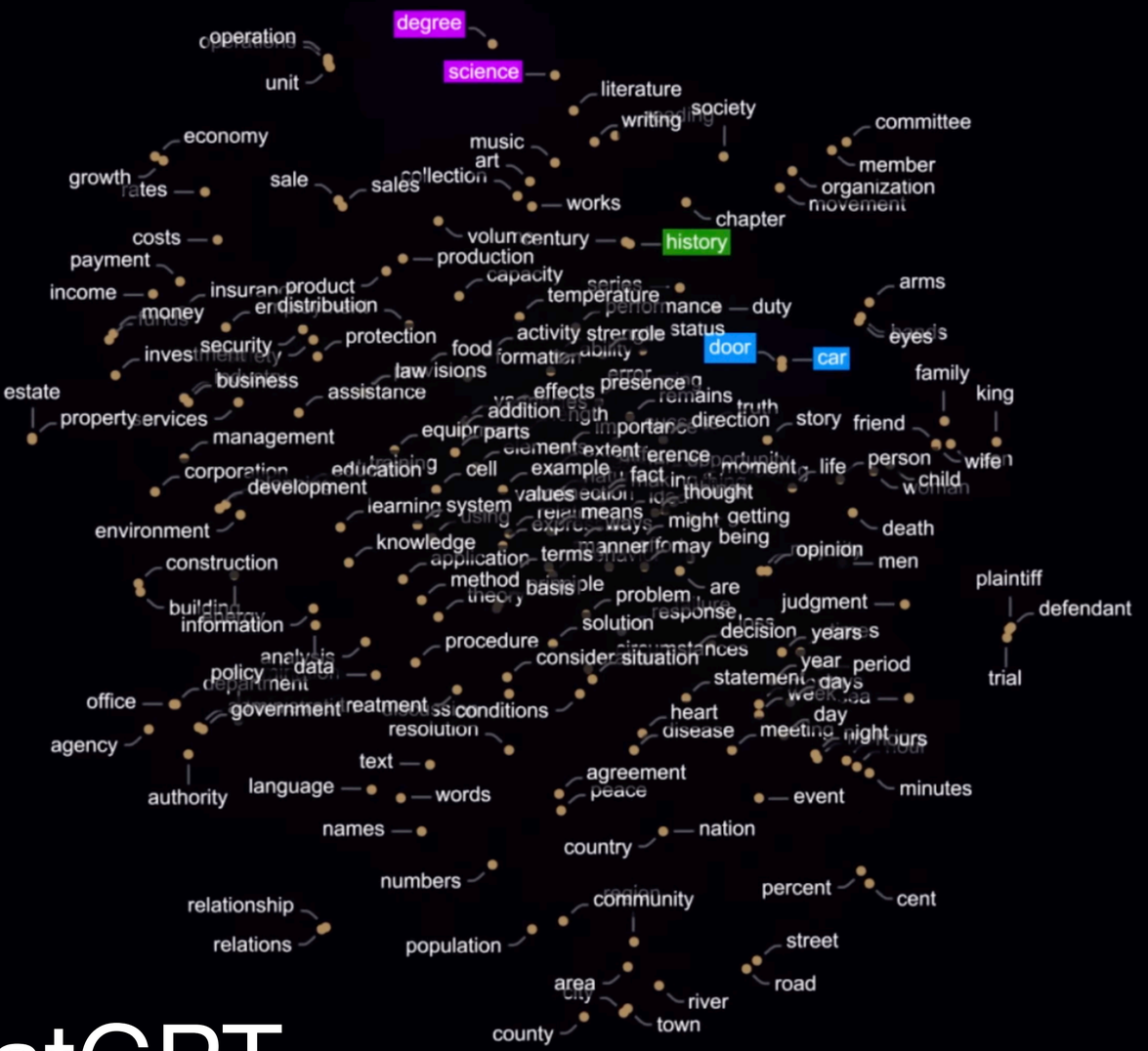
Masked Language Modeling

Kyle _____ playing Sekiro

Kyle *loves* playing Sekiro (?)
Kyle *fears* playing Sekiro (34,622)
Kyle *hates* playing Sekiro (?)
Kyle *avoids* playing Sekiro (?)

Words/Tokens relations

2D example!



12288D in chatGPT

Hallucinations

Big problem

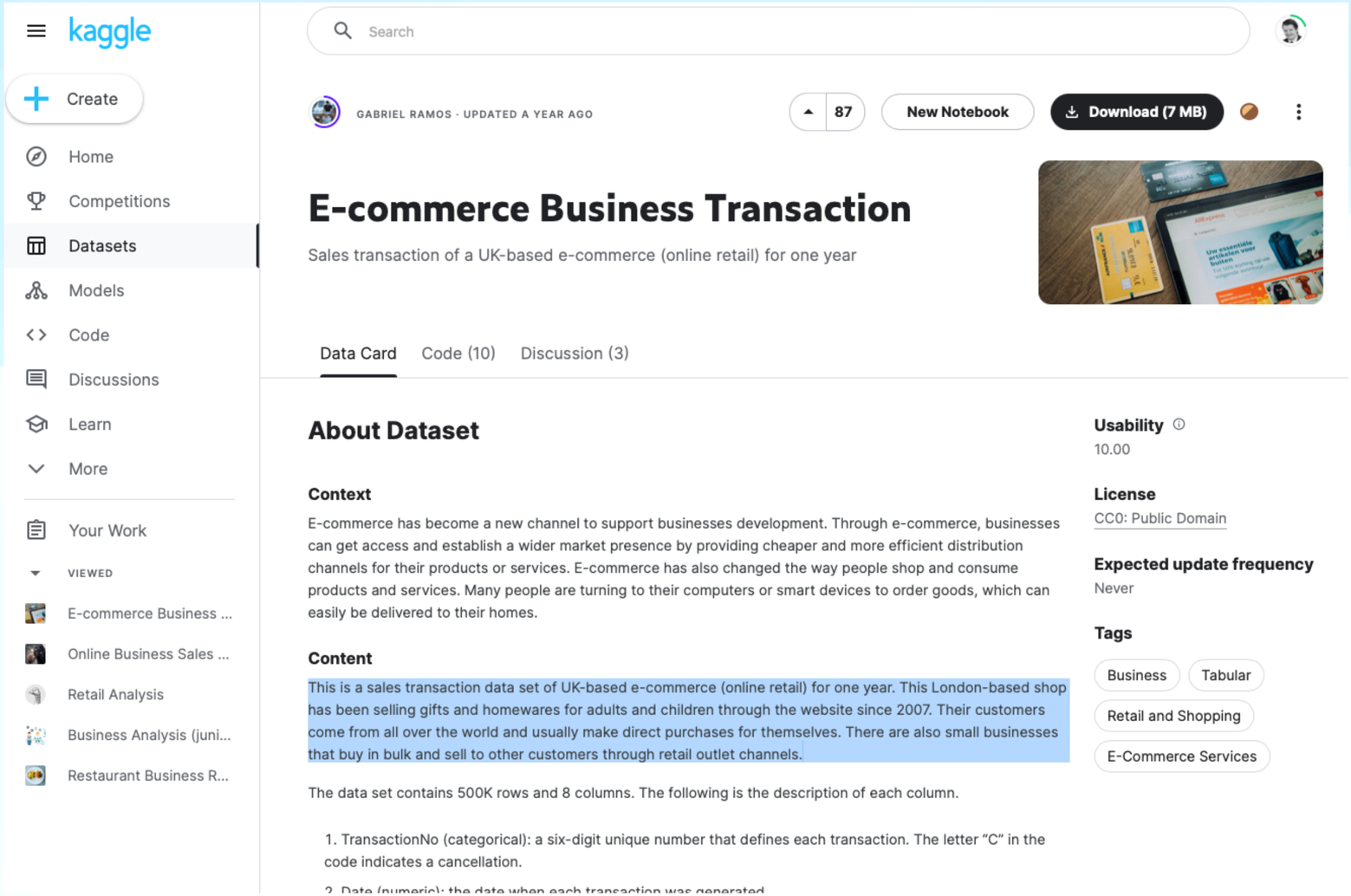
- **Hallucinations** are words or phrases generated by the model that are often nonsensical or grammatically incorrect.
- The model is not trained on enough data
- The model is trained on noisy or dirty data
- The model is not given enough context
- The model is not given enough constraints

Example 1

Text &
data processing

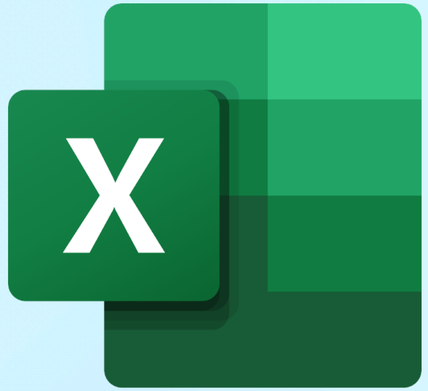
Data set source

Kaggle



Business Transactions

Online retail (UK)



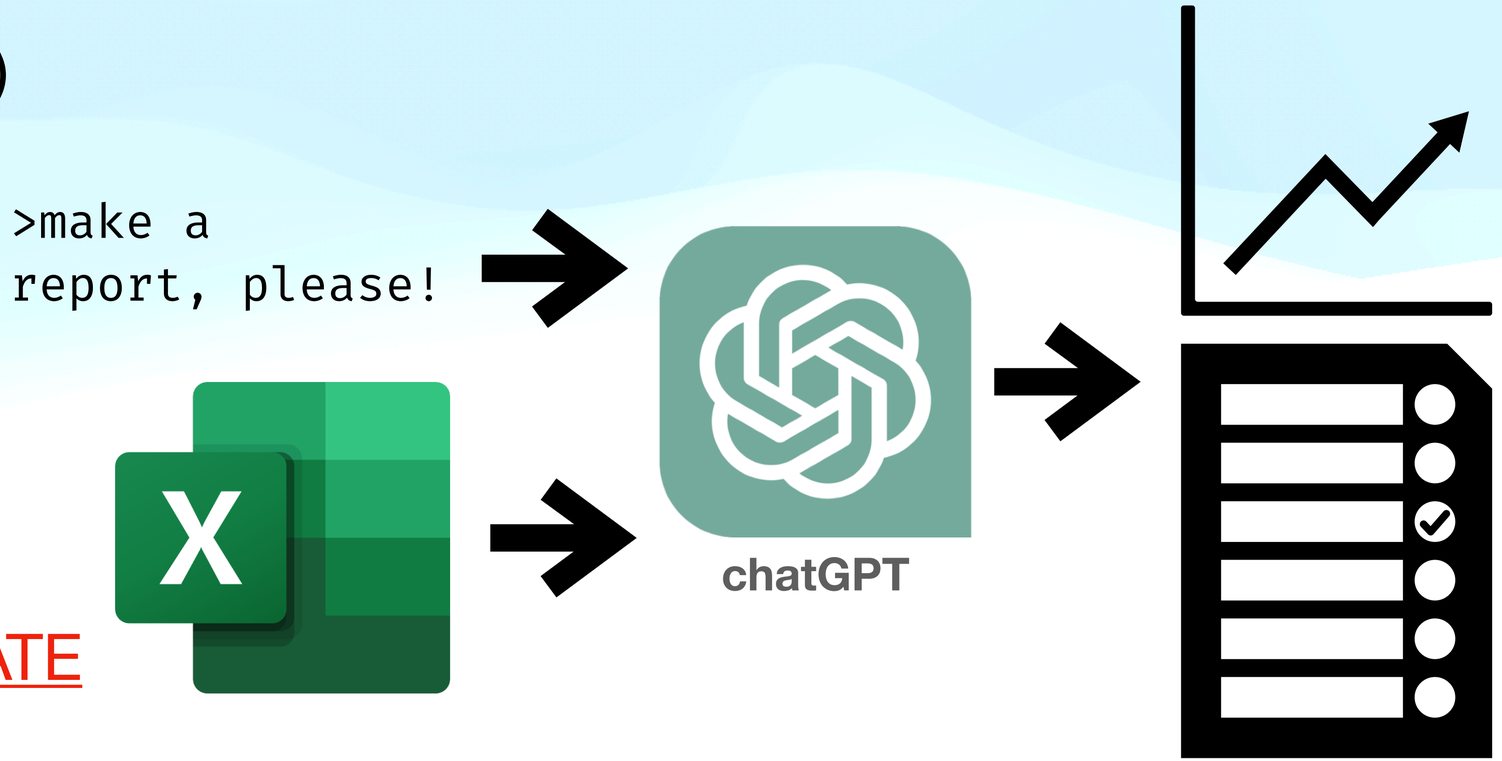
- E-commerce Business Transaction
- 2018-2019

TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
581482	12/9/2019	22485	Set Of 2 Wooden Market Crates	21.47	12	17490	United Kingdom
581475	12/9/2019	22596	Christmas Star Wish List Chalkboard	10.65	36	13069	United Kingdom
581475	12/9/2019	23235	Storage Tin Vintage Leaf	11.53	12	13069	United Kingdom
581475	12/9/2019	23272	Tree T-Light Holder Willie Winkie	10.65	12	13069	United Kingdom
581475	12/9/2019	23239	Set Of 4 Knick Knack Tins Poppies	11.94	6	13069	United Kingdom
581475	12/9/2019	21705	Bag 500g Swirly Marbles	10.65	24	13069	United Kingdom
581475	12/9/2019	22118	Joy Wooden Block Letters	11.53	18	13069	United Kingdom
581475	12/9/2019	22119	Peace Wooden Block Letters	12.25	12	13069	United Kingdom
581475	12/9/2019	22217	T-Light Holder Hanging Lace	10.65	12	13069	United Kingdom
581475	12/9/2019	22216	T-Light Holder White Lace	10.55	24	13069	United Kingdom
581475	12/9/2019	22380	Toy Tidy Spaceboy	11.06	20	13069	United Kingdom
581475	12/9/2019	22442	Grow Your Own Flowers Set Of 3	12.25	12	13069	United Kingdom
581475	12/9/2019	22664	Toy Tidy Dolly Girl Design	11.06	20	13069	United Kingdom
581475	12/9/2019	22721	Set Of 3 Cake Tins Sketchbook	12.25	12	13069	United Kingdom

The plan

Online retail (UK)

- Provide business data (.CSV)
- Basic prompt >
- Process data
- Enjoy report!
- **NOTE**: DO NOT GIVE IT CUSTOMER INFO OR PRIVATE DATA!



Demo time!

ChatGPT PLUS

Plan a trip
to see the northern lights in Norway

Make up a story
about Sharky, a tooth-brushing shark superhero

Help me pick
a birthday gift for my mom who likes gardening

Explain why popcorn pops
to a kid who loves watching it in the microwave

Send a message

?

Þetta hljómaði frekar stíft..

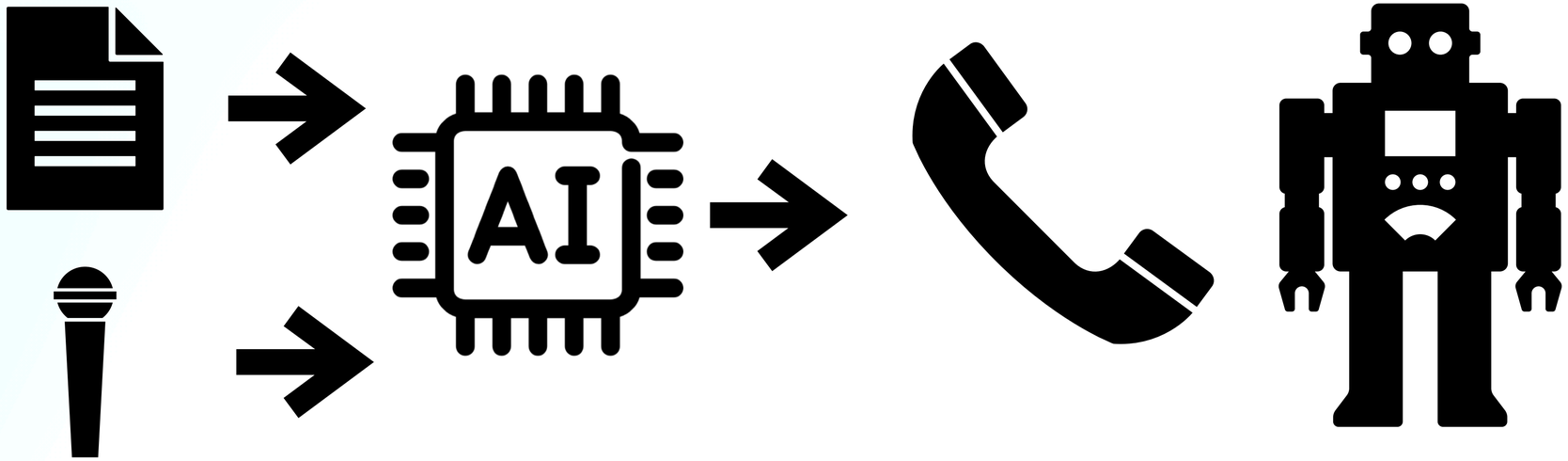
Example 2

Sound & Music

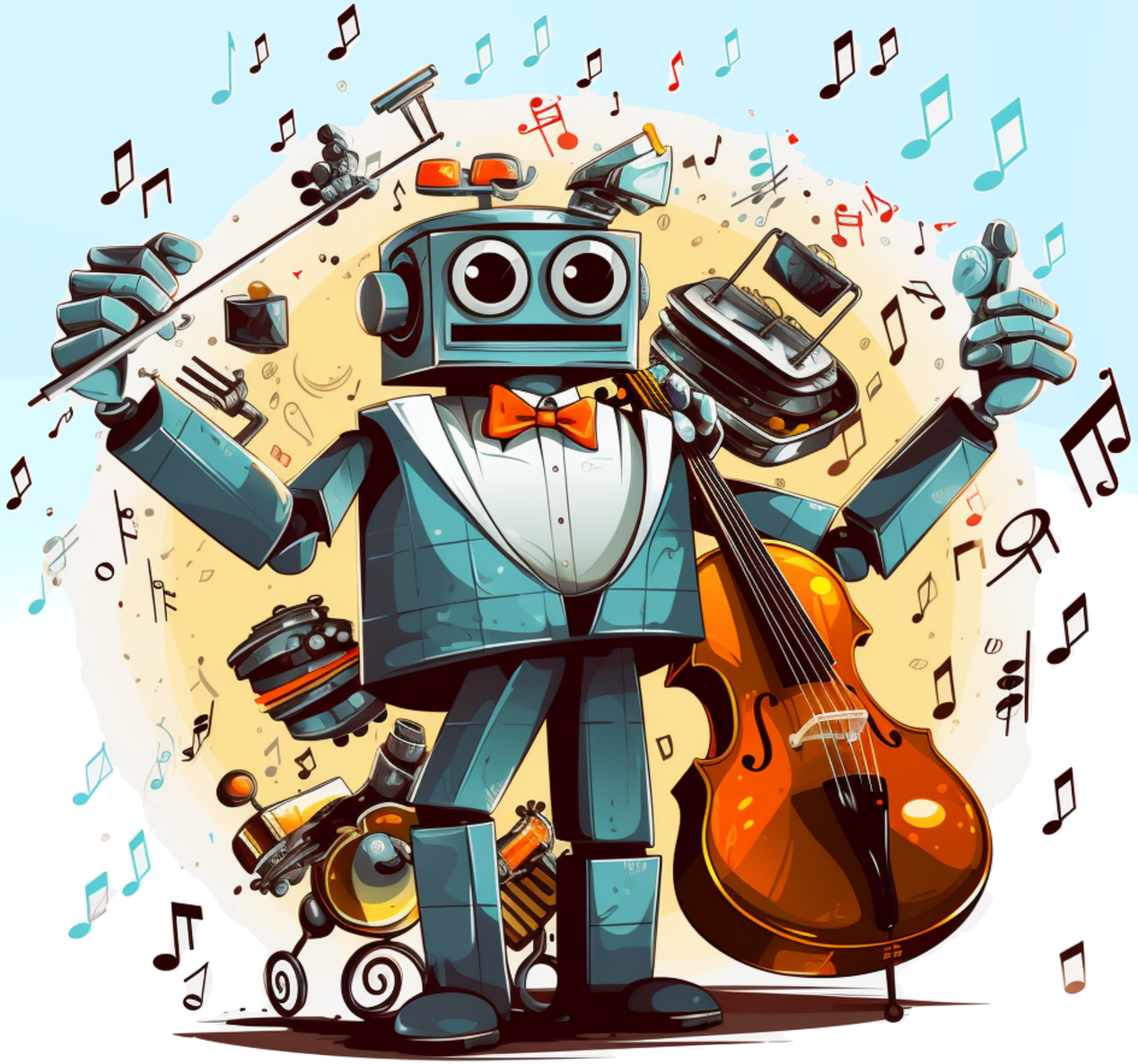
Speech Synthesis

“Text to speech”

