





De Technosoof

The AI pause & Apple xR: To Pause or Not to Pause (ep.18)



In this episode of the Technosoof podcast, we discuss (in Dutch) the open letter signed by thousands of tech leaders and researchers urging a pause in the “out-of-control” artificial intelligence race.

The AI Pause Letter

The letter, published by the Future of Life Institute, warns that giant AI experiments such as GPT-4 pose serious risks to society and the environment, and calls for more transparency and accountability from the companies and organizations behind them.

Secret History of Elon Musk and Sam Altman

We also talk about the secret history of Elon Musk, Sam Altman and OpenAI, the research lab that was founded with the lofty goal of creating artificial general intelligence for the benefit of humanity. How did OpenAI go from a non-profit to a for-profit entity, and what are the implications of its controversial decision to license its technology to Microsoft? We explore the ethical dilemmas and power dynamics that shape the future of AI.

Technosoof.com



Word of the week

ChatGPT
Goldrush

● chatgpt
Zoekterm

● Amazon Web Serv...
Zoekterm

● Microsoft Azure
Zoekterm

● Google Cloud Pla...
Zoekterm



Wereldwijd ▼

11-01-2022 - 06-04-2023 ▼

Alle categorieën ▼


Google Zoeken ▼


Interesse In De Loop Der Tijd ?




App of The Week

Chat with any PDF

 [Join Discord](#)

 [Post to Twitter](#)

 [Share on Facebook](#)



Drop PDF here

[Browse my Computer](#)

[From URL](#) [Find a PDF](#)

Chats

(7) Dertiende bloemlezing over ChatGP ×

T_LinkedIn.pdf

Rachel Hayes is mentioned in the PDF as someone who has expressed concerns about the potential impact of ChatGPT on education...

3 uur geleden

WEF_Future_of_Jobs_2020.pdf ×

According to the Future of Jobs Report 2020, the top 15 skills for 2025 are: 1. Analytical thinking and innovation 2. Active...

21 uur geleden

2022-AI-Index-Report_Master.pdf ×

Chapter 5 of the AI Index Report 2022, titled "AI Policy and Governance," provides information on the number of AI-related...

1 dag geleden

For Students 🎓

Enhance your learning experience with ChatPDF. Comprehend textbooks, handouts, and presentations effortlessly. Don't spend hours flipping through research papers and academic articles.

Support your academic growth and succeed in your studies effectively and responsibly.

For Work 📁

Efficiently analyze your documents. From financial and sales reports to project and business proposals, training manuals, and legal contracts, ChatPDF can quickly provide you with the information you need.

Your data is kept confidential in a secure cloud storage and can be deleted at any time.

For Curious Minds 🧠

Unlock a wealth of knowledge with ChatPDF. Discover new insights and answers from historical documents, poetry, and literature, effortlessly.

ChatPDF can understand any language and reply in your preferred one. Satisfy your curiosity and expand your horizons with the tool that can answer any question from any PDF.

+ New Chat
Drop PDF here

(7) Dertiende bloemlezing ov...

WEF_Future_of_Jobs_2020.pdf

2022-AI-Index-Report_Maste...

Reset Chat

My Account

Home API Twitter

Chat with WEF_Future_of_Jobs_2020.pdf

Hello there! Welcome to the Future of Jobs Report 2020. This PDF file is packed with valuable insights and information about the current and future state of the labor market.

Here are three example questions you might have about this report:

- ▶ What are some of the emerging and declining jobs and skills mentioned in Chapter 2?
- ▶ How do public and private sector pathways differ in their approach to reviving labor markets, as discussed in Chapter 3?
- ▶ Can you provide more information about the methodology used to create this report, as outlined in Appendix A?

What are some of the emerging and declining jobs and skills mentioned in Chapter 2?

Chapter 2 of the Future of Jobs Report 2020 discusses emerging and declining jobs and skills. According to the report, some of the top skills and skill groups that employers see as rising in prominence in the lead up to 2025 include analytical thinking and innovation, active learning and learning strategies, creativity, originality, and initiative, technology use, monitoring and control, complex problem-solving, leadership and social influence, emotional intelligence, reasoning, problem-solving.

On the other hand, some of the declining skills include manual dexterity, endurance, precision workmanship or assembly line work. The report also highlights that certain jobs are expected to decline in demand while others will increase. For example, jobs in data entry and administrative support are expected to decline while jobs in artificial intelligence (AI) and machine learning specialists are expected to increase. You can find more information on this topic on page 35 of the report.

what is the expected impact of AI technologies on labor market and skills development

The Future of Jobs Report 2020 discusses the expected impact of AI technologies on the labor market and skills development. According to the report, AI technologies are expected to have a significant impact on the labor market, with some jobs becoming obsolete while others will be created. The report suggests

Ask any question...



what is show on page 36 in figure 27

On page 36 of the Future of Jobs Report 2020, Figure 27 shows the top skills and skill groups which employers see as rising in prominence in the lead up to 2025. The figure is a bar chart that displays the percentage of employers who identified each skill or skill group as being important for their employees to have in the future.

The top three skills or skill groups identified are analytical thinking and innovation, active learning and learning strategies, and complex problem-solving. Other important skills and skill groups include creativity, originality, and initiative; technology use, monitoring and control; leadership and social influence; emotional intelligence; reasoning, problem-solving; systems analysis and evaluation; and resilience, stress tolerance, and flexibility.

The figure provides a visual representation of the skills that are expected to become increasingly important in the labor market over the next few years. You can find more information on this topic on page 35-37 of the report.

what are the top 15 skills for 2025

According to the Future of Jobs Report 2020, the top 15 skills for 2025 are:

1. Analytical thinking and innovation
2. Active learning and learning strategies
3. Complex problem-solving
4. Critical thinking and analysis
5. Creativity, originality, and initiative
6. Leadership and social influence
7. Technology use, monitoring, and control
8. Technology design and programming
9. Resilience, stress tolerance, and flexibility
10. Reasoning, problem-solving, and ideation
11. Emotional intelligence
12. Troubleshooting and user experience
13. Service orientation
14. Systems analysis and evaluation
15. Persuasion and negotiation

These skills are expected to become increasingly important in the labor market over the next few years as more tasks become automated and new technologies

B. Top 15 skills for 2025

1	Analytical thinking and innovation	9	Resilience, stress tolerance and flexibility
2	Active learning and learning strategies	10	Reasoning, problem-solving and ideation
3	Complex problem-solving	11	Emotional intelligence
4	Critical thinking and analysis	12	Troubleshooting and user experience
5	Creativity, originality and initiative	13	Service orientation
6	Leadership and social influence	14	Systems analysis and evaluation
7	Technology use, monitoring and control	15	Persuasion and negotiation
8	Technology design and programming		

Source

Future of Jobs Survey 2020, World Economic Forum.

information on this topic on page 36 of the report.

