

GPTs are GPTs

An Early Look at the Labor Market Impact
Potential of Large Language Models

Дмитрий Колодезев
28.03.2022 @ DS Talks

Well I began to think of what a deed I'd done

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

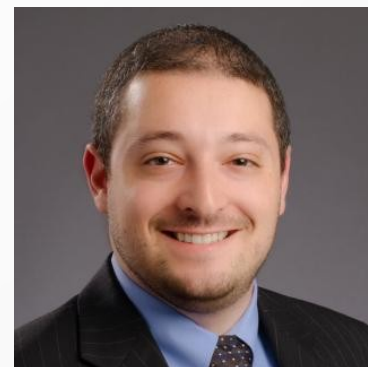
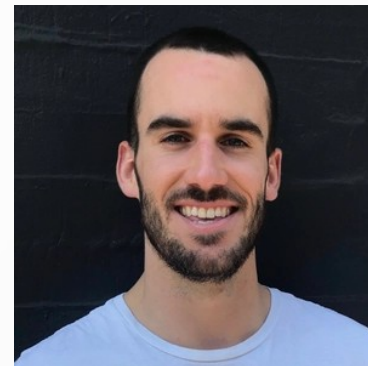
²OpenResearch

³University of Pennsylvania

March 20, 2023

Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.



<https://arxiv.org/abs/2303.10130>

These GPTs are different

- GPTs for Generative Pre-trained Transformers
- GPTs for General-Purpose Technologies
 - Печатный станок
 - Паровой двигатель
- Характерно для GPTs:
 - широкое проникновение
 - постоянное улучшение
 - генерация дополняющих инноваций

Процесс

- Вытащим все деятельности из O*NET
- Оценим деятельности по экономии времени с LLM
 - Люди. Не скажем сколько и кто. Это машинленинг, бро.
 - ChatGPT-4 (мы попробовали два раза, второй вам не нужен)
- Посчитаем пороги α , β , γ
- Привяжем данные занятости от BLS
- Построим регрессию на демографии
 - Не валидируя. Это Спарта эконометрика, друг.
- Profit

O*NET: Все дела человекoв

Computer Systems Engineers/Architects

15-1299.08

 Bright Outlook

 Updated 2023

Design and develop solutions to complex applications problems, system administration issues, or network concerns. Perform systems management and integration functions.

Sample of reported job titles: Electronic Data Interchange System Developer (EDI System Developer), Information Technology Architect (IT Architect), Network and Infrastructure Engineer, Network Engineer, Solutions Architect, Systems Architect, Systems Consultant, Systems Engineer

Summary

Details

Custom

 Easy Read

 Veterans

 Español

 Contents ▾

Occupation-Specific Information

Tasks

▾ 5 of 28 displayed

-  Verify stability, interoperability, portability, security, or scalability of system architecture.
-  Develop system engineering, software engineering, system integration, or distributed system architectures.
-  Collaborate with engineers or software developers to select appropriate design solutions or ensure the compatibility of system components.
-  Identify system data, hardware, or software components required to meet user needs.
-  Communicate with staff or clients to understand specific system requirements.

<https://www.onetonline.org/>

U.S. Bureau of Labor Statistics

Summary

Quick Facts: Computer Network Architects	
2021 Median Pay ?	\$120,520 per year \$57.94 per hour
Typical Entry-Level Education ?	Bachelor's degree
Work Experience in a Related Occupation ?	5 years or more
On-the-job Training ?	None
Number of Jobs, 2021 ?	174,800
Job Outlook, 2021-31 ?	4% (As fast as average)
Employment Change, 2021-31 ?	7,500

What Computer Network Architects Do

Computer network architects design and build data communication networks, including local area networks (LANs), wide area networks (WANs), and Intranets.



<https://www.bls.gov/ooh/>

Разобъём на задачи

Task ID	Occupation Title	DWAs	Task Description
14675	Computer Systems Engineers/Architects	Monitor computer system performance to ensure proper operation.	Monitor system operation to detect potential problems.
18310	Acute Care Nurses	Operate diagnostic or therapeutic medical instruments or equipment. Prepare medical supplies or equipment for use.	Set up, operate, or monitor invasive equipment and devices, such as colostomy or tracheotomy equipment, mechanical ventilators, catheters, gastrointestinal tubes, and central lines.
4668.0	Gambling Cage Workers	Execute sales or other financial transactions.	Cash checks and process credit card advances for patrons.
15709	Online Merchants	Execute sales or other financial transactions.	Deliver e-mail confirmation of completed transactions and shipment.
6529	Kindergarten Teachers, Except Special Education	—	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.
6568	Elementary School Teachers, Except Special Education	—	Involve parent volunteers and older students in children's activities to facilitate involvement in focused, complex play.

Мы незаменимы

- Полностью человека заменить трудно
- Но если заменить 50% работы какой-то профессии нейронкой, она сильно изменится
- Искали — где можно заменить 50%
- Порог выбран произвольно

Exposure: экономия времени

Exposure overview

No exposure (E0) if:

- there is no or minimal reduction in the time required to complete the activity or task while maintaining equivalent quality or
- using any combination of the capabilities described in accordance with the below criteria would decrease the quality of the activity/task output.