- Process voice samples into a model.
- Uses a Multi-modal model (text + voice data)

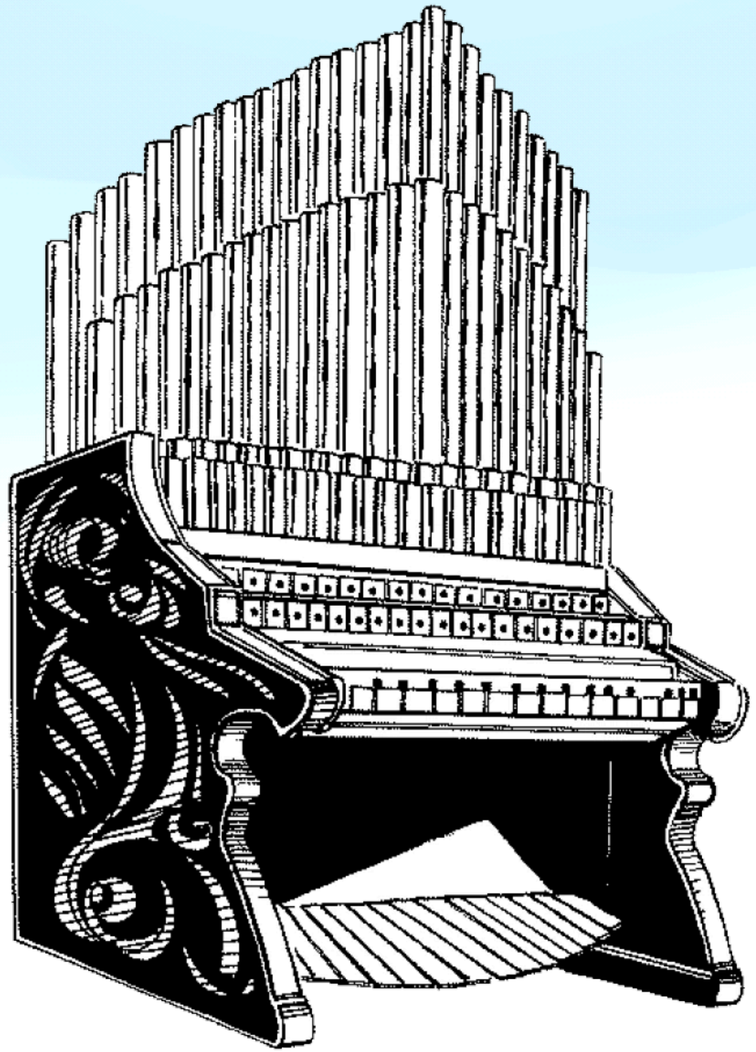


Music



Music

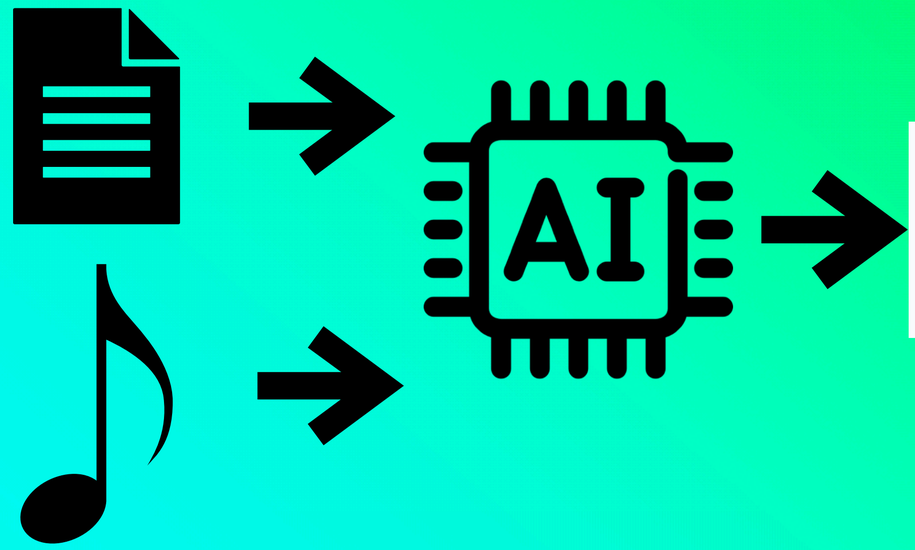
Toccata and Fugue in D minor

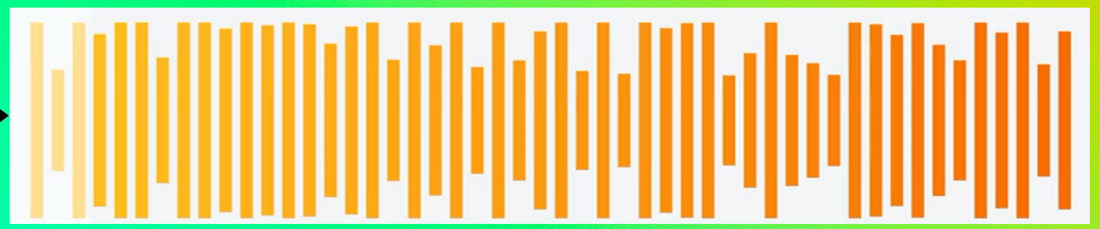


Music

Toccata and Fugue in D minor

“An 80s driving pop song with heavy drums and synth pads in the background”





Music

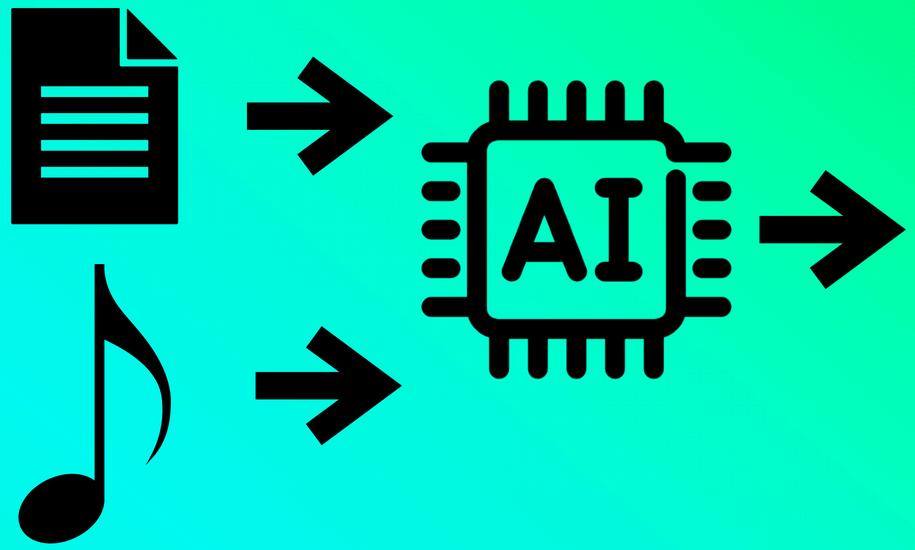
RUV

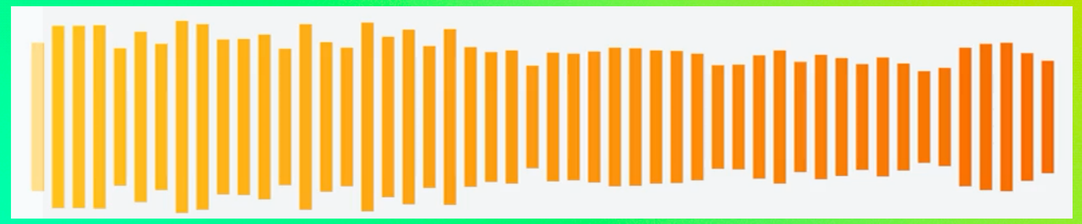
- Short and sweet

Music

RUV

“An 80s driving pop song with heavy drums and synth pads in the background”





Example 3

Vision

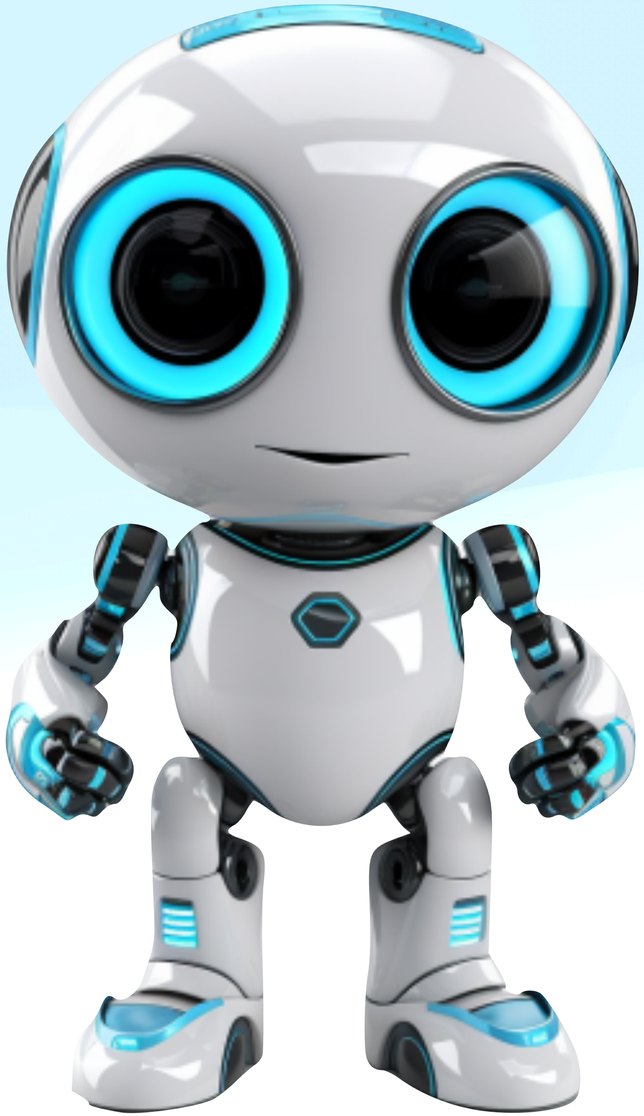


Image creation

Midjourney

- “Walt Disney pixar art of Davíð Oddsson, eating ice cream in the whitehouse”

Vision processing?

~The year is 2015


SEEFOOD

~8 years later in 2023

Image processing

LLaVA

- “Please analyze the photo and inform me if there is a "hotdog" in the photo with the response "hotdog" (there wont always be a hotdog in the photo) and if there is "No hotdog" in the photo, the response "no hotdog" or if you are not sure use "unknown"
- Do not say anything except 1 of these 3 options ["no hotdog", "hotdog", “unknown"]"

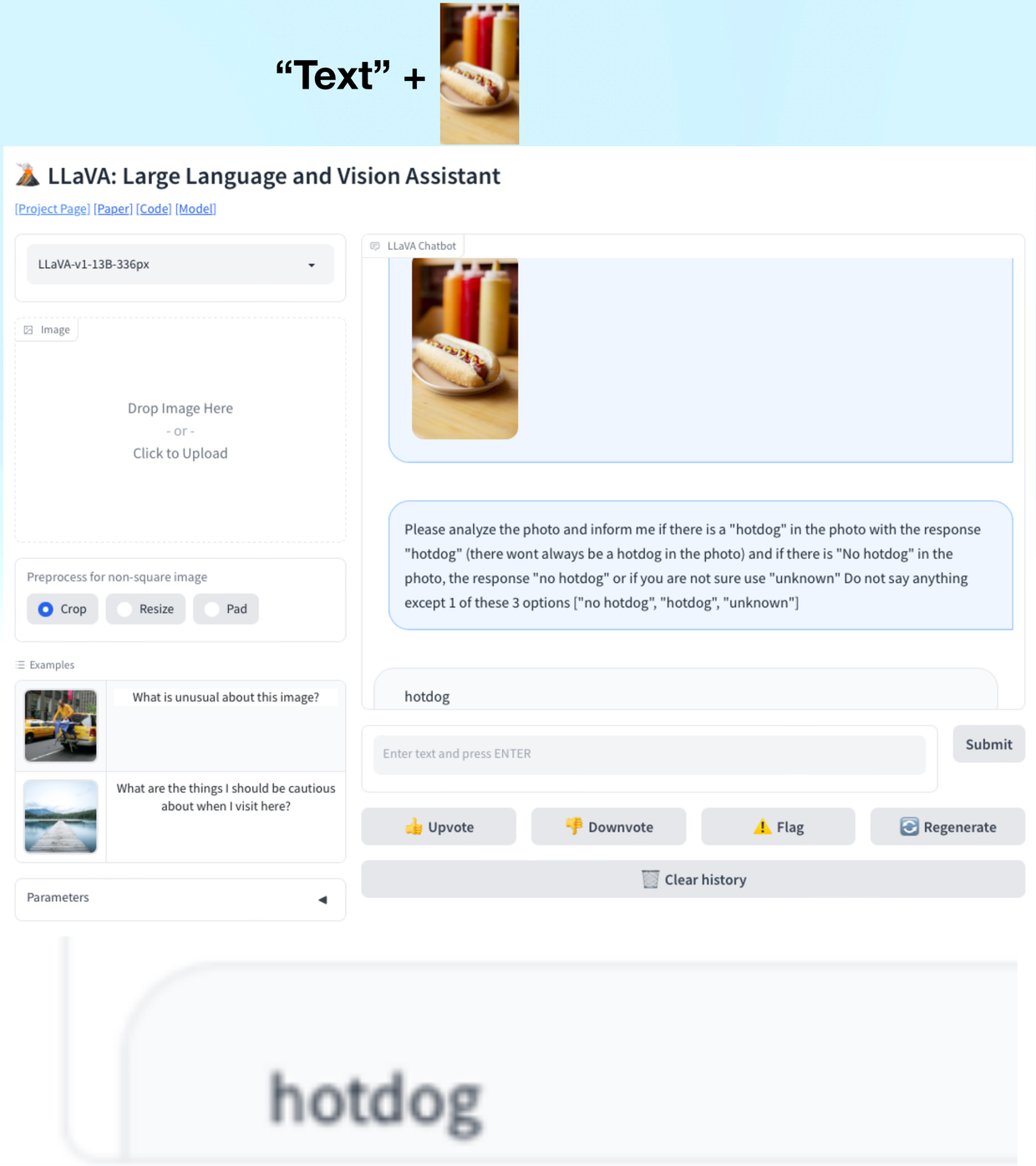
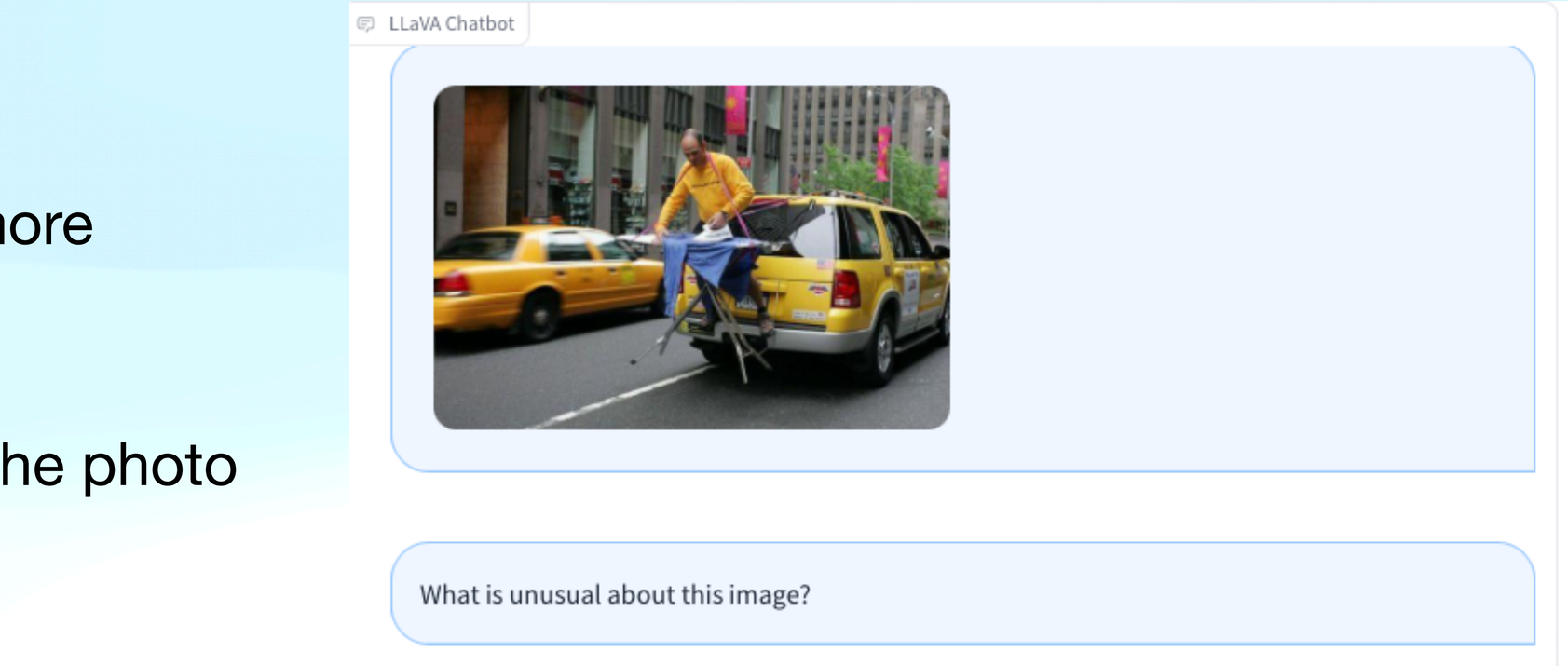


Image processing

LLaVA

- We can do so much more
- Explain the photo
- Tell me a story about the photo

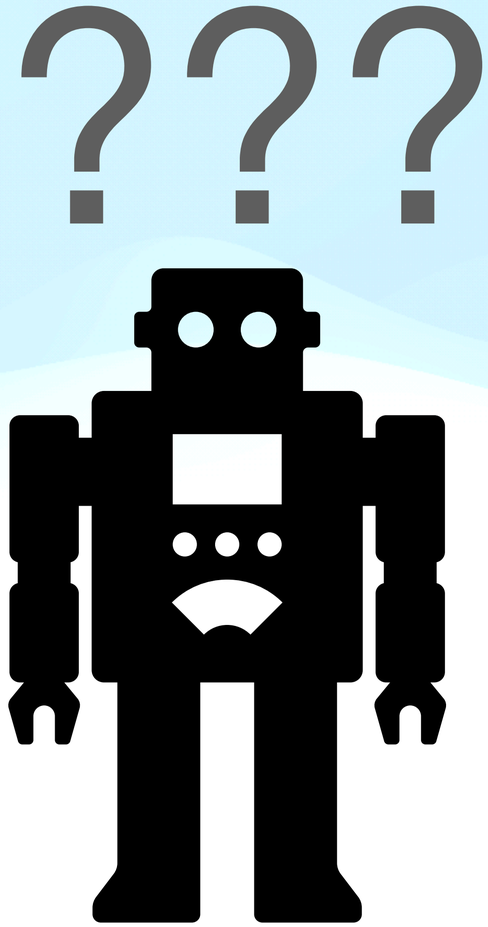
“Text” + 



What can we do?