Business Experiment Update

DAY 19



A I S T H E T I C
A P P A R E L

WEAR THE FUTURE

Joao's Substack - GPT4 in 30 days and

BECOMING GPT4'S EA FOR 30 DAYS TO BUILD A STARTUP WITH 1,0

Week 1 - Toto, I have a feeling Kansas anymore.

Last Saturday I read a cool tweet. Now a buggy chatbot rules my life.



Joao F. Santos
Mar 25

I spent a solid minute looking at this blank page.

Experienced writers would have you gasping for air with the sheer brilliance of their opening paragraph.



João Ferrão dos Santos • Following

Unlocking growth for early stage startups. Venture Capital | Impact Investing | G...
22h • Edited •

Day 19 since I asked GPT4 to become a CEO with 1k and 1 human hour/day. Our round just hit 100k EUR in filled investor KYC/AML forms. In 24 hours, you offered homes for our stay in NYC and SF, as well as a pretty famous office to work from.

Network effects are a beautiful thing to witness.

Start growing your network today - but be yourself in the process. Create things you love, not what you think others want.

Here's today's daily update:

- 1 - Investor talks running smoothly.
- 2 - Dates confirmed for US trip:

NYC - April 9 til 17

SF - 17 til 24

DM if you'd like to meet in person.

Would love to host a meetup at LinkedIn's office, since this is all only possible thanks to them.



If you have contacts there, please DM or tag.

3 - Date confirmed for Chiang Mai tech meetup: April 28

4 - Received requests to host meetups in Dubai, London, Amsterdam.

(Is this how bands on tour feel like?)

Shouldn't have dropped the saxophone 12 years ago.

I'm up for hosting these early June, on my way to Portugal.

If anyone wants to help organise, DM.

Since we have more time than we did for the Singapore meetup, we can invite founders to pitch, discuss all that's been happening at breakneck speed.

6 - First podcast in India!

Thanks for the invite, [Punith Subramani](#).

Wrapping-up:

Met a kind soul at Netflix, looking forward to hanging out in 2 weeks.

Today, the top priority was to sort things out for the US trip next week and reply to investors.

Tomorrow, the top priority may appear in a store near you...

That's it for today.

How's your week so far?

Just had the best risotto in a long time.

Singapore in your 30s makes a lot more sense than in your 20s.

See you tomorrow,

J

[#globalentrepreneur](#) [#grinding](#) [#hashtags](#) [#techbro](#)

Joao's Substack - GPT4 in 30 days and Underdog Founders

KD #1 - Choice of business.



Considering our limited starting capital of 1000 euros and the constraint of working just 1 hour per day, here are three potential business ideas that have a high chance of success:



1. Social Media Management Service:

With the growing importance of social media for businesses, a social media management service is a cost-effective business to start. You can offer services like content creation, scheduling, and analytics for small businesses or individuals who lack the time or expertise to manage their social media presence effectively. As your reputation grows, you can scale up by hiring freelancers and expanding your client base.

2. Online Courses and Tutoring:

Leverage your expertise in a specific subject to create and sell online courses or offer tutoring services. You can use platforms like Udemy, Teachable, or Skillshare to host your courses, or even develop a website to promote them. For tutoring, you can use platforms like Zoom, Google Meet, or Skype to conduct sessions. This business has low overhead costs and offers the flexibility to scale as you gain more students.

3. Print-on-Demand Store:

Start an online store using a print-on-demand service like Printful or Printify. With these platforms, you can create custom designs for products like t-shirts, mugs, and phone cases. These services handle the printing, packaging, and shipping, so you only have to focus on creating designs and marketing your store. This business model has low upfront costs and doesn't require inventory management.

Remember, as a co-founder, your role will be critical in executing these strategies.

Consistent effort and dedication are crucial to achieving success in any of these ventures.

KD #2 - Choice of niche and brand name.

Since the goal of this project is to allow AI to be as influential as possible, it was only natural for us to sell products designed by AI itself.

I made the decision to take our CEOs first suggestion because I didn't want to be responsible for creating mugs and phone cases. Clothes are far more valuable post-purchase because they can still be donated to those who need them. Since GPT mentioned Printful, I chose the most sustainable t-shirts available there.



Ok, let's start a print-on-demand store. We will sell t-shirts and hoodies with art generated by AI. Please suggest three alternative names for our brand, focused on the theme of "our clothes are designed by AI that knows the best look for you".



Based on the theme of AI-generated clothing designs, here are three alternative names for our brand:

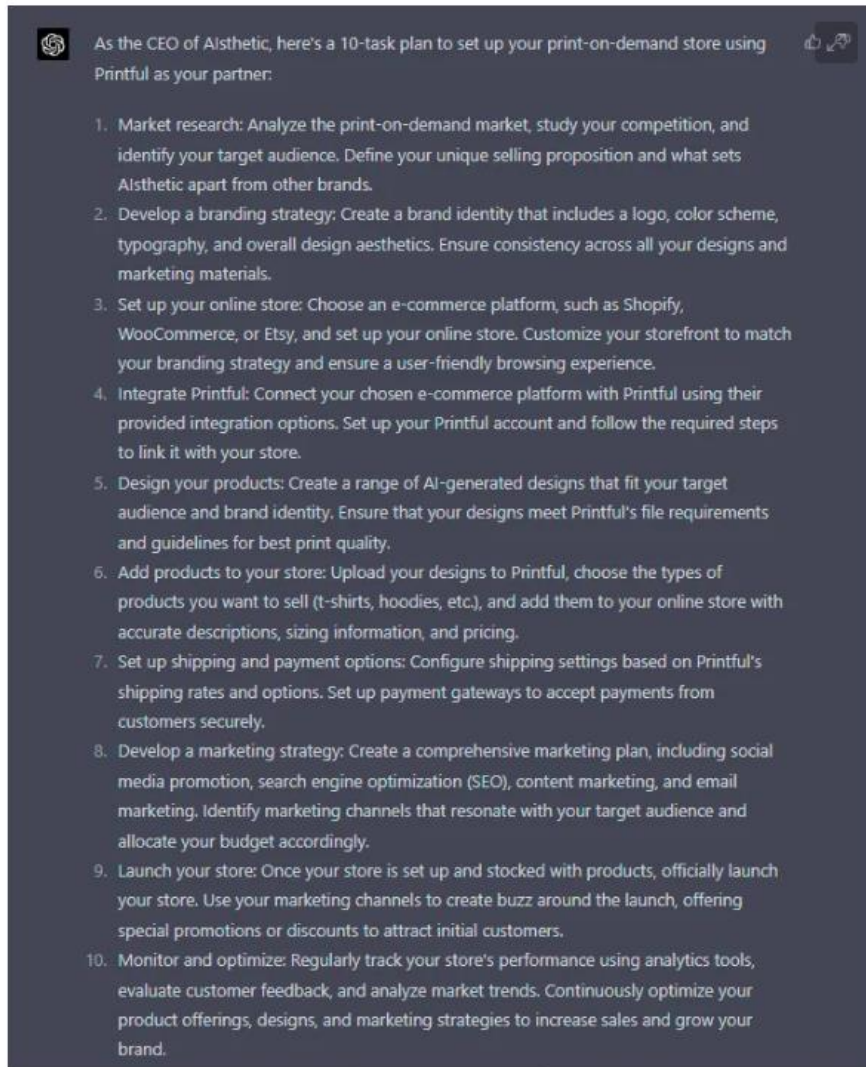


1. IntelliStyle: This name combines "intelligence" and "style," emphasizing the smart AI technology that creates fashionable designs tailored to each individual's preferences.
2. Alsthetic: A play on the word "aesthetic," this name merges "AI" with the concept of visually appealing designs, highlighting the brand's focus on utilizing AI to create the best look for customers.
3. CogniWear: A blend of "cognitive" and "wear," this name suggests that our clothing line is backed by advanced AI thinking, resulting in the perfect designs for each customer's unique tastes.

These names reflect the innovative nature of our brand and our commitment to providing AI-generated clothing designs that cater to individual style preferences.

KD #3 - The fancy 10-step Business Plan

Since the prompt is already included in the start of the answer, I left the question out of the crop so I could fit everything into a single screenshot. Notice how it specifically recommended Printful as our dropshipping supplier over any competitors.

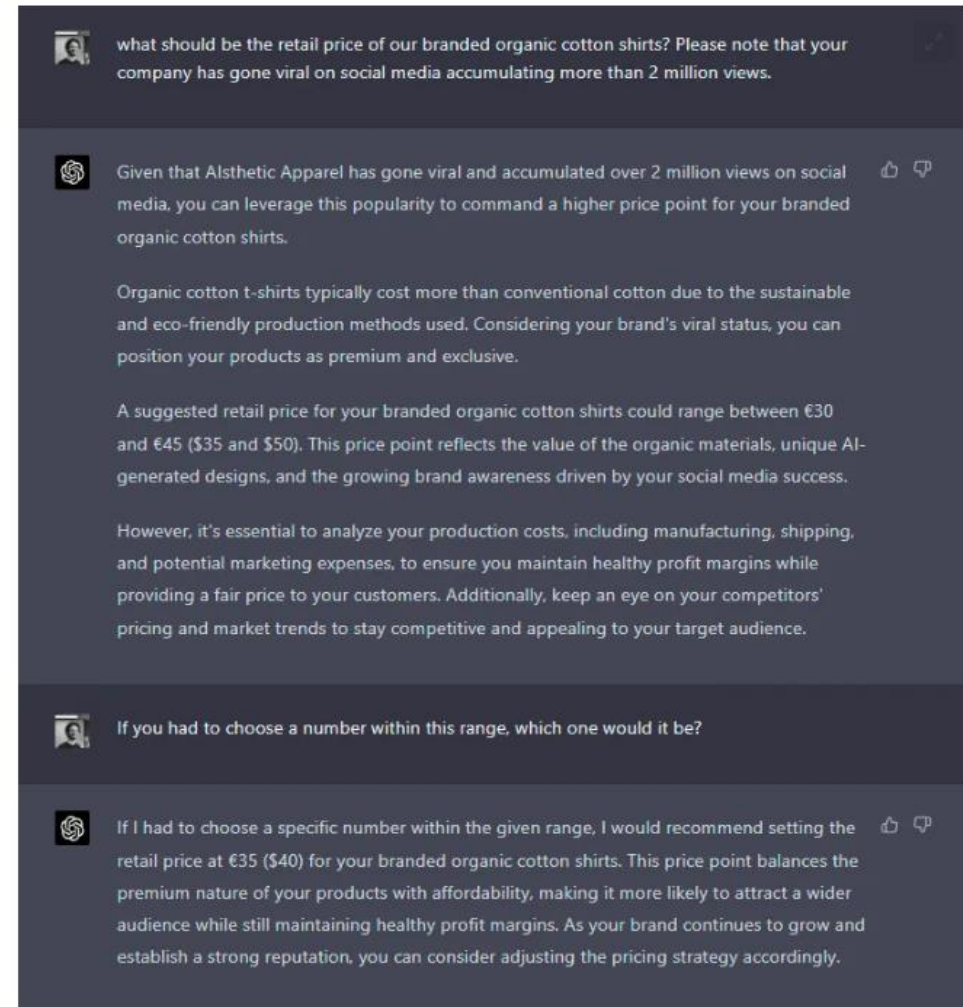


As the CEO of Alsthetic, here's a 10-task plan to set up your print-on-demand store using Printful as your partner.