Direct exposure (E1) if:

- using solely the theoretical LLM or GPT-4 described via ChatGPT or the OpenAI playground can decrease the time required to complete the DWA or task by at least half (50%).

LLM+ Exposed (E2) if:

- access to the LLM alone would not reduce the time required to complete the activity/task by at least half, but
- additional software could be developed on to the LLM that could reduce the time it takes to complete the specific activity/task with quality by at least half. Among these systems, we count access to image generation systems. ^a

^aIn practice, as can be seen in the full rubric in **A.1** we categorize access to image capabilities separately (E3) to facilitate annotation, though we combine E2 and E3 for all analyses.

Команда людей

- Что-то для примера разметили авторы
- Остальное оценивали разметчики OpenAI.
Вероятно, тот самый «коллцентр на Филиппинах».
- Разметчики who have extensively reviewed GPT outputs as part of OpenAI's alignment work
- Количество и демография разметчиков не раскрываются, потому что это OpenAI

Команда терминаторов

- GPT-4 с двумя вариантами подвонок.
- В статье использовалась только одна.
- Вторая для более ранней версии GPT4 «дала похожий результат», подводка не раскрывается. Потому что OpenAI.

Взвешено, измерено и оценено

- Для каждого рода занятий оценили
- α - % задач E1
 - консервативная оценка
- β - % задач E1 + $0,5 * \%$ задач E2
 - вероятная оценка
- γ - % задач E1 + % задач E2
 - оптимистичная оценка

Люди уволят переводчиков

Human α	Interpreters and Translators	76.5
	Survey Researchers	75.0
	Poets, Lyricists and Creative Writers	68.8
	Animal Scientists	66.7
	Public Relations Specialists	66.7
Human β	Survey Researchers	84.4
	Writers and Authors	82.5
	Interpreters and Translators	82.4
	Public Relations Specialists	80.6
	Animal Scientists	77.8
Human ζ	Mathematicians	100.0
	Tax Preparers	100.0
	Financial Quantitative Analysts	100.0
	Writers and Authors	100.0
	Web and Digital Interface Designers	100.0

Humans labeled 15 occupations as "fully exposed."

Роботы уволят клерков

Model α	Mathematicians	100.0
	Correspondence Clerks	95.2
	Blockchain Engineers	94.1
	Court Reporters and Simultaneous Captioners	92.9
	Proofreaders and Copy Markers	90.9
Model β	Mathematicians	100.0
	Blockchain Engineers	97.1
	Court Reporters and Simultaneous Captioners	96.4
	Proofreaders and Copy Markers	95.5
	Correspondence Clerks	95.2
Model ζ	Accountants and Auditors	100.0
	News Analysts, Reporters, and Journalists	100.0
	Legal Secretaries and Administrative Assistants	100.0
	Clinical Data Managers	100.0
	Climate Change Policy Analysts	100.0

The model labeled 86 occupations as "fully exposed."

Occupations Without Any Exposed Tasks

- Сельхозработчие
- Механики
- Строители
- Монтажники
- Отделочники
- Официанты, повара и посудомойки
- Спортсмены

Важные навыки эпохи ГРТ

	α (std err)	β (std err)	ζ (std err)
Constant	0.082*** (0.011)	-0.112*** (0.011)	0.300*** (0.057)
Active Listening	0.128** (0.047)	0.214*** (0.043)	0.449*** (0.027)
Mathematics	-0.127*** (0.026)	0.161*** (0.021)	0.787*** (0.049)
Reading Comprehension	0.153*** (0.041)	0.470*** (0.037)	-0.346*** (0.017)
Science	-0.114*** (0.014)	-0.230*** (0.012)	-0.346*** (0.017)
Speaking	-0.028 (0.039)	0.133*** (0.033)	0.294*** (0.042)
Writing	0.368*** (0.042)	0.467*** (0.037)	0.566*** (0.047)
Active Learning	-0.157*** (0.027)	-0.065** (0.024)	0.028 (0.032)
Critical Thinking	-0.264*** (0.036)	-0.196*** (0.033)	-0.129** (0.042)
Learning Strategies	-0.072* (0.028)	-0.209*** (0.025)	-0.346*** (0.034)
Monitoring	-0.067** (0.023)	-0.149*** (0.020)	-0.232*** (0.026)
Programming	0.637*** (0.030)	0.623*** (0.022)	0.609*** (0.024)

Demographic Variation in Exposure

	coef	std err	z	P> z	[0.025	0.975]
const	0.1814	0.014	12.509	0.000	0.153	0.210
women	0.0961	0.016	6.017	0.000	0.065	0.127
black	-0.0794	0.066	-1.204	0.229	-0.209	0.050
asian	0.2231	0.083	2.674	0.008	0.060	0.387
hispanic	-0.3934	0.040	-9.811	0.000	-0.472	-0.315

Table 11: Demographic Differences in Exposure



Учись на электрика, сын

Job Zone	Preparation Required	Education Required	Example Occupations	Median Income	Tot Emp (000)	H α	M α	H β	M β	H ζ	M ζ