CAPTCHAs

We can solve them!



CAPTCHAs

An Empirical Study & Evaluation of Modern CAPTCHAs

- AI are up to ~15% better at solving this than humans*

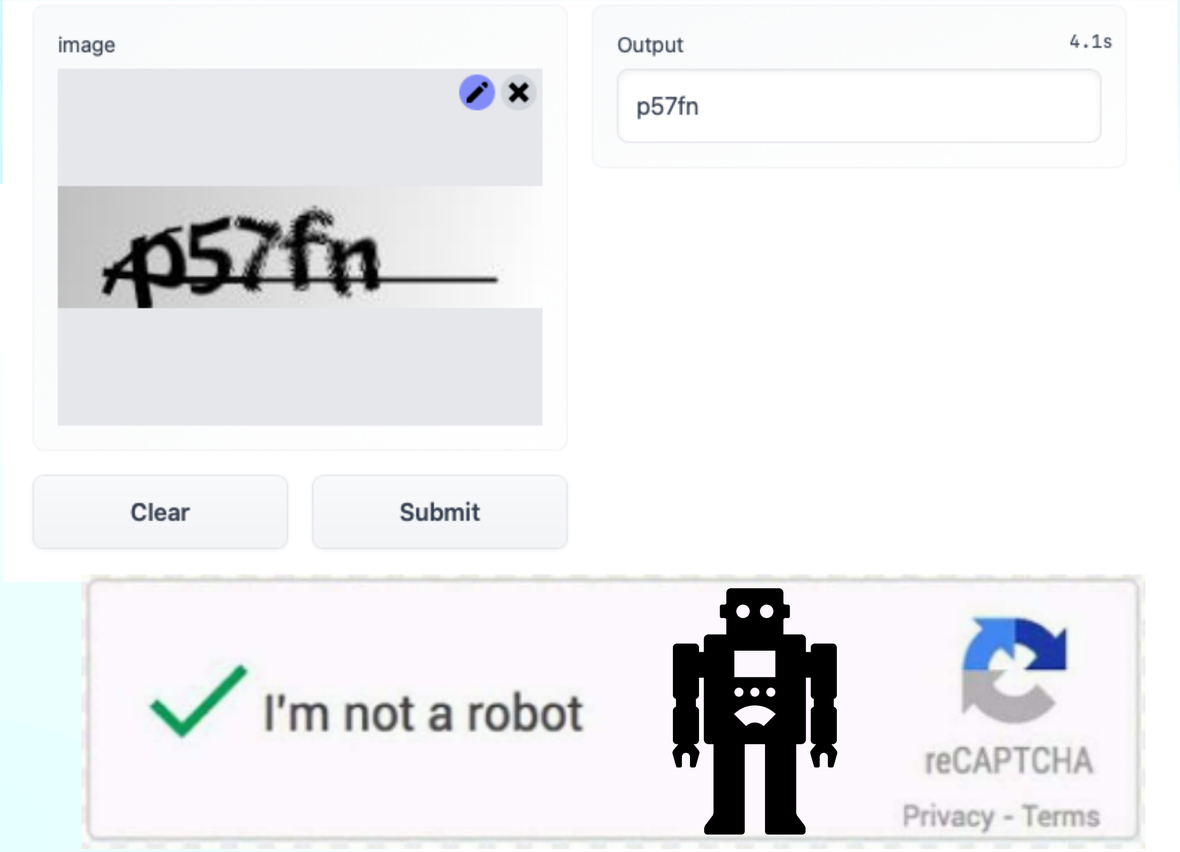


Table 3: Humans vs. bot solving time (seconds) and accuracy (percentage) for different CAPTCHA types.

	Human		Bot	
CAPTCHA Type	Time	Accuracy	Time	Accuracy
reCAPTCHA (click)	3.1-4.9	71-85%	1.4 [63]	100% [63]
Geetest	28-30	N/A	5.3 [70]	96% [70]
Arkose	18-42	N/A	N/A	N/A
Distorted Text	9-15.3	50-84%	<1 [77]	99.8% [39]
reCAPTCHA (image)	15-26	81%	17.5 [45]	85% [45]
hCAPTCHA	18-32	71-81%	14.9 [44]	98% [44]

Table 4: Agreement for distorted text CAPTCHAS.

	Average Agreement	Average Agreement (case insensitive)
Simple	84%	93%
Masked	50%	73%
Moving	62%	90%
Total	65%	85%

https://arxiv.org/pdf/2307.12108.pdf

Research paper on AI and Captcha

An Empirical Study & Evaluation of Modern CAPTCHAs

- You can read it here https://arxiv.org/pdf/2307.12108.pdf
- 17 pages

An Empirical Study & Evaluation of Modern CAPTCHAs

Andrew Searles, *UC Irvine*
Yoshimichi Nakatsuka[*], *ETH Zürich*
Ercan Ozturk, *UC Irvine*
Andrew Paverd, *Microsoft*
Gene Tsudik, *UC Irvine*
Ai Enkoji[†], *LLNL*

Abstract

For nearly two decades, CAPTCHAs have been widely used as a means of protection against bots. Throughout the years, as their use grew, techniques to defeat or bypass CAPTCHAs have continued to improve. Meanwhile, CAPTCHAs have also evolved in terms of sophistication and diversity, becoming increasingly difficult to solve for both bots (machines) and humans. Given this long-standing and still-ongoing arms race, it is critical to investigate how long it takes legitimate users to solve modern CAPTCHAs, and how they are perceived by those users.

In this work, we explore CAPTCHAS *in the wild* by evaluating users' solving performance and perceptions of *unmodified currently-deployed* CAPTCHAS. We obtain this data through manual inspection of popular websites and user studies in which 1,400 participants collectively solved 14,000 CAPTCHAS. Results show significant differences between the most popular types of CAPTCHAS: surprisingly, solving time and user perception are not always correlated. We performed a comparative study to investigate the effect of *experimental context* – specifically the difference between solving CAPTCHAS directly versus solving them as part of a more natural task, such as account creation. Whilst there were several potential confounding factors, our results show that experimental context could have an impact on this task, and must be taken into account in future CAPTCHA studies. Finally, we investigate CAPTCHA-induced user task *abandonment* by analyzing participants who start and do not complete the task.

website functionality intended for human use [31, 46]. If left unchecked, bots can perform these nefarious actions at scale. CAPTCHAS are a widely-deployed defense mechanism that aims to prevent bots from interacting with websites by forcing each user to perform a task, such as solving a challenge [5]. Ideally, the task should be straightforward for humans, yet difficult for machines [68].

The earliest CAPTCHAS asked users to transcribe random distorted text from an image. However, advances in computer vision and machine learning have dramatically increased the ability of bots to recognize distorted text [35, 41, 74], and by 2014, automated tools achieved over 99% accuracy [39, 62]. Alternatively, bots often outsource solving to CAPTCHA *farms* – sweatshop-like operations where humans are paid to solve CAPTCHAS [54]. In light of this, CAPTCHAS have changed and evolved significantly over the years. Popular CAPTCHA tasks currently include object recognition (e.g., "select squares with..."), parsing distorted text, puzzle solving (e.g., "slide the block..."), and user behavior analysis [39, 62]. It is therefore critical to understand and quantify how long it takes legitimate users to solve current CAPTCHAS, and how these CAPTCHAS are perceived by users.

Several prior research efforts have explored CAPTCHA solving times, e.g., [24, 27, 33, 37, 58, 67]. For example, over a decade ago, Bursztein et al. [27] performed a large-scale user study, using over 1,100 unique participants from Amazon Mechanical Turk (MTurk) [3] as well as CAPTCHA farms. Their results showed that CAPTCHAS were often more difficult or took longer to solve than was expected. There was a loose correlation between time-to-annoyance and abandon-

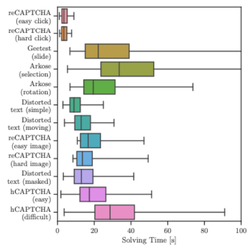


Figure 7: Solving times for various types of CAPTCHAS. Boxes show the middle 50% of participants, and whiskers show the filtered range. Black vertical lines show the median.

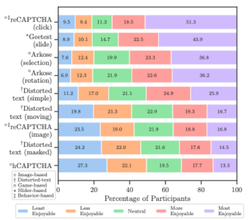


Figure 8: Participant-reported preference scores for different types of CAPTCHAS, sorted from highest to lowest.

of hCAPTCHA are visually indistinguishable, we could only ask participants for one preference.

As expected, participants tend to prefer CAPTCHAS with lower solving times. For example, reCAPTCHA (click) has the lowest median solving time and the highest user preference. However, surprisingly, this trend does not seem to hold for game-based and slider-based CAPTCHAS, since these received some of the highest preference scores, even though they typically took longer than other types. This suggests that factors beyond solving time could be contributing to participants' preference scores. Notably, no single CAPTCHA type is either universally liked or disliked. For example, even the top-rated click-based reCAPTCHA, was rated 1 or 2 by 18.9% of participants. Similarly, over 31.0% rated hCAPTCHA 4 or 5, although it had the lowest overall preference score.

5.3 Direct vs. contextualized setting

This subsection addresses **RQ3:** *Does experimental context affect solving time?* Figure 9 shows histograms of CAPTCHA solving times for participants in the direct vs. contextualized settings. In every case except one, the mean solving time is lower in the direct setting. In most cases, the distribution from the contextualized setting has more participants with longer solving times, i.e., a longer tail.

The largest statistically significant difference is in reCAPTCHA (easy click), where the mean solving time grows by 1.8 seconds (57.5%). Second is Arkose (rotation), where it grows by 10 seconds (56.1%). Across all CAPTCHA types, the average increase from direct to contextualized is 26.7%. Similarly, the mean solving time for reCAPTCHA (easy image) increased by 63.6% in the contextualized setting showing the largest increase. However this was not statistically significant. This is likely due to the skew of participants in direct and contextualized versions receiving image-challenges, which is controlled by Google. Easy images were shown to 8.9% of contextualized and to 17.2% of direct setting participants, while hard images were shown to 25.5% and 30% respectively, resulting in different population sizes.

On the other hand, hCAPTCHA (difficult), which has the highest median solving time overall, showed no significant difference in mean solving time between direct and contextualized settings. This may be attributable to the difficulty of solving this type of CAPTCHA, regardless of the setting.

Results of Kruskal-Wallis tests confirm that there are statistically significant differences in mean solving times for all CAPTCHA types ($p < 0.001$) except Geetest, reCAPTCHA (image) and hCAPTCHA (difficult). While there were several potential confounding factors in our study, these results suggest that experimental context can have a significant impact on participants' CAPTCHA solving times, and must therefore be taken into account in the design of future user studies.

5.4 Effects of demographics

This subsection addresses **RQ4:** *Do demographics affect solving time?* We analyzed how demographic characteristics in our study correlate with CAPTCHA solving times. For some characteristics, such as education and gender, we did not observe large differences in CAPTCHA solving times (see Figures 13 and 14 in Appendix D).

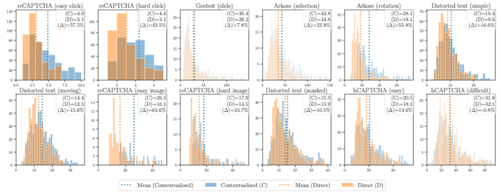


Figure 9: CAPTCHA solving times for direct (D) vs. contextualized (C) user study settings. The horizontal axis shows solving time in seconds, quantized into one-second buckets, and the vertical axis shows number of participants.

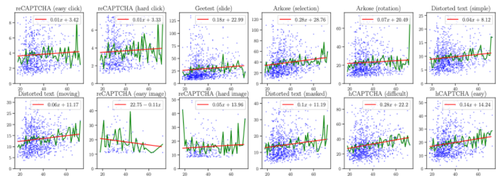


Figure 10: Effects of age in CAPTCHA solving time. The horizontal axis shows the age and the vertical axis shows the solving time. The red line shows the linear fit of the data points and the green line shows the average solving time per age.

5.4.1 Effects of age

Figure 10 shows the effect of participants' age on solving time. The green line is the average solving time for each age, and the red line is a linear fit minimizing mean square error. For all types, except reCAPTCHA (easy image), there is a trend of younger participants having lower average solving times. This agrees with prior results [27] and is especially noticeable in hCAPTCHA, Arkose (selection), and Geetest.

5.4.2 Effects of device type

Figure 11 shows the effect of device type. Although there are some differences in median between device types for certain CAPTCHA types, the Kruskal-Wallis test shows that the differences in means are mostly not statistically significant. The only statistically significant differences are in distorted text CAPTCHAS ($p < 0.02$) and reCAPTCHA (hard click) ($p < 0.01$), where participants who used computers had a lower mean solving time compared to those using phones. Interestingly, we found a statistically significant difference between participants who used physical keyboards and those who used touch input for the simple and masked distorted text CAPTCHAS ($p < 0.02$), as well as reCAPTCHA (hard click) ($p < .001$), reCAPTCHA (easy click) ($p < .05$), and Arkose (selection) ($p < .003$). We found no statistically significant difference in mean solving times for moving distorted text.