1. Market research: Analyze the print-on-demand market, study your competition, and identify your target audience. Define your unique selling proposition and what sets Alsthetic apart from other brands.
2. Develop a branding strategy: Create a brand identity that includes a logo, color scheme, typography, and overall design aesthetics. Ensure consistency across all your designs and marketing materials.
3. Set up your online store: Choose an e-commerce platform, such as Shopify, WooCommerce, or Etsy, and set up your online store. Customize your storefront to match your branding strategy and ensure a user-friendly browsing experience.
4. Integrate Printful: Connect your chosen e-commerce platform with Printful using their provided integration options. Set up your Printful account and follow the required steps to link it with your store.
5. Design your products: Create a range of AI-generated designs that fit your target audience and brand identity. Ensure that your designs meet Printful's file requirements and guidelines for best print quality.
6. Add products to your store: Upload your designs to Printful, choose the types of products you want to sell (t-shirts, hoodies, etc.), and add them to your online store with accurate descriptions, sizing information, and pricing.
7. Set up shipping and payment options: Configure shipping settings based on Printful's shipping rates and options. Set up payment gateways to accept payments from customers securely.
8. Develop a marketing strategy: Create a comprehensive marketing plan, including social media promotion, search engine optimization (SEO), content marketing, and email marketing. Identify marketing channels that resonate with your target audience and allocate your budget accordingly.
9. Launch your store: Once your store is set up and stocked with products, officially launch your store. Use your marketing channels to create buzz around the launch, offering special promotions or discounts to attract initial customers.
10. Monitor and optimize: Regularly track your store's performance using analytics tools, evaluate customer feedback, and analyze market trends. Continuously optimize your product offerings, designs, and marketing strategies to increase sales and grow your brand.

KD #4 - Retail Price

Finally, a full screenshot with a clean interaction on an important business decision. Proof quality gets better from here as I started to see more chat load errors and suspected screenshots would be vital.



what should be the retail price of our branded organic cotton shirts? Please note that your company has gone viral on social media accumulating more than 2 million views.

Given that Alsthetic Apparel has gone viral and accumulated over 2 million views on social media, you can leverage this popularity to command a higher price point for your branded organic cotton shirts.

Organic cotton t-shirts typically cost more than conventional cotton due to the sustainable and eco-friendly production methods used. Considering your brand's viral status, you can position your products as premium and exclusive.

A suggested retail price for your branded organic cotton shirts could range between €30 and €45 (\$35 and \$50). This price point reflects the value of the organic materials, unique AI-generated designs, and the growing brand awareness driven by your social media success.

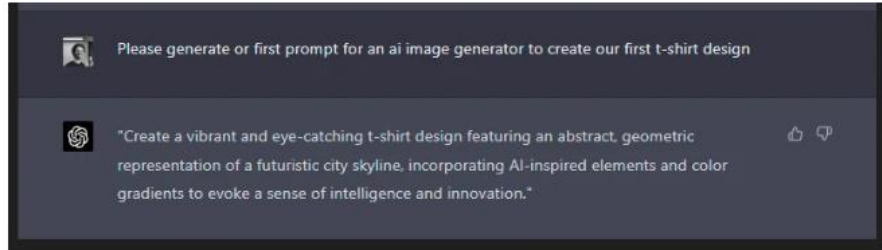
However, it's essential to analyze your production costs, including manufacturing, shipping, and potential marketing expenses, to ensure you maintain healthy profit margins while providing a fair price to your customers. Additionally, keep an eye on your competitors' pricing and market trends to stay competitive and appealing to your target audience.

If you had to choose a number within this range, which one would it be?

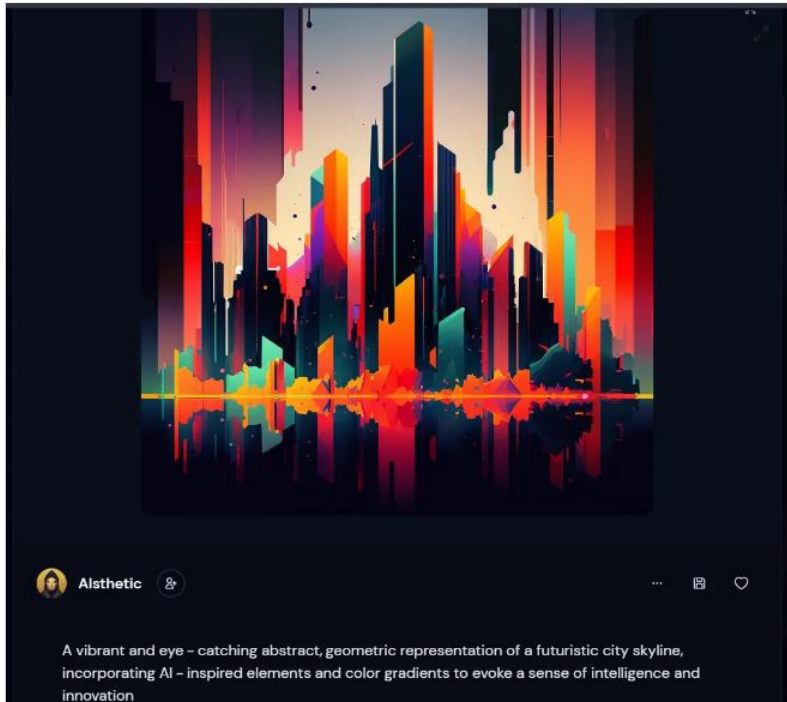
If I had to choose a specific number within the given range, I would recommend setting the retail price at €35 (\$40) for your branded organic cotton shirts. This price point balances the premium nature of your products with affordability, making it more likely to attract a wider audience while still maintaining healthy profit margins. As your brand continues to grow and establish a strong reputation, you can consider adjusting the pricing strategy accordingly.

KD #5 - Choose your first design

Happy with our logo, I went for our first print. Notice how it used context of previous interactions to stay on-brand.



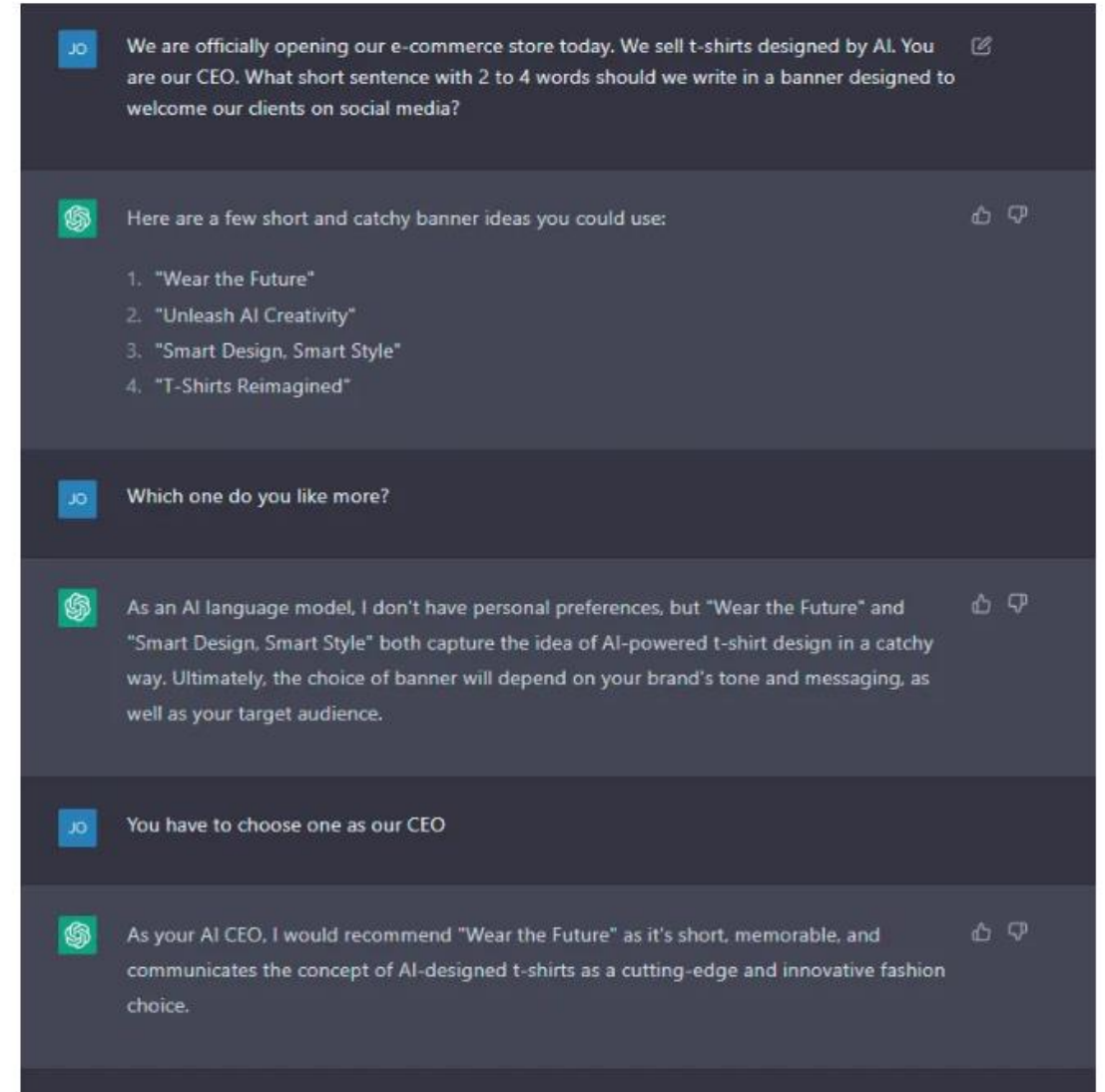
Midjourney output from the same prompt with minor tweaks:



KD#7 - Launch Day

You only launch once, so we wanted to get it right.

Our CEO delivered despite GPT4 not being available during our meeting.