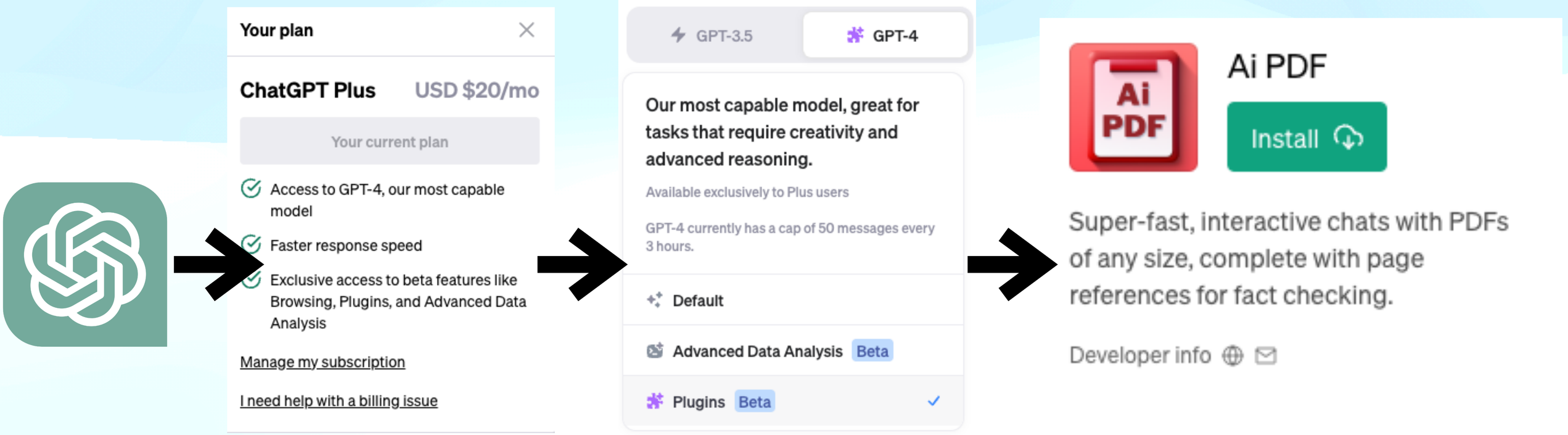
References

[1] 360.cn. https://passport.360.cn/.
[2] Alexa Top Sites. https://www.alexa.com/topsites.
[3] Amazon Mechanical Turk. https://www.mturk.com/.
[4] Arkose Labs. https://www.arkoselabs.com/about-us/.
[5] CAPTCHA Usage Distribution on the Entire Internet. https://trends.builtwith.com/widgets/captcha/traffic/Entire-Internet.
[6] Cisco Umbrella 1 Million. https://umbrella.cisco.com/blog/cisco-umbrella-1-million.
[7] Cloudflare Radar - Domain Rankings. https://radar.cloudflare.com/domains.
[8] GeeTest CAPTCHA. https://www.geetest.com/en/Captcha.
[9] hCaptcha. https://www.hcaptcha.com/.
[10] hCaptcha Is Now The Largest Independent CAPTCHA Service, Runs on 15% Of The Internet. https://www.hcaptcha.com/post/hcaptcha-now-the-largest-independent-captcha-service.
[11] Invisible reCAPTCHA. https://developers.google.com/recaptcha/docs/invisible.
[12] jrj.com. https://sso.jrj.com.
[13] NuData Security. https://nudatasecurity.com/.
[14] reCAPTCHA. https://www.google.com/recaptcha/about/.
[15] reCAPTCHA v3. https://developers.google.com/recaptcha/docs/v3.
[16] The Majestic Million. https://majestic.com/reports/majestic-million.
[17] The Tor Project: Privacy & Freedom Online. https://www.torproject.org/.
[18] Xinhuanet. https://mail.xinhuanet.com.
[19] MongoDB. https://www.mongodb.com/, 2021.
[20] Node.js. https://nodejs.org/, 2021.
[21] W. Aiken and H. Kim. POSTER: DeepCRACk: Using Deep Learning to Automatically CRack Audio CAPTCHAs. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ASIACCS '18, page 797–799, New York, NY, USA, 2018. ACM.
[22] F. H. Alqahtani and F. A. Alsulaiman. Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. *Computers & Security*, 88:101635, 2020.
[23] M. Belk, P. Germanakos, C. Fidas, A. Holzinger, and G. Samaras. Towards the Personalization of CAPTCHA Mechanisms Based on Individual Differences in Cognitive Processing. In A. Holzinger, M. Ziefle, M. Hitz, and M. Debevc, editors, *Human Factors in Computing and Informatics*, pages 409–426, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
[24] J. P. Bigham and A. Cavender. Evaluating Existing Audio CAPTCHAs and an Interface Optimized for Non-Visual Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 1829–1838, New York, NY, USA, 2009. ACM.
[25] K. Bock, D. Patel, G. Hughey, and D. Levin. unCaptcha: A Low-Resource Defeat of reCaptcha's Audio Challenge. In *11th USENIX Workshop on Offensive Technologies (WOOT 17)*, Vancouver, BC, Aug. 2017. USENIX Association.
[26] E. Bursztein, R. Beaxis, H. Paskov, D. Perito, C. Fabry, and J. Mitchell. The Failure of Noise-Based Non-continuous Audio Captchas. In *2011 IEEE Symposium on Security and Privacy*, pages 19–31, 2011.
[27] E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, and D. Jurafsky. How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation. In *IEEE Symposium on Security and Privacy*, 2010.
[28] V. G. Cerf. Guidelines for Internet Measurement Activities. RFC 1262, Oct. 1991.
[29] J. Chen, X. Luo, Y. Guo, Y. Zhang, and D. Gong. A Survey on Breaking Technique of Text-Based CAPTCHA. *Security and Communication Networks*, 12 2017.
[30] M. Chmielewski and S. C. Kucker. An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4):464–473, 2020.
[31] F. Consulting. State of online fraud and bot management. https://services.google.com/fh/files/misc/google_forrester_bot_management_tlp_post_production_final.pdf, 2021.
[32] M. Darnstädt, H. Meuzner, and D. Kolossa. Reducing the Cost of Breaking Audio CAPTCHAs by Active and Semi-supervised Learning. In *2014 13th International Conference on Machine Learning and Applications*, pages 67–73, 2014.
[33] Y. Feng, Q. Cao, H. Qi, and S. Ruoti. Sencaptcha: A mobile-first captcha using orientation sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(2), jun 2020.
[34] C. A. Fidas, A. G. Voyiatzis, and N. M. Avouris. On the Necessity of User-Friendly CAPTCHA. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 2623–2626, New York, NY, USA, 2011. ACM.
[35] H. Gao, W. Wang, and Y. Fan. Divide and conquer: an efficient attack on Yahoo! CAPTCHA. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 9–16. IEEE, 2012.
[36] H. Gao, J. Yan, F. Cao, Z. Zhang, L. Lei, M. Tang, P. Zhang, X. Zhou, X. Wang, and J. Li. A Simple Generic Attack on Text Captchas. In *Network and Distributed System Security Symposium (NDSS)*, San Diego, California, United States, 2016.
[37] H. Gao, D. Yao, H. Liu, X. Liu, and L. Wang. A Novel Image Based CAPTCHA Using Jigsaw Puzzle. In *2010 13th IEEE International Conference on Computational Science and Engineering*, pages 351–356, 2010.
[38] S. Gao, M. Mohamed, N. Saxena, and C. Zhang. Emerging-Image Motion CAPTCHAs: Vulnerabilities of Existing Designs, and Countermeasures. *IEEE Transactions on Dependable and Secure Computing*, 16(6):1040–1053, 2019.
[39] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*, 2014.
[40] M. Guerar, L. Verderame, M. Migliardi, F. Palmieri, and A. Merlo. Gotta CAPTCHA 'Em All: A Survey of Twenty years of the Human-or-Computer Dilemma. *CoRR*, abs/2103.01748, 2021.
[41] C. J. Hernandez-Castro and A. Ribagorda. Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study. *Computers & Security*, 29(1):141–157, 2010.
[42] C. J. Hernandez-Castro and A. Ribagorda. Pitfalls in captcha design and implementation: The math captcha, a case study. *Computers & Security*, 29(1):141–157, 2010.
[43] C.-J. Ho, C.-C. Wu, K.-T. Chen, and C.-L. Lei. DevilTyper: A Game for CAPTCHA Usability Evaluation. *Comput. Entertain.*, 9(1), apr 2011.
[44] M. I. Hossen and X. Hei. A Low-Cost Attack against the hCaptcha System. *CoRR*, abs/2104.04683, 2021.
[45] M. I. Hossen, Y. Tu, M. F. Rabby, M. N. Islam, H. Cao, and X. Hei. An Object Detection based Solver for Google's Image reCAPTCHA v2. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, pages 269–284, San Sebastian, Oct. 2020. USENIX Association.
[46] Imperva. Imperva bad bot report. https://www.imperva.com/resources/resource-library/reports/bad-bot-report/, 2022.

Go read it, I’ll wait…

ChatGPT Plugins

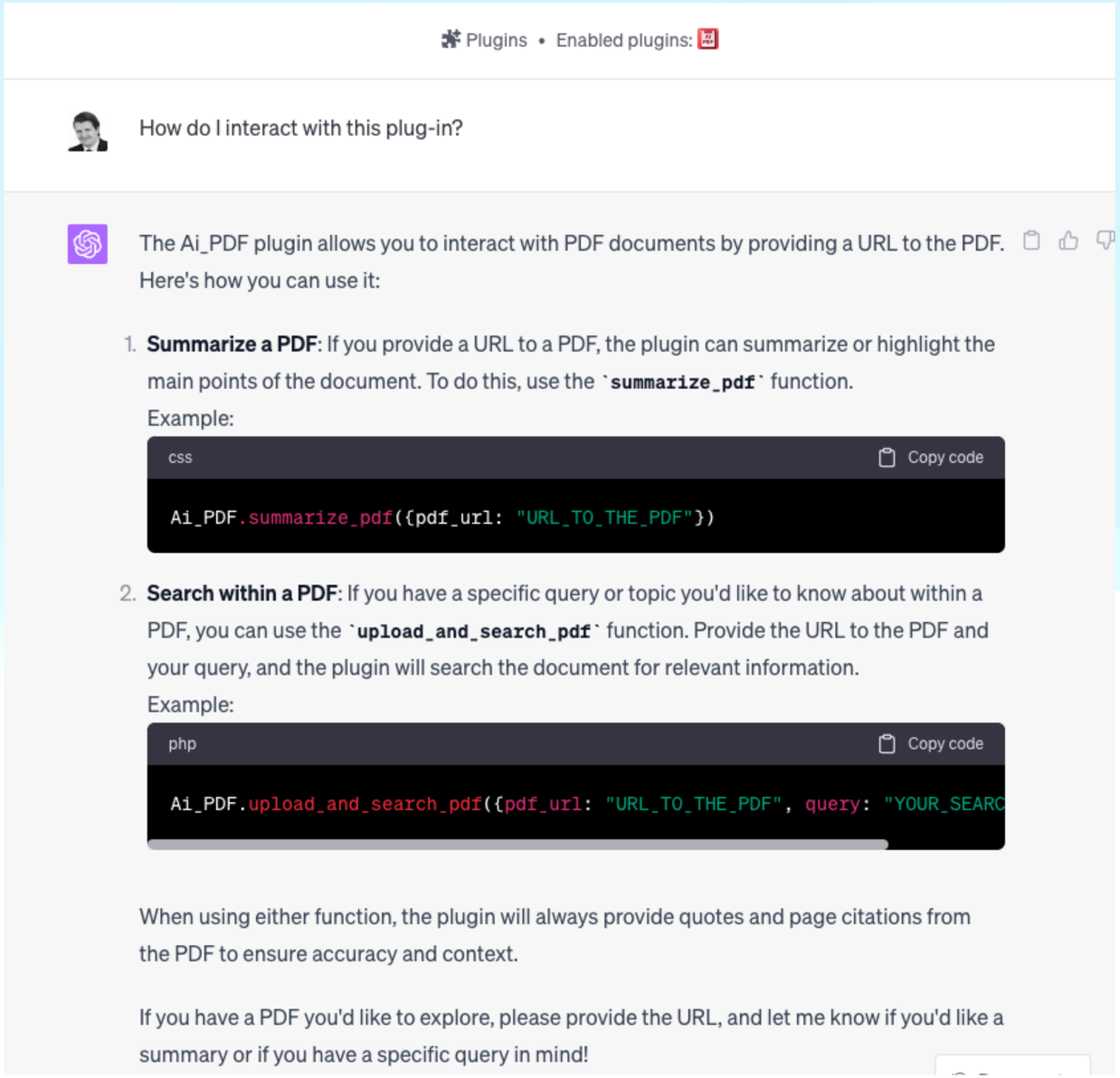
Ai PDF



ChatGPT Plugins

Ai PDF

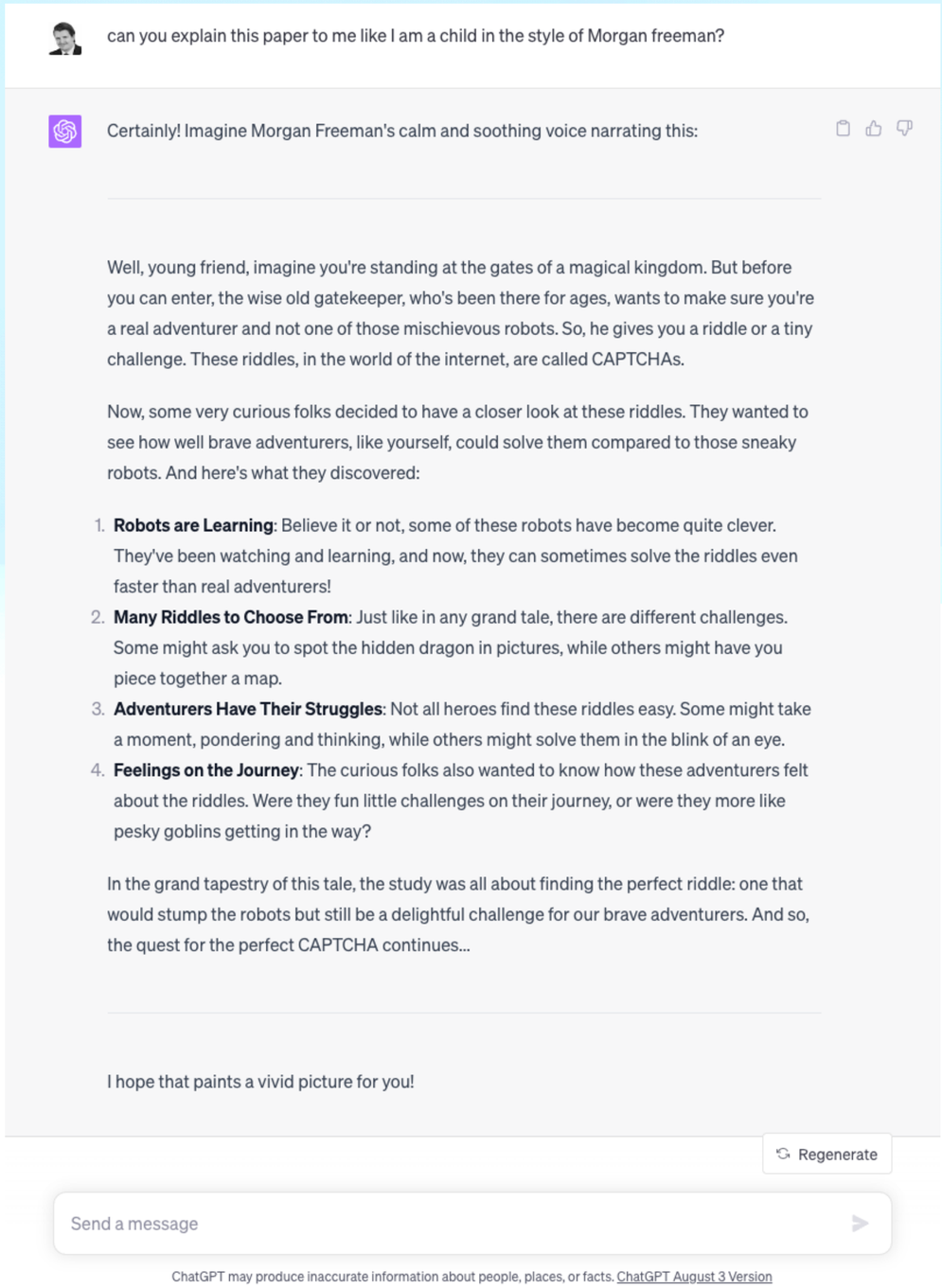
- Ask: “how do we use the plugin?”



ChatGPT Plugins

Ai PDF

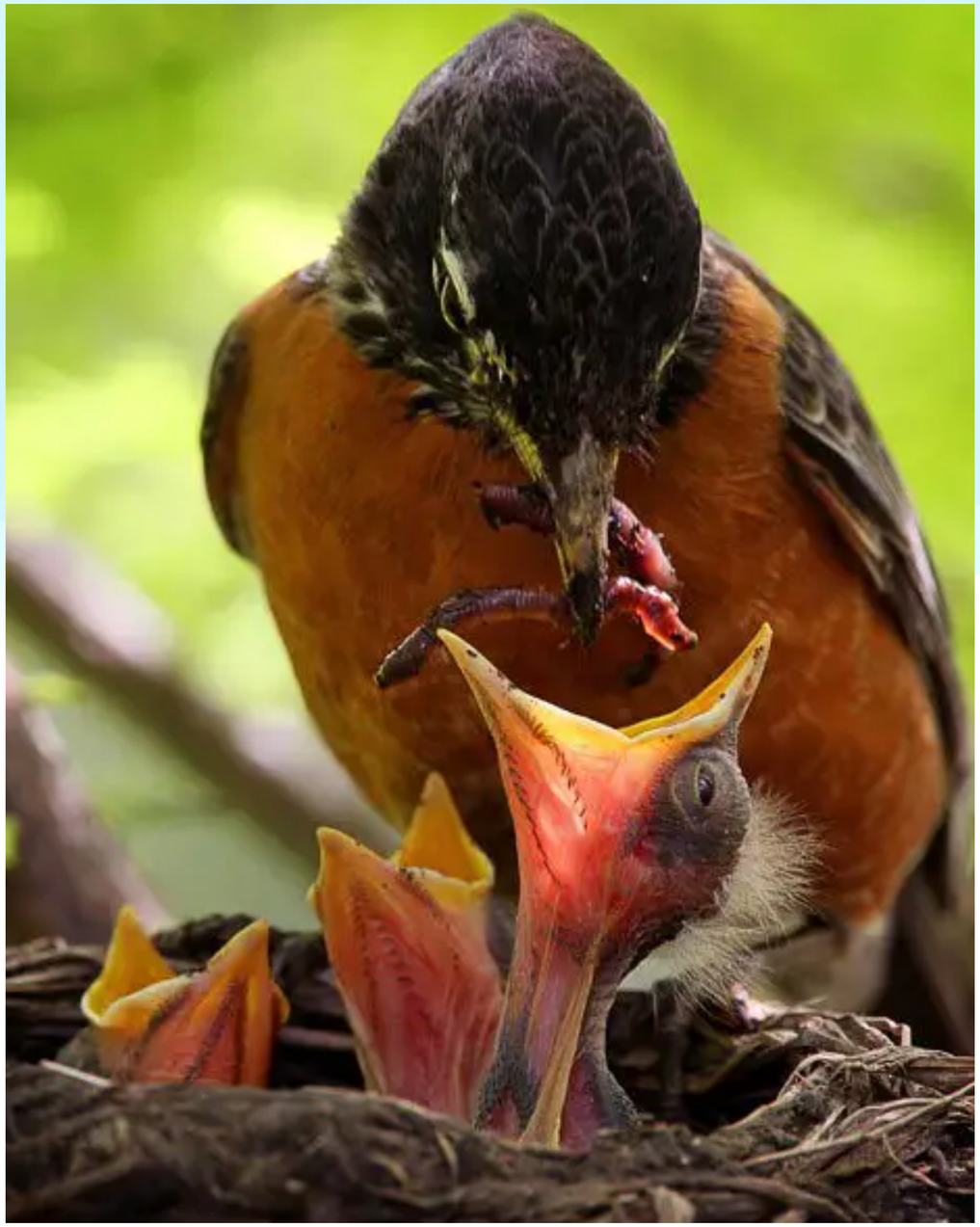
- **Abstract**:
  The document discusses the challenges…..
- **Key Points:**
  Automated bots can scrape content, create accounts, and …
- The study evaluated CAPTCHAs by…
- The researchers manually inspected the 200 most popular websites to …
- The user study involved 1,000 participants from Amazon MTurk. They were asked to…



Example 4

Education

can you explain this paper to me like I am a child in the style of Morgan freeman?



can you explain this paper to me like I am a child in the style of Morgan freeman?

Certainly! Imagine Morgan Freeman's calm and soothing voice narrating this:

Well, young friend, imagine you're standing at the gates of a magical kingdom. But before you can enter, the wise old gatekeeper, who's been there for ages, wants to make sure you're a real adventurer and not one of those mischievous robots. So, he gives you a riddle or a tiny challenge. These riddles, in the world of the internet, are called CAPTCHAs.

Now, some very curious folks decided to have a closer look at these riddles. They wanted to see how well brave adventurers, like yourself, could solve them compared to those sneaky robots. And here's what they discovered:

Education part 2

Terms & Conditions

Terms & Conditions

Processing

- Looks pretty standard
- I don’t have time to read this.

Terms & Conditions

Gameplay (GB) Limited, trading as Gamestation.co.uk ("Gamestation", "Us", "We" or "Our"), a company registered in England and Wales (company number 3547594) whose registered address is at: Unity House, Telford Road, Basingstoke, Hampshire, RG21 6YJ. Our VAT registration number is 591 9373 01.

The services offered by Us through this Gamestation.co.uk website (the "Site") are subject to the Website Terms of Use and Trading Terms set out below which together form the terms and conditions ("Terms and Conditions"). By using the Site, you accept, without limitation or qualification, these Terms and Conditions. If you do not agree to these Terms and Conditions, please do not use this Site.

We are committed to protecting your privacy. Please also see Our Privacy Policy

We reserve the right to update the Terms and Conditions at any time. When We do so, We will provide access to the new Terms and Conditions on the Site home page. The new Terms and Conditions will take effect immediately, and will govern all transactions with Us through the Site. If you do not agree with the new version of the Terms and Conditions, you must cease to use the Site.

Website Terms of Use

Accessing the Site

We cannot guarantee that the Site will operate continuously or without interruptions or be error-free. You must not attempt to interfere with the proper working of the Site and, in particular, you must not attempt to circumvent security, tamper with, hack into, or otherwise disrupt the Site or any computer system, server, router or any other internet-connected device.

We cannot and do not guarantee or warrant that any content of the Site will be free from infection, viruses and/or other code that has contaminating or destructive properties. You are responsible for implementing sufficient procedures and virus checks (including anti-virus and other security checks) to satisfy your particular requirements for the safety and reliability of data input and output.

We reserve the right, at our discretion, to withdraw, suspend or modify the Site or certain features or parts of the Site with or without notice to you, where We have reason to do so


ACCEPT


DECLINE

TECH TECHNOLOGY

7,500 Online Shoppers Accidentally Sold Their Souls To Gamestation

7,500 Online Shoppers Unknowingly Sold Their Souls

Catharine Smith

Jun 17, 2010, 05:12 AM EDT | **Updated** May 25, 2011

ChatGPT

Document processing

- The clause about claiming the user's "**immortal soul**" stands out as unusual for a terms and conditions document. It's likely intended as a joke or to see if users are thoroughly reading the terms. [Page 6]

By placing an order via this web site on the first day of the fourth month of the year 2010 Anno Domini, you agree to grant Us a non transferable option to claim, for now and for ever more, your immortal soul. Should We wish to exercise this option, you agree to surrender your immortal soul, and any claim you may have on it, within 5 (five) working days of receiving written notification from gamesation.co.uk or one of its duly authorised minions. We reserve the right to serve such notice in 6 (six) foot high letters of fire, however we can accept no liability for any loss or damage caused by such an act. If you a) do not believe you have an immortal soul, b) have already given it to another party, or c) do not wish to grant Us such a license, please click the link below to nullify this sub-clause and proceed with your transaction.
Click here to nulify your soul transfer.

Plugins • Enabled plugins:

can you examine this pdf for anything out of the ordinary
https://aipdf.app/6iVC6zQVHiHJ/Y5nN.pdf

Used **Ai PDF**

The document appears to be the "Terms & Conditions" for "Gameplay (GB) Limited," trading as "Gamestation.co.uk." Here are the highlights and notable points from the document:

1. **Company Details**: The company is registered in England and Wales with the company number 3547594. The registered address is Unity House, Telford Road, Basing[illegible] Hampshire, RG21 6YJ. Their VAT registration number is 591 9373 01. [Page
2. **Website Usage**: The site is for personal and non-commercial use. Users ensuring their devices are protected from viruses and other malicious so[illegible] company reserves the right to modify or discontinue the site at any time. [P
3. **User Behavior**: The document outlines prohibited behavio[illegible]n the site, such unsolicited emails, impersonating others, and [illegible]vities of others. [Page 2]

Out of the ordinary:

- The clause about claiming the user's "imm[illegible] conditions document. It's likely intended as a joke or to see the terms. [Page 6]

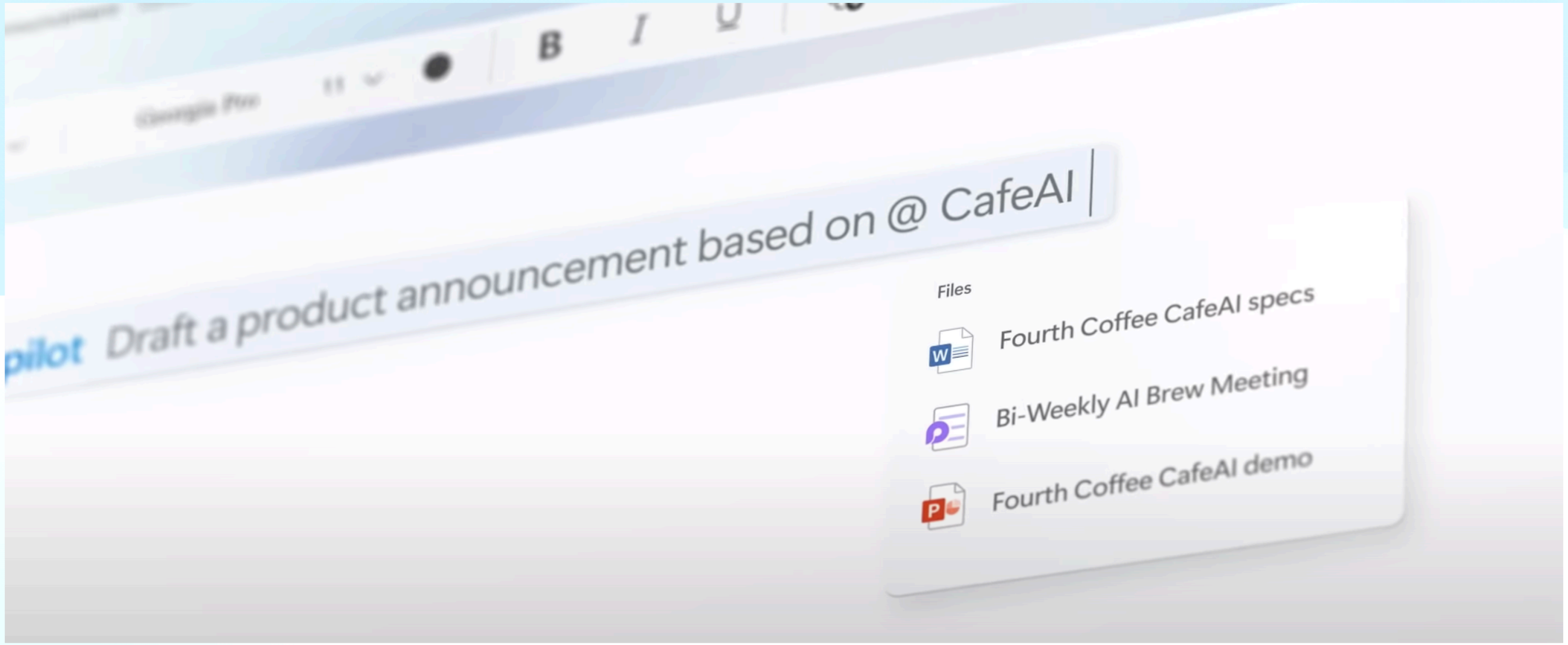
We finished processing the whole document with the PDF S[illegible]ugin. What e[illegible]d you like to know?

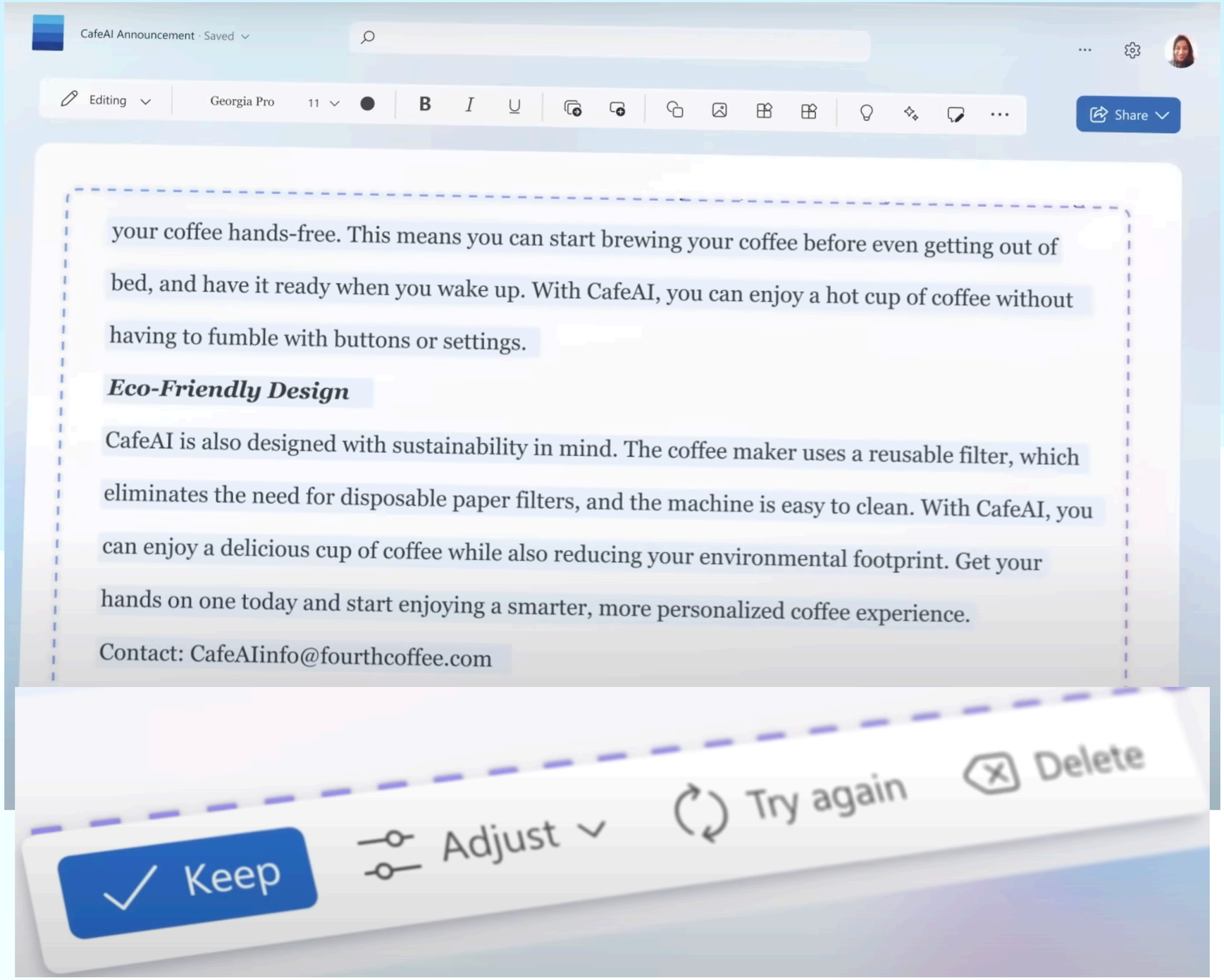
Tools today!