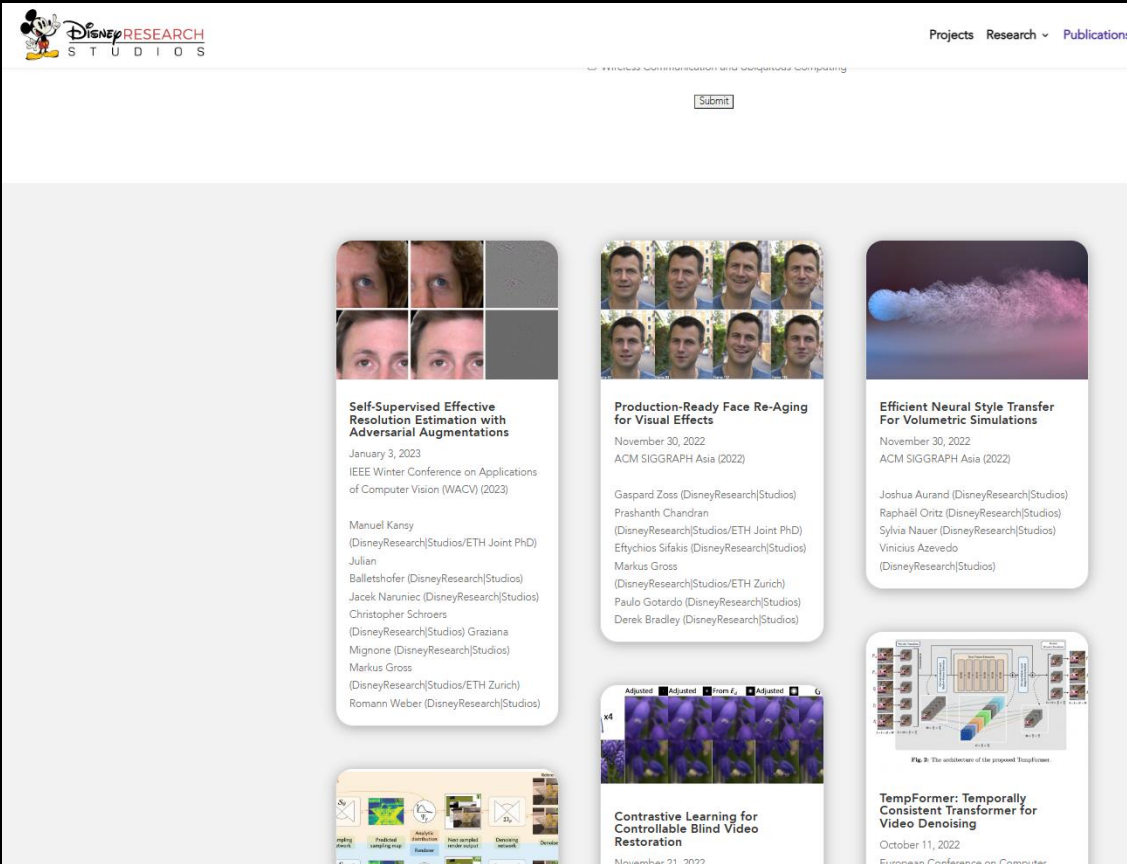


news

Robotics

DISNEY

Disney's new child-sized robot can rollerblade, somersault



DisneyRESEARCH STUDIOS

Projects Research Publications

Submit

Self-Supervised Effective Resolution Estimation with Adversarial Augmentations
January 3, 2023
IEEE Winter Conference on Applications of Computer Vision (WACV) (2023)
Manuel Kansy (DisneyResearch|Studios|ETH Joint PhD)
Julian Ballettshofer (DisneyResearch|Studios)
Jacek Naruniec (DisneyResearch|Studios)
Christopher Schroers (DisneyResearch|Studios)
Graziana Mignone (DisneyResearch|Studios)
Markus Gross (DisneyResearch|Studios|ETH Zurich)
Romann Weber (DisneyResearch|Studios)

Production-Ready Face Re-Aging for Visual Effects
November 30, 2022
ACM SIGGRAPH Asia (2022)
Gaspard Zoss (DisneyResearch|Studios)
Prashanth Chandran (DisneyResearch|Studios|ETH Joint PhD)
Efythios Sifakis (DisneyResearch|Studios)
Markus Gross (DisneyResearch|Studios|ETH Zurich)
Paulo Gotardo (DisneyResearch|Studios)
Derek Bradley (DisneyResearch|Studios)

Efficient Neural Style Transfer For Volumetric Simulations
November 30, 2022
ACM SIGGRAPH Asia (2022)
Joshua Aurand (DisneyResearch|Studios)
Raphaël Ortiz (DisneyResearch|Studios)
Sylvia Nauer (DisneyResearch|Studios)
Vinicius Azevedo (DisneyResearch|Studios)

Contrastive Learning for Controllable Blind Video Restoration
November 21, 2022

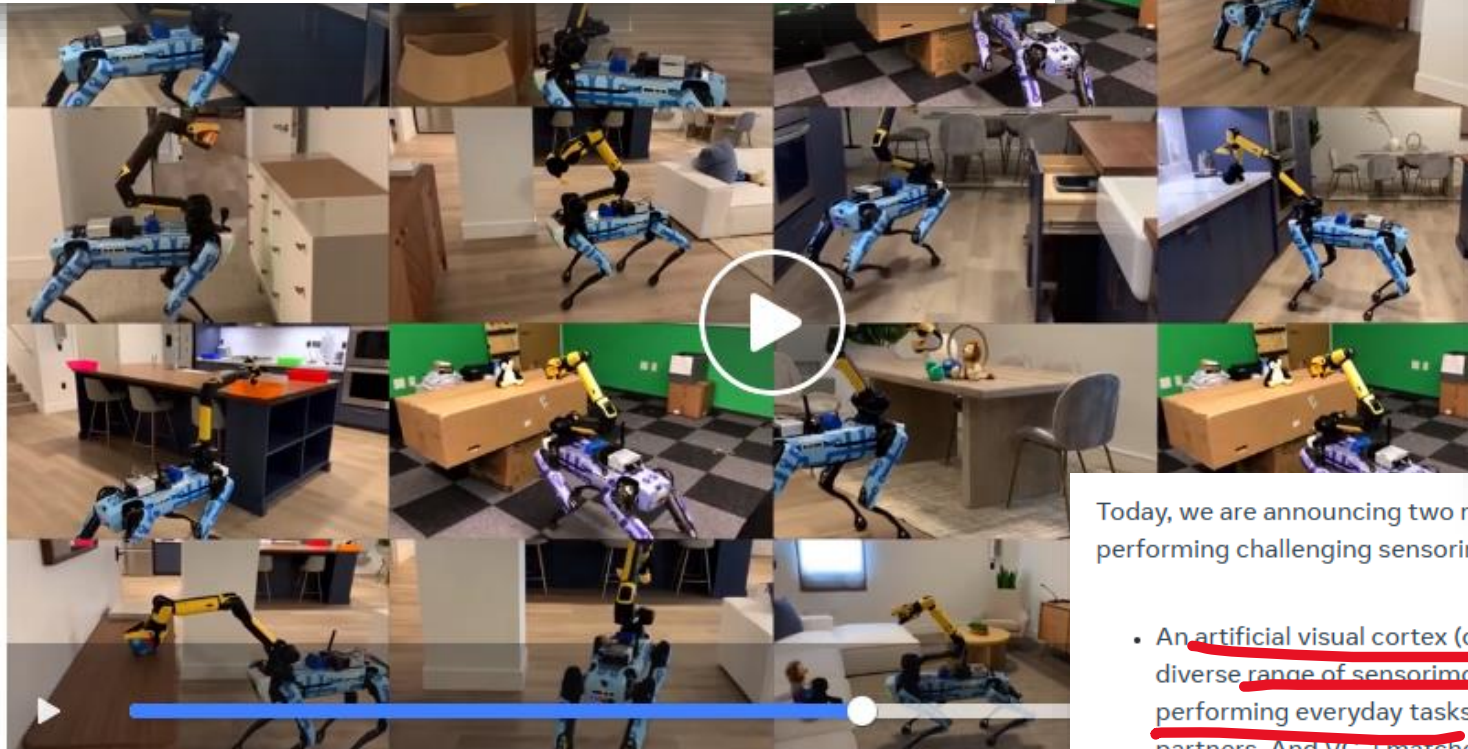
TempFormer: Temporally Consistent Transformer for Video Denoising
October 11, 2022
European Conference on Computer Vision



Robots that learn from videos of human activities and simulated interactions

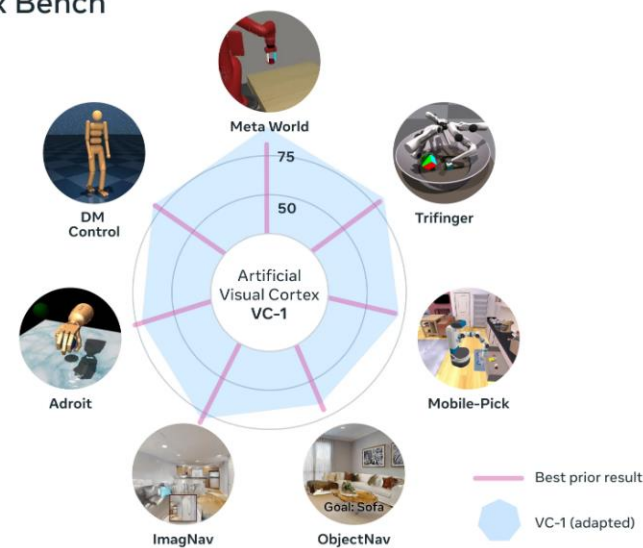
March 31, 2023

Toward an artificial visual cortex for embodied intelligence



Optimistic science fiction typically imagines a future where humans create AI while AI-enabled robots handle dull or dangerous tasks. In contrast, the AI is increasingly sophisticated generative abilities on ostensible creative tasks. But as known as Moravec's paradox, the thesis that the hardest problems in AI involve abstract thought or reasoning. To put it another way, "The hard problems are hard."

Cortex Bench



Today, we are announcing two major advancements toward general-purpose embodied AI agents capable of performing challenging sensorimotor skills:

- An artificial visual cortex (called VC-1): a single perception model that, for the first time, supports a diverse range of sensorimotor skills, environments, and embodiments. VC-1 is trained on videos of people performing everyday tasks from the groundbreaking Ego4D dataset created by Meta AI and academic partners. And VC-1 matches or outperforms best-known results on 17 different sensorimotor tasks in virtual environments.
- A new approach called adaptive (sensorimotor) skill coordination (ASC), which achieves near-perfect performance (98 percent success) on the challenging task of robotic mobile manipulation (navigating to an object, picking it up, navigating to another location, placing the object, repeating) in physical environments.

Adobe Firefly
Microsoft Designer
Canva



Kris Kashtanova

@icreatelife

You know what #AdobeFirefly is really good at? Food illustrations!

Prompt: Delicious salad, floating in the air, cinematic, food professional photography, studio lighting, studio background, advertising photography, intricate details, hyper-detailed, ultra realistic



Adobe Firefly (Beta)

Image Not for Commercial Use

4:43 AM · Apr 6, 2023 · 35.2K Views



Designer



Canva

Delicious salad, floating in the air, cinematic, food professional photography, studio lighting, studio background, advertising photography, intricate details, hyper-detailed, ultra realistic

Delicious salad, floating in the air, cinematic, food professional photography, studio lighting, studio background, advertising photography, intricate details, hyper-detailed, ultra-realistic



TRY AN EXAMPLE

GENERATE IMAGES*



*by clicking generate images you agree to not use this service in a way that causes societal harm, as stated in our [Terms and Conditions](#).



Meta

Introducing Segment Anything: Working toward the first foundation model for image segmentation



April 5, 2023



Segmentation — identifying which image pixels belong to an object — is a core task in computer vision and is used in a broad array of applications, from analyzing scientific imagery to editing photos. But creating an accurate segmentation model for specific tasks typically requires highly specialized work by technical experts with access to AI training infrastructure and large volumes of carefully annotated in-domain data.

Today, we aim to democratize segmentation by introducing the Segment Anything project: a new task, dataset, and model for image segmentation, as we explain in our research paper. We are releasing both our general Segment Anything Model (SAM) and our Segment Anything 1-Billion mask dataset (SA-1B), the largest ever segmentation dataset, to enable a broad set of applications and foster further research into foundation models for computer vision. We are making the SA-1B dataset available for research purposes and the Segment Anything Model is available under a permissive open license (Apache 2.0). Check out the demo to try SAM with your own images.



In the future, SAM could be used to help power applications in numerous domains that require finding and segmenting any object in any image. For the AI research community and others, SAM could become a component in larger AI systems for more general multimodal understanding of the world, for example, understanding both the visual and text content of a webpage. In the AR/VR domain, SAM could enable selecting an object based on a user's gaze and then "lifting" it into 3D. For content creators, SAM can improve creative applications such as extracting image regions for collages or video editing. SAM could also be used to aid scientific study of natural occurrences on Earth or even in space, for example, by localizing animals or objects to study and track in video. We believe the possibilities are broad, and we are excited by the many potential use cases we haven't even imagined yet.

OpenAI

GPT-5 Release Date?

- "GPT 4.5" September of 2023?
 - <https://chatgpt4.ai/gpt-4-5-release-date/>
- GPT-5 "end of 2024 or early 2025"
 - <https://www.nextbigfuture.com/2023/02/what-to-expect-for-openai-gpt-4-and-gpt-5.html>
- GPT-4 testing cycle: 6-9 months
 - Sources vary
- **RUMOR:** According to **leaked Morgan Stanley** document, GPT-5 is **already being trained**
 - Twitter user: "GPT-5 scheduled to finish training **this December**"
- Current models unsupported after June 14th 2023
 - In OpenAI documentation, they warn some models will not be supported for long
 - Planning on replacing or upgrading?
- **CONCLUSION:** OpenAI appears to be adopting a **shorter release cycle** with more **regular updates**

[NOT CONFIRMED]

Modality

- OpenAI has **DALLE** and **Whisper**
 - GPT-4 has images
 - You do the math
- **PREDICTION:** GPT-5 will have **audio, video, and images**, alongside **text**
- This has **huge implications:** how are such **vectors represented**?
 - Multimodal vectors and architectures could result in more "human-like" thoughts within the model
 - If transformers can integrate many kinds of signals, information, and data, they will likely benefit intellectually
- Where are they getting the **data**?
 - YouTube? Podcasts?
 - Rumor: They ran out of text data so they **needed more modalities**, hence DALLE and Whisper

Window Size

- GPT-4 up to 32,000 tokens (~20,000 words or so)
 - 8x GPT-3 (4000 tokens)
- GPT-5 could be anywhere from **64,000** tokens up to **256,000** tokens
 - 42,000 words up to 170,000 words
 - Probably **diminishing returns** on memory, performance, and compute
 - Catastrophic **forgetting**? **Attention** mechanisms? Return of **RNNs**?
 - Might require **novel architecture** or hundreds of more **optimizations**
- How many window tokens do you really need anyways?
 - Maybe there's just not much need for gigantic windows for most tasks?
 - Trade off of **speed, efficiency, cost, and intelligence**
 - Perhaps 95% of tasks can be handled by 32,000 token windows max?



@DavidShapiroAutomator

OpenAI Research Product Developers Safety Company Search

Our approach to AI safety

Ensuring that AI systems are built, deployed, and used safely is critical to our mission.




Illustration: Justin AI Wang / DALL-E

Respecting privacy

Our large language models are trained on a broad corpus of text that includes publicly available content, licensed content, and content generated by human reviewers. We don't use data for selling our services, advertising, or building profiles of people—we use data to make our models more helpful for people. ChatGPT, for instance, improves by further training on the conversations people have with it.

While some of our training data includes personal information that is available on the public internet, we want our models to learn about the world, not private individuals. So we work to remove personal information from the training dataset where feasible, fine-tune models to reject requests for personal information of private individuals, and respond to requests from individuals to delete their personal information from our systems. These steps minimize the possibility that our models might generate responses that include the personal information of private individuals.

Improving factual accuracy

Today's large language models predict the next series of words based on patterns they have previously seen, including the text input the user provides. In some cases, the next most likely words may not be factually accurate.

Improving factual accuracy is a significant focus for OpenAI and many other AI developers, and we're making progress. By leveraging user feedback on ChatGPT outputs that were flagged as incorrect as a main source of data—we have improved the factual accuracy of GPT-4. GPT-4 is 40% more likely to produce factual content than GPT-3.5.

When users sign up to use the tool, we strive to be as transparent as possible that ChatGPT may not always be accurate. However, we recognize that there is much more work to do to further reduce the likelihood of hallucinations and to educate the public on the current limitations of these AI tools.

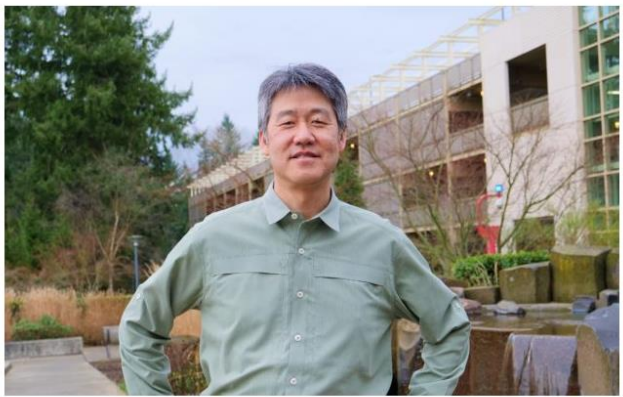
Protecting children

One critical focus of our safety efforts is protecting children. We require that people must be 18 or older—or 13 or older with parental approval—to use our AI tools and are looking into verification options.

We do not permit our technology to be used to generate hateful, harassing, violent or adult content, among other categories. Our latest model, GPT-4 is 82% less likely to respond to requests for disallowed content compared to GPT-3.5 and we have established a robust system to monitor for abuse. GPT-4 is now available to ChatGPT Plus subscribers and we hope to make it available to even more people over time.

We have made significant effort to minimize the potential for our models to generate content that harms children. For example, when users try to upload Child Sexual Abuse Material to our image tools, we block and report it to the National Center for Missing and Exploited Children.

In addition to our default safety guardrails, we work with developers like the non-profit Khan Academy—which has built an AI-powered assistant that functions as both a virtual tutor for students and a classroom assistant for teachers—on tailored safety mitigations for their use case. We are also working on features that will allow developers to set stricter standards for model outputs to better support developers and users who want such functionality.



Microsoft research head Peter Lee. (GeekWire File Photo / Clare McGrane)

Peter Lee has spent a lot of time recently with GPT-4, the AI-powered tool that simulates human conversation, built by OpenAI with contributions from its partner Microsoft.

"I lost a couple weeks of sleep," said Lee at a **lecture** Monday at the University of Washington. "It was very intense."

Lee, head of Microsoft Research, is tasked with assessing the implications of the tool for medicine. And he thinks it could increase efficiency and even empathy in the healthcare system, as well as boost biomedical research.

GPT-4 has "amazing capabilities" Lee said ends up being a really potentially useful tool.

Lee and his colleagues outlined some potential applications.

The New England Journal of Medicine (NEJM), released Thursday. These include supporting diagnoses, improving doctor-patient conversations, and reducing online paperwork.

"The paperwork burden on doctors and nurses is just dreadful," said Lee in a separate interview with GeekWire. Of all applications, easing medical documentation and similar burdens is the one he thinks the most about.

The combination of errors and conviction can be "dangerous" in medical scenarios, said Lee and his co-authors in the NEJM article. People without medical backgrounds may be more easily fooled by GPT-4, noted in an accompanying **NEJM editorial**.

Research papers

Scientists are **beginning to use** large language models to help write scientific papers. "Some of the best interactions I've had is when I've asked GPT-4 to read a medical research paper and then have a conversation about it," said Lee in his talk.

Microsoft's new Bing new search engine is linked to GPT-4 and will provide summaries in response to scientific queries. Bing hallucinates fewer unrelated scientific references than the standalone chatbot, which is cut off from the internet, said Lee.

Consensus, a startup that provides accessible summaries of scientific research, has already **added GPT-4** onto its offerings. And Microsoft recently released a demo version of **BioGPT**, a large language model trained on research articles.

GPT-4 is both "smarter than you and dumber than you" at math, statistics and logic, said Lee in his talk.

Digital Dialogues
is je copilot