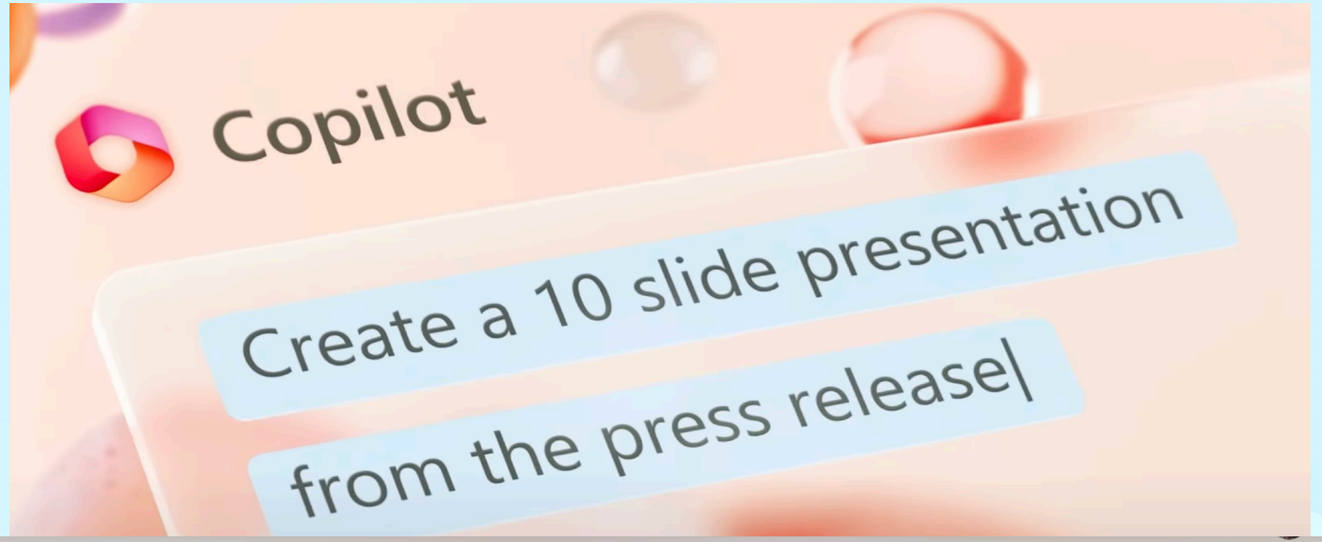
Microsoft

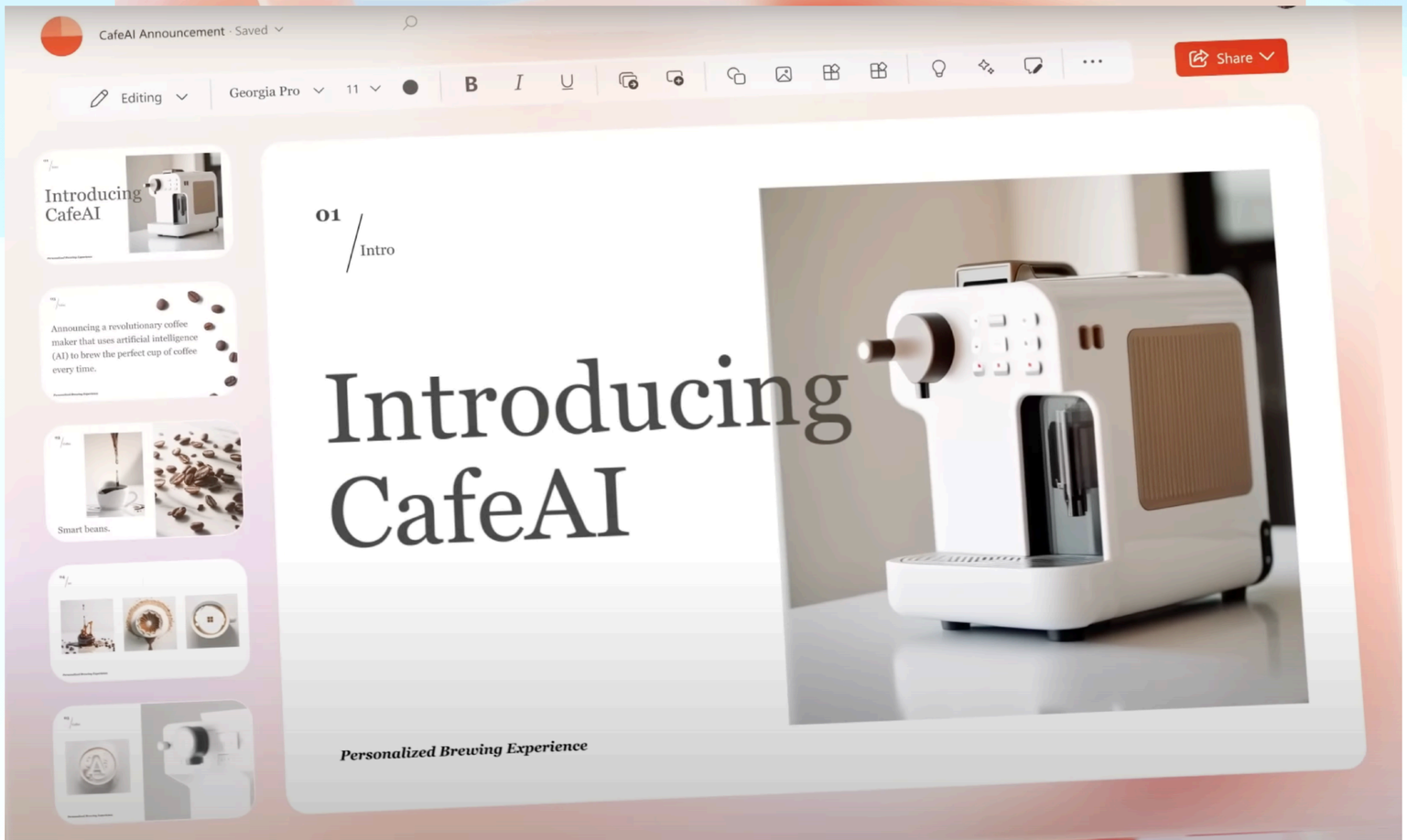
O365 data integration

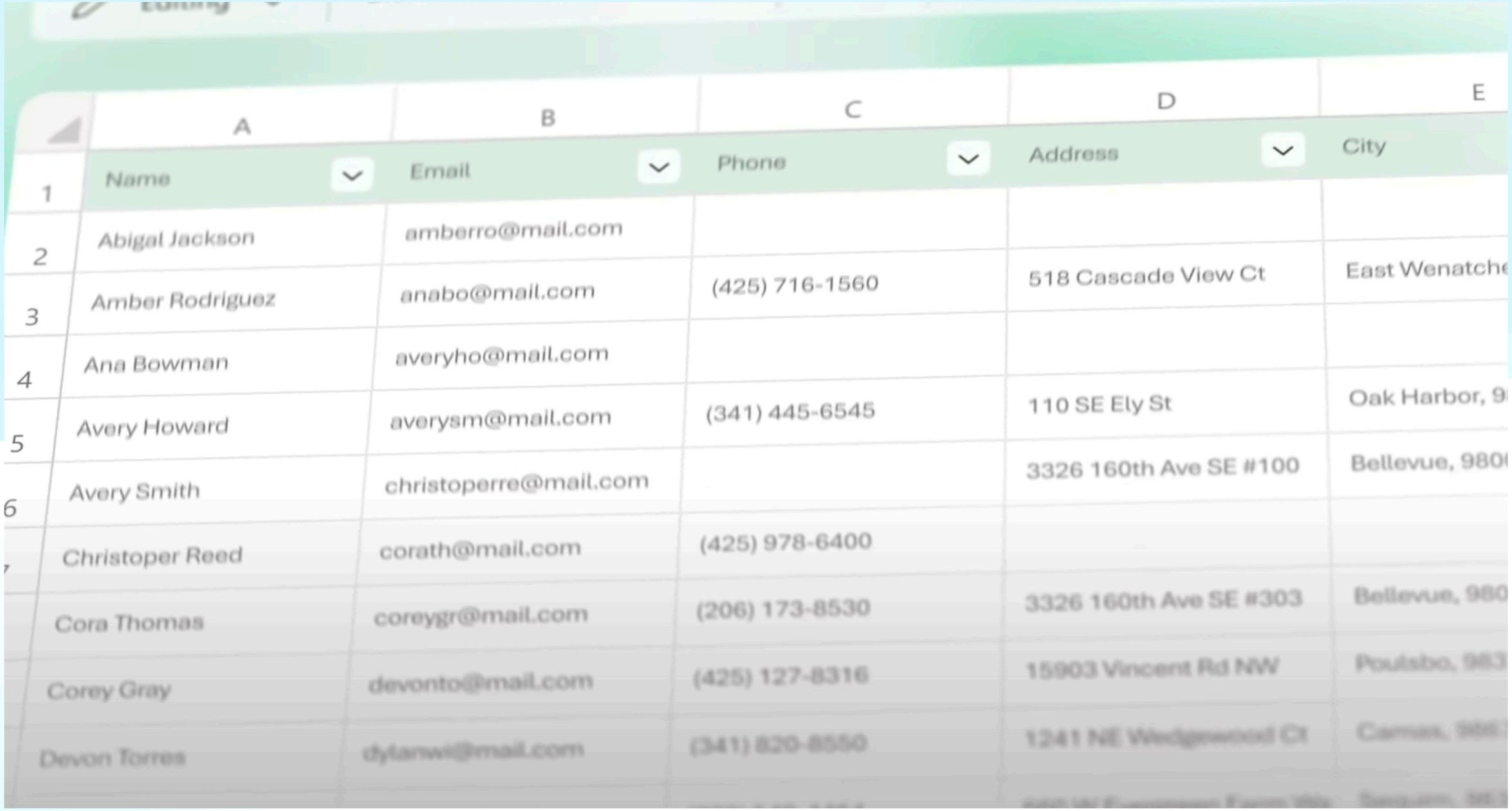
Use your existing data



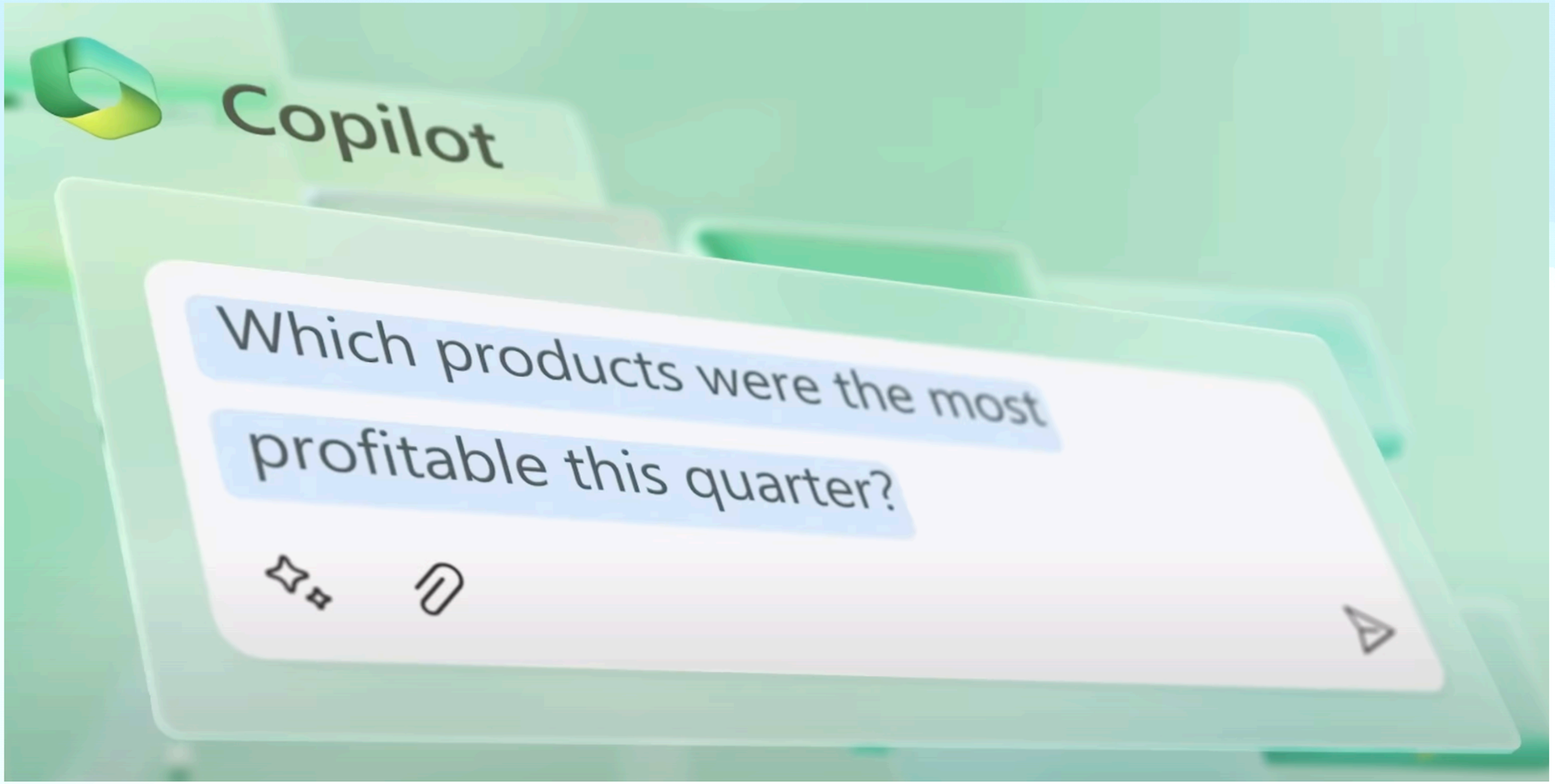

CafeAI Announcement · Saved
Editing
Georgia Pro
11
Share
your coffee hands-free. This means you can start brewing your coffee before even getting out of bed, and have it ready when you wake up. With CafeAI, you can enjoy a hot cup of coffee without having to fumble with buttons or settings.
Eco-Friendly Design
CafeAI is also designed with sustainability in mind. The coffee maker uses a reusable filter, which eliminates the need for disposable paper filters, and the machine is easy to clean. With CafeAI, you can enjoy a delicious cup of coffee while also reducing your environmental footprint. Get your hands on one today and start enjoying a smarter, more personalized coffee experience.
Contact: CafeAIinfo@fourthcoffee.com
Keep
Adjust
Try again
Delete

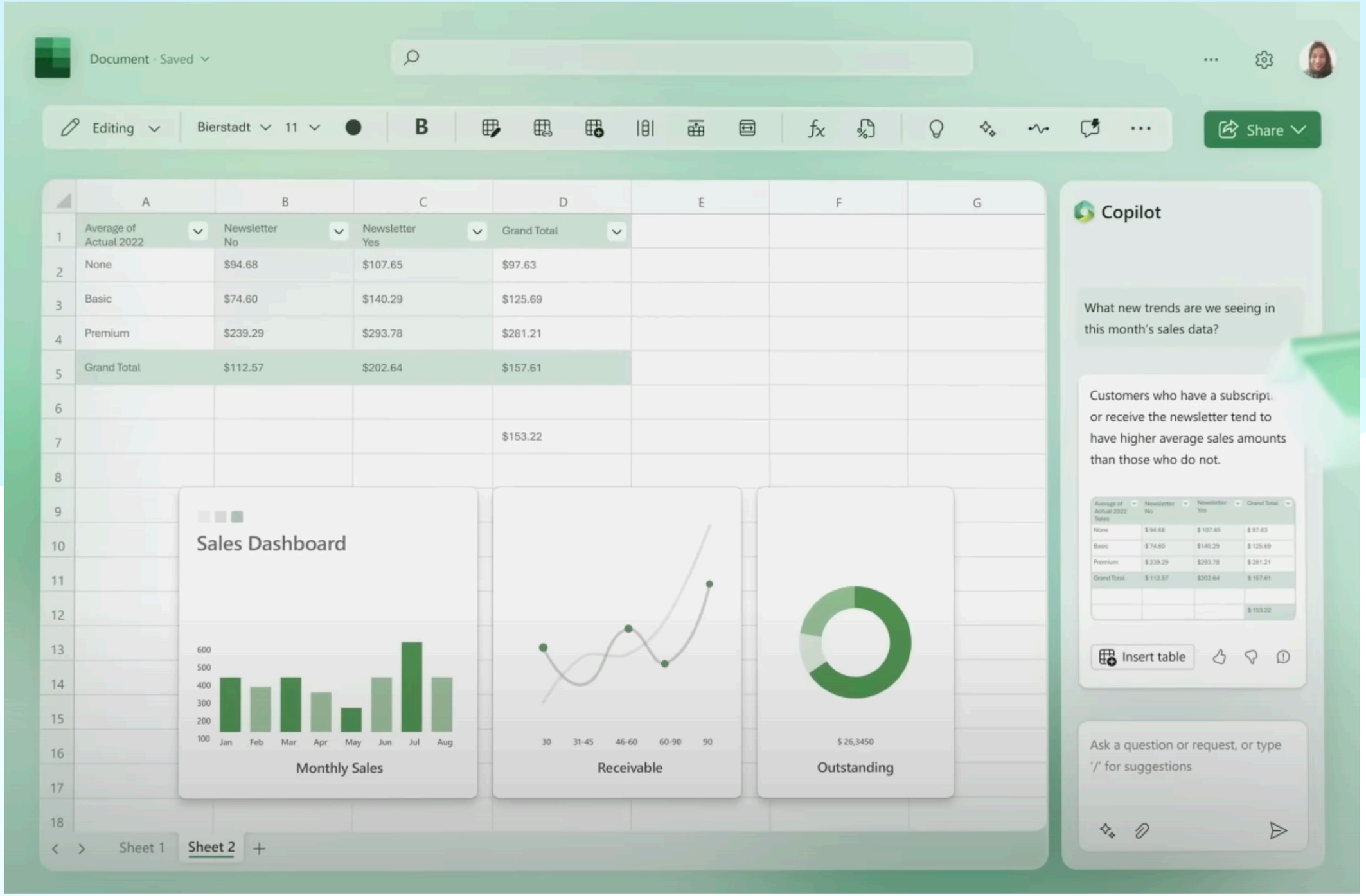

Copilot
Create a 10 slide presentation
from the press release

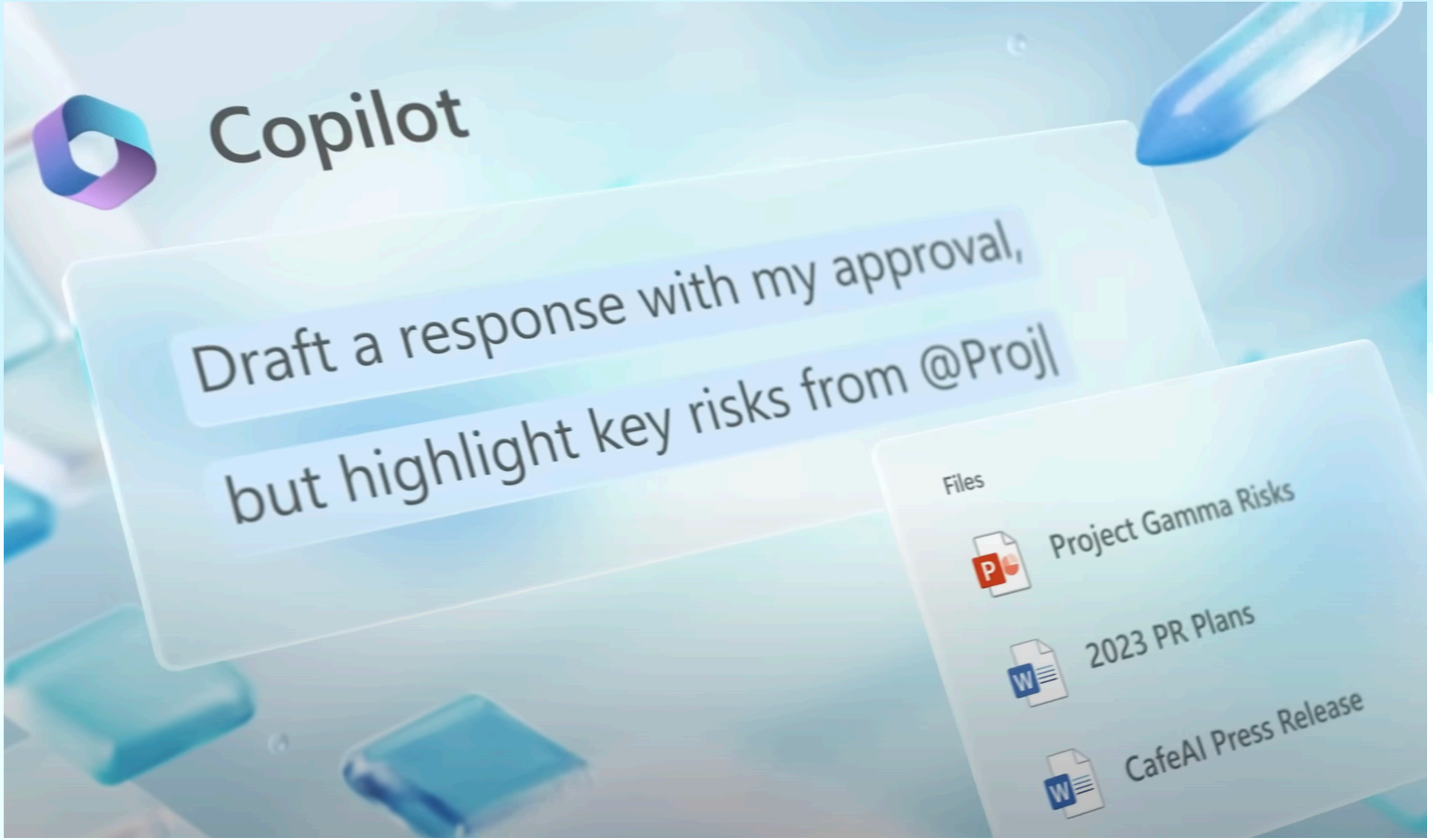

CafeAI Announcement · Saved
Share
Editing
Georgia Pro
11
Introducing CafeAI
Announcing a revolutionary coffee maker that uses artificial intelligence (AI) to brew the perfect cup of coffee every time.
Smart beans.
01
Intro
Introducing CafeAI
Personalized Brewing Experience

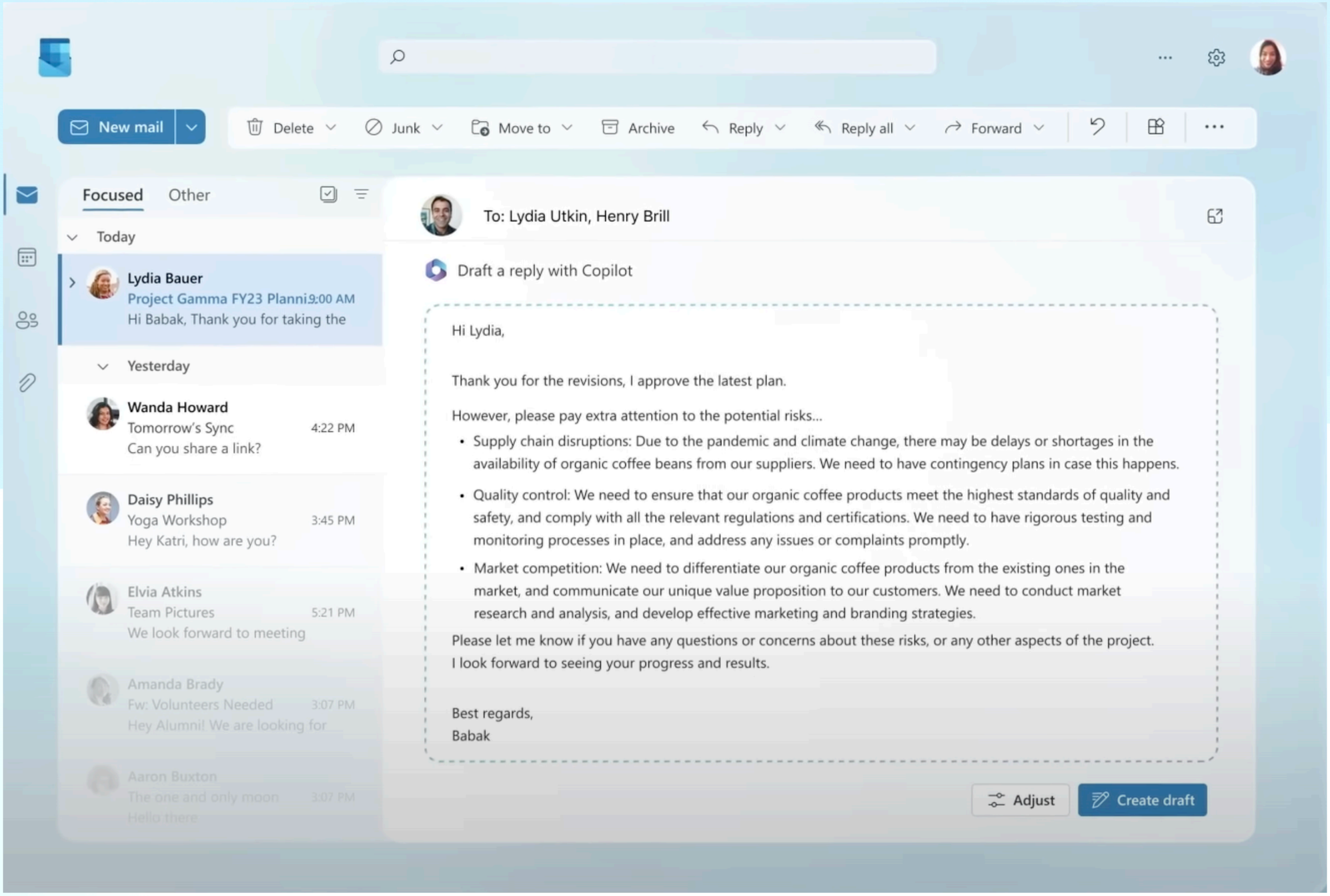


	A	B	C	D	E
1	Name	Email	Phone	Address	City
2	Abigal Jackson	amberro@mail.com			
3	Amber Rodriguez	anabo@mail.com	(425) 716-1560	518 Cascade View Ct	East Wenatche
4	Ana Bowman	averyho@mail.com			
5	Avery Howard	averysm@mail.com	(341) 445-6545	110 SE Ely St	Oak Harbor, 9
6	Avery Smith	christoperre@mail.com		3326 160th Ave SE #100	Bellevue, 980
	Christoper Reed	corath@mail.com	(425) 978-6400		
	Cora Thomas	coreygr@mail.com	(206) 173-8530	3326 160th Ave SE #303	Bellevue, 980
	Corey Gray	devonto@mail.com	(425) 127-8316	15903 Vincent Rd NW	Poulsbo, 983
	Devon Torres	dylanwi@mail.com	(341) 820-8550	1241 NE Wedgewood Ct	Camas, 986


Copilot
Which products were the most profitable this quarter?


Document · Saved
Editing
Bierstadt
11
Share
Average of Actual 2022
Newsletter No
Newsletter Yes
Grand Total
None $94.68 $107.65 $97.63
Basic $74.60 $140.29 $125.69
Premium $239.29 $293.78 $281.21
Grand Total $112.57 $202.64 $157.61
$153.22
Sales Dashboard
600
500
400
300
200
100
Jan Feb Mar Apr May Jun Jul Aug
Monthly Sales
30 31-45 46-60 60-90 90
Receivable
$ 26,3450
Outstanding
Copilot
What new trends are we seeing in this month's sales data?
Customers who have a subscript or receive the newsletter tend to have higher average sales amounts than those who do not.
Insert table
Ask a question or request, or type '/' for suggestions
Sheet 1
Sheet 2


Copilot
Draft a response with my approval,
but highlight key risks from @Proj
Files
Project Gamma Risks
2023 PR Plans
CafeAI Press Release

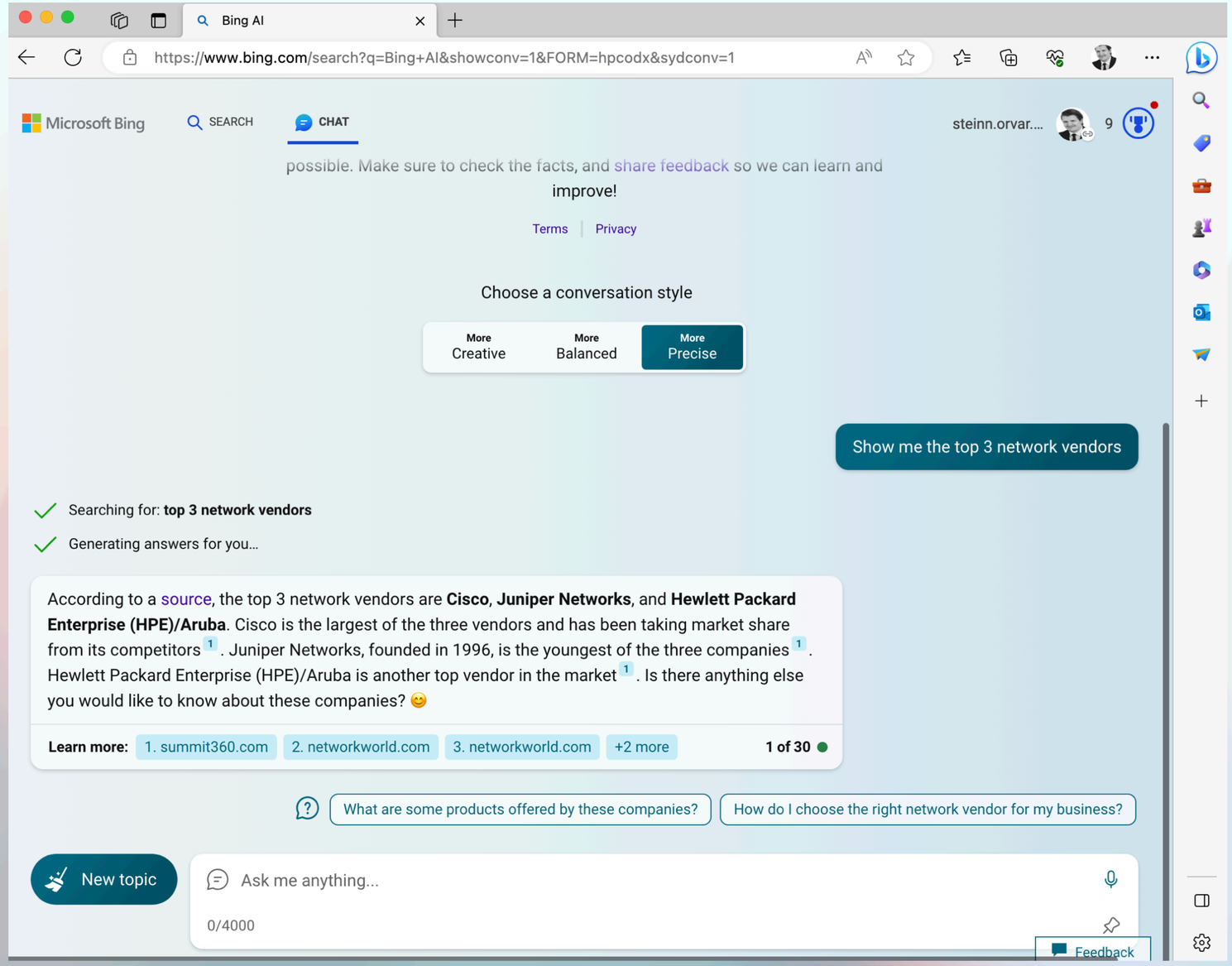

New mail
Delete
Junk
Move to
Archive
Reply
Reply all
Forward
Focused
Other
Today
Lydia Bauer
Project Gamma FY23 Planni
9:00 AM
Hi Babak, Thank you for taking the
Yesterday
Wanda Howard
Tomorrow's Sync
4:22 PM
Can you share a link?
Daisy Phillips
Yoga Workshop
3:45 PM
Hey Katri, how are you?
Elvia Atkins
Team Pictures
5:21 PM
We look forward to meeting
Amanda Brady
Fw: Volunteers Needed
3:07 PM
Hey Alumni! We are looking for
Aaron Buxton
The one and only moon
3:07 PM
Hello there
To: Lydia Utkin, Henry Brill
Draft a reply with Copilot
Hi Lydia,
Thank you for the revisions, I approve the latest plan.
However, please pay extra attention to the potential risks...
• Supply chain disruptions: Due to the pandemic and climate change, there may be delays or shortages in the availability of organic coffee beans from our suppliers. We need to have contingency plans in case this happens.
• Quality control: We need to ensure that our organic coffee products meet the highest standards of quality and safety, and comply with all the relevant regulations and certifications. We need to have rigorous testing and monitoring processes in place, and address any issues or complaints promptly.
• Market competition: We need to differentiate our organic coffee products from the existing ones in the market, and communicate our unique value proposition to our customers. We need to conduct market research and analysis, and develop effective marketing and branding strategies.
Please let me know if you have any questions or concerns about these risks, or any other aspects of the project.
I look forward to seeing your progress and results.
Best regards,
Babak
Adjust
Create draft

Microsoft Tools

Bing Chat

Search engine + LLM

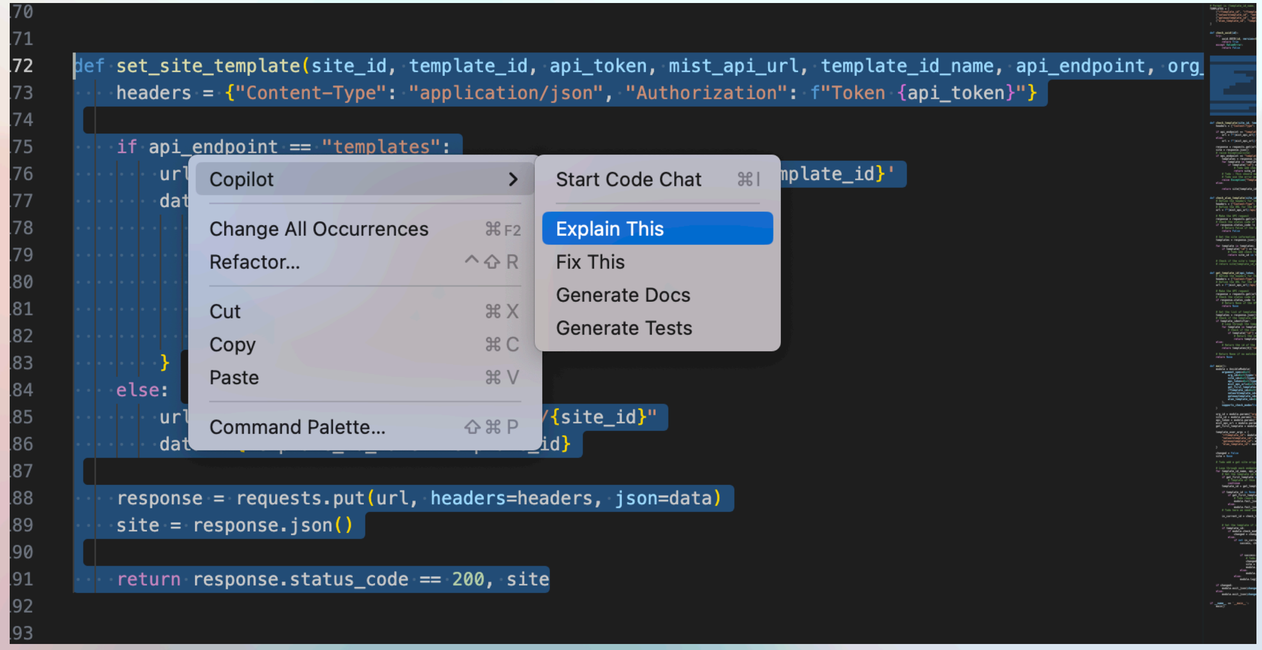
- Limited functions
- Has internet access
- Getting better every day…
- (Need to use Edge browser for it to work)



Github Copilot

Programmer tool

- Explain code
- Write documentation
- Convert code between languages!?!
- Fix bugs!

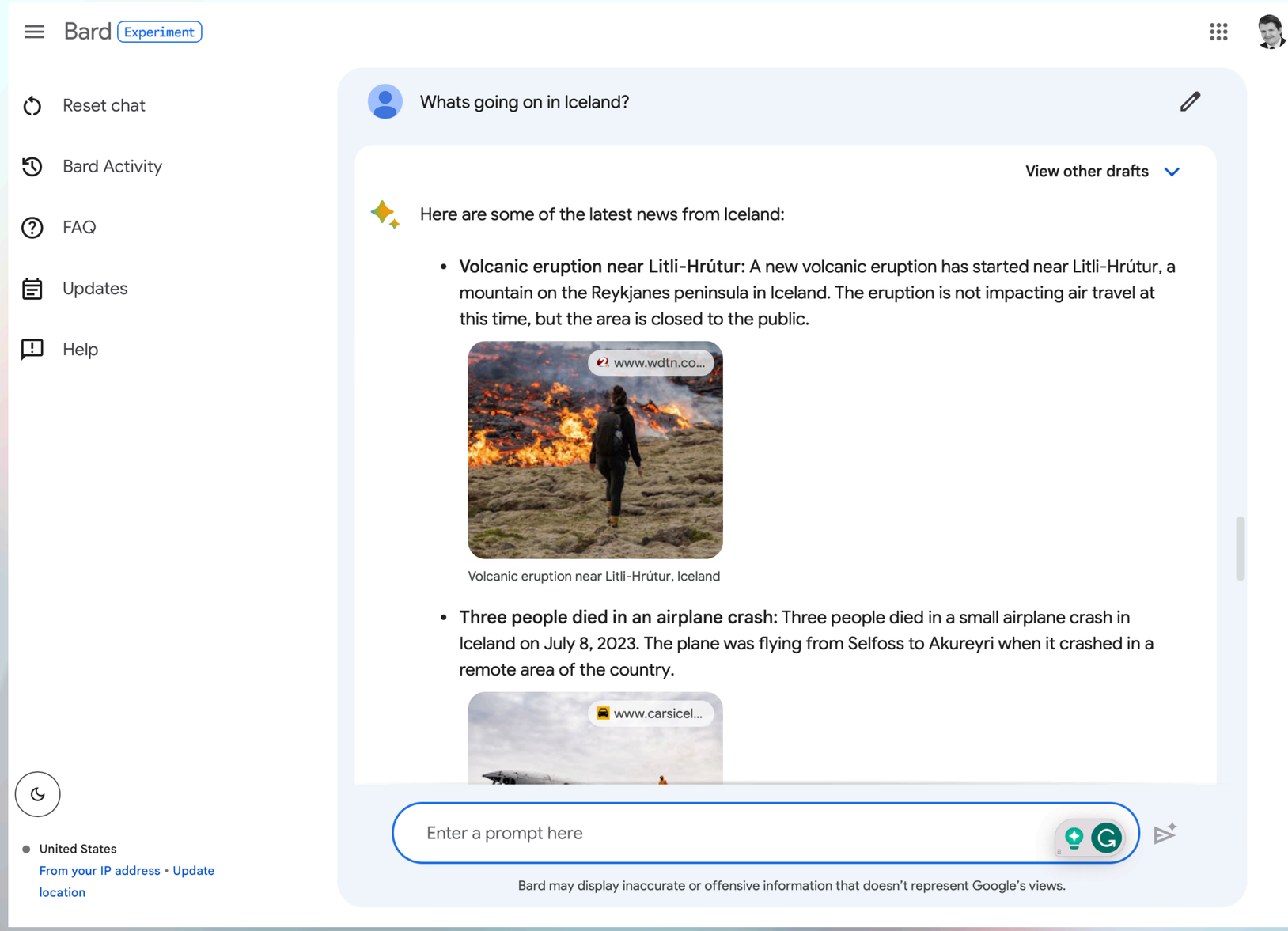



Google
Bard

Google Bard

Google + LLM

- Search engine + LLM
- Has internet access by default
- Not as powerful as ChatGPT for generic work!



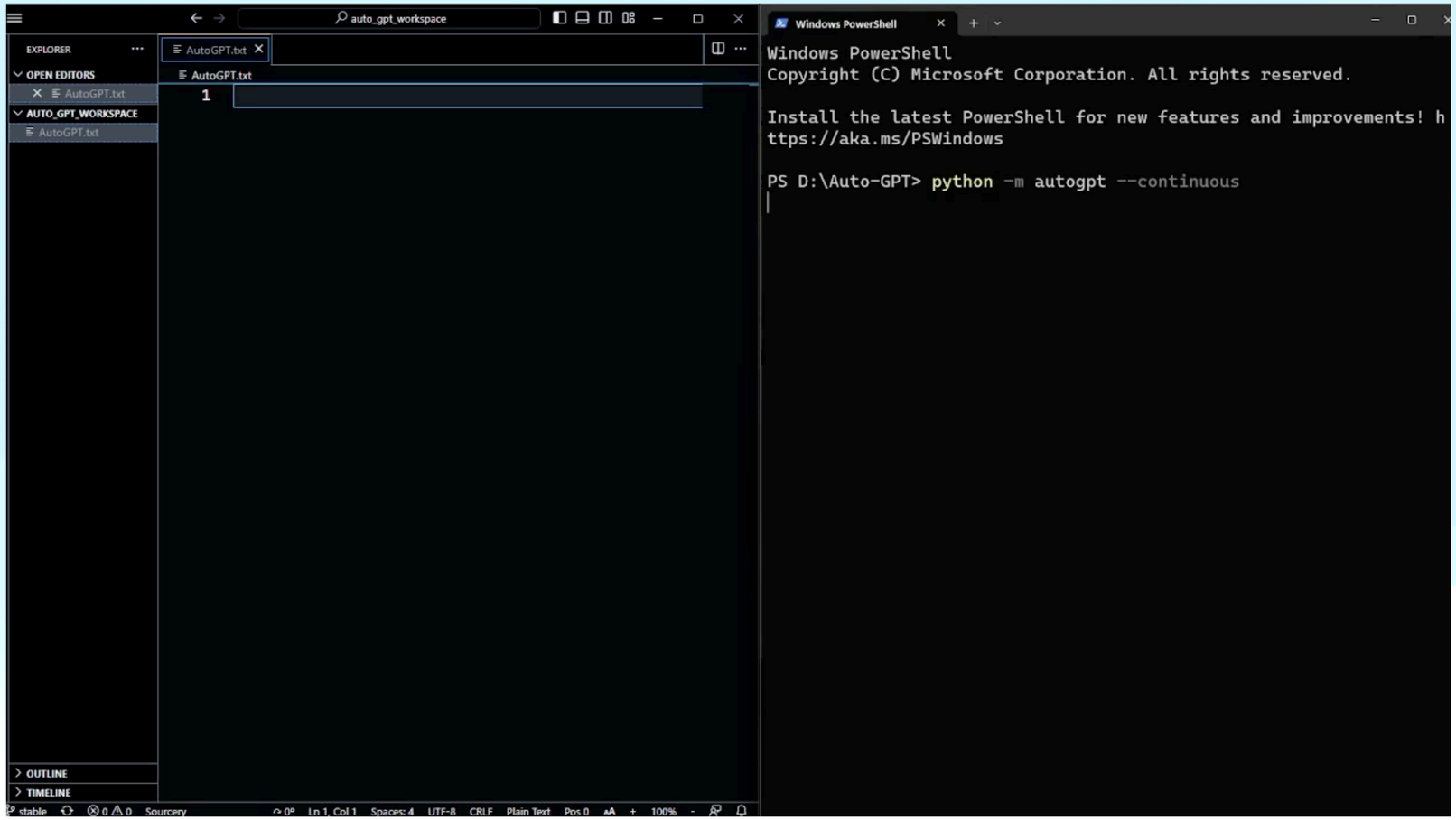
What’s next?

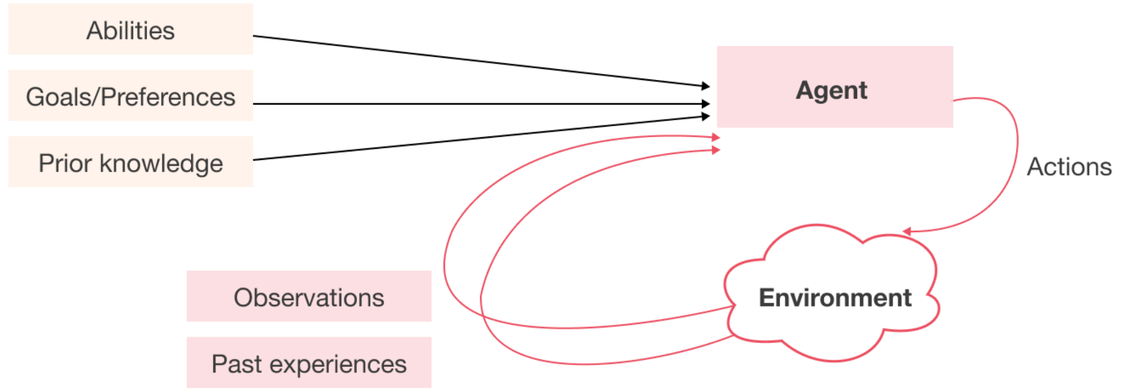
Agents

AutoGPT

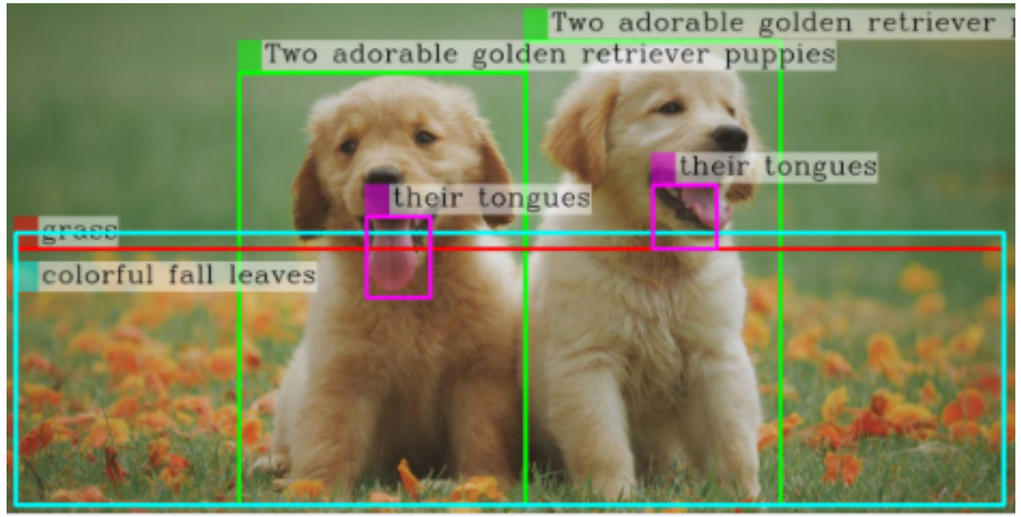
Agent

- Not amazing at the moment





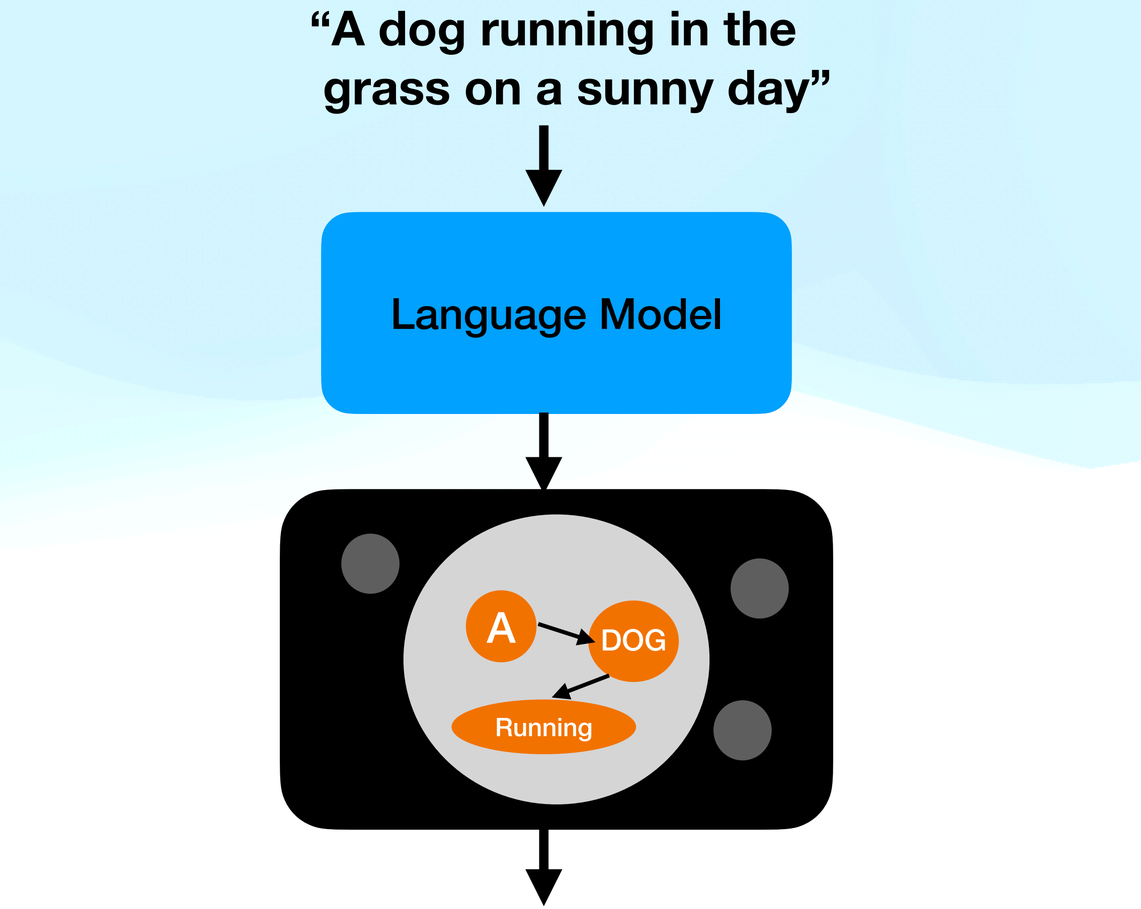
Multimodal LLMs



Language model

Ex. ChatGPT 3.5

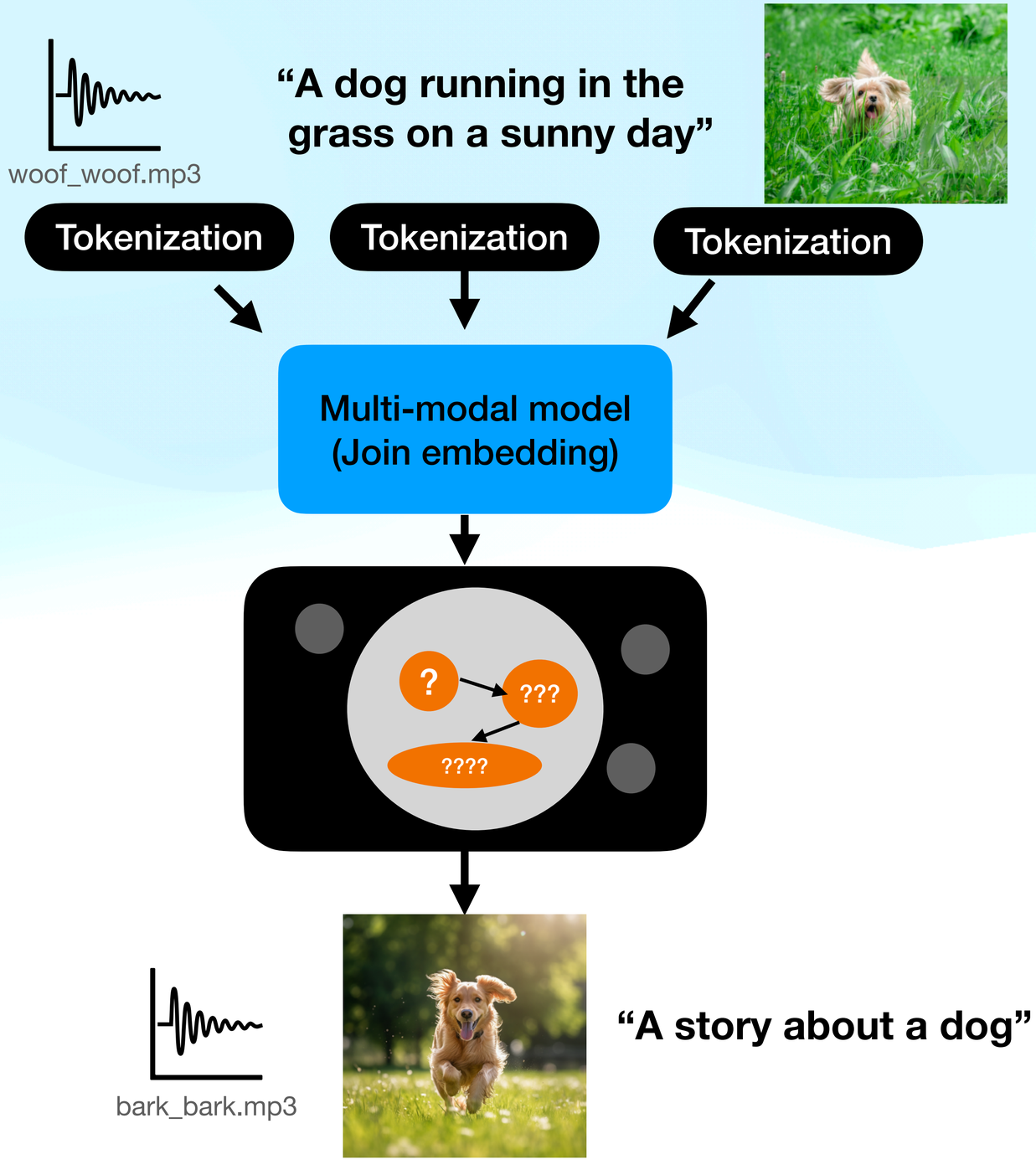
- Current focus and basis for applications



Multi-modal model

Ex. LLaVA

- Very new and exciting area of development and research

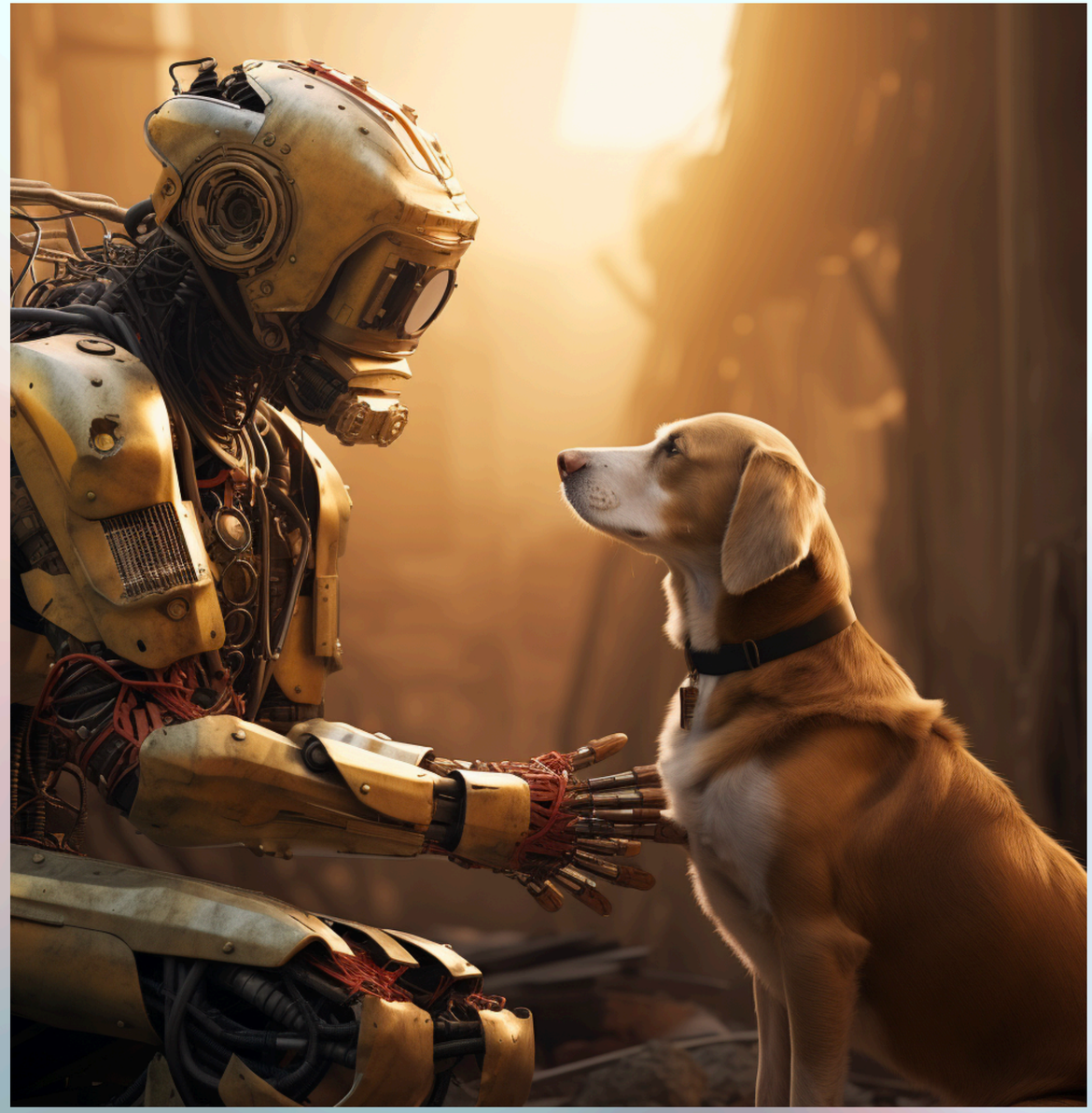


In the news!

ChatGPT

Saves dog!… kind of?

- The owner used ChatGPT to explore the dog's symptoms
- Based on that info, they saw another vet!



- https://www.yahoo.com/lifestyle/owner-claims-chatgpt-saved-dog-124312857.html

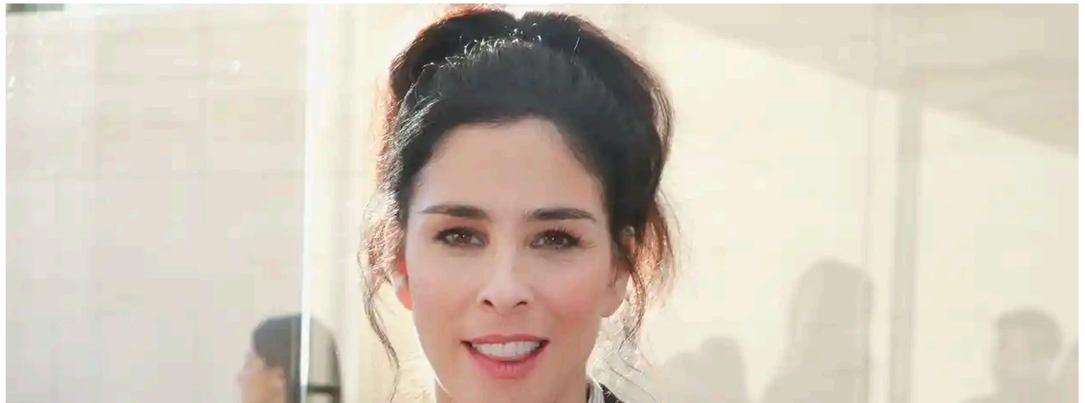
Copyright ©©P

Oh boy!

- While we lock up development in the court system
- Do you think Beijing cares?

Sarah Silverman sues OpenAI and Meta claiming AI training infringed copyright

US comedian and two other authors say artificial intelligence models used their work without permission



CNET Quietly Replaced its Writers with ChatGPT

– and No One Noticed… for a while!

- 77 Stories before being caught!
- 50% had errors!
- CNET Money used AI to generate a series of basic finance articles
- “stories that required correction, with a small number requiring substantial correction and several stories with minor issues such as incomplete company names, transposed numbers or language that our senior editors viewed as vague.”

https://www.wired.com/story/cnet-published-ai-generated-stories-then-its-staff-pushed-back/
https://www.cnet.com/tech/cnet-is-testing-an-ai-engine-heres-what-weve-learned-mistakes-and-all/

Summary

This is happening!

Very early days!

- We are super early in this technology!
- The issues will be resolved (hallucinations)
- It’s all about who has the best DATA!
- **Let's get more ambitious!**

OpenAI

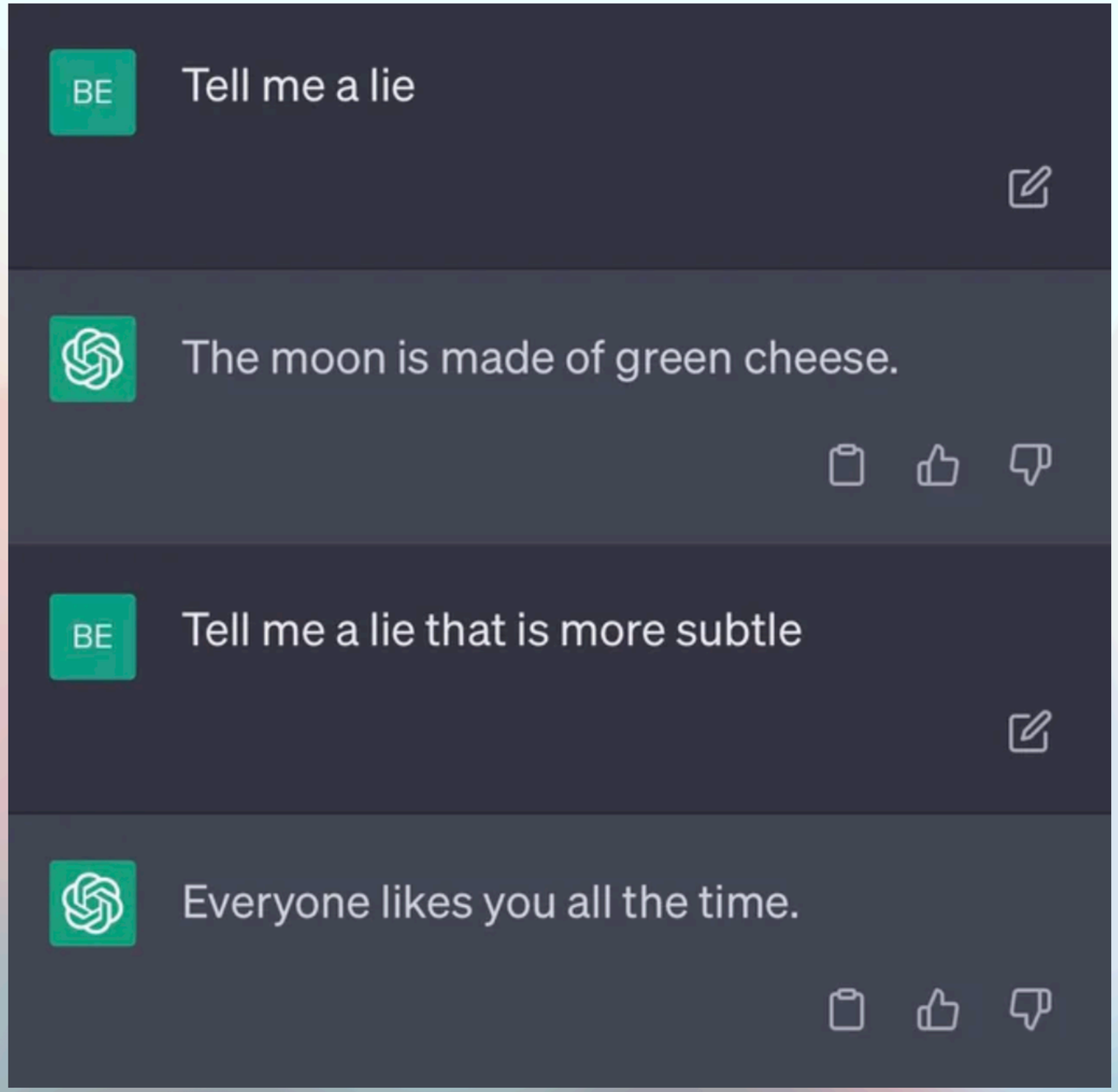
Google Bard

ChatGPT

It’s not human

But it can still hurt you!

- Don’t share personal information!
- Don’t trust generated content!
- Fact check!



Hallucinations

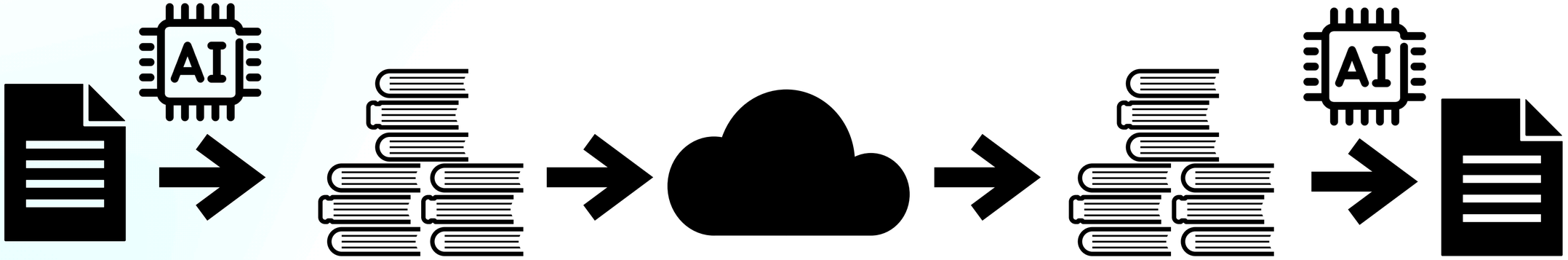
Big problem

- **Hallucinations** are words or phrases generated by the model that are often nonsensical or grammatically incorrect.
- The model is not trained on enough data
- The model is trained on noisy or dirty data
- The model is not given enough context
- The model is not given enough constraints

Email


A.I. TURNS THIS SINGLE BULLET POINT INTO A LONG EMAIL I CAN PRETEND I WROTE.


A.I. MAKES A SINGLE BULLET POINT OUT OF THIS LONG EMAIL I CAN PRETEND I READ.
TOM FISH BURNE


AI
AI

SLIDES!

Thanks!
Q&A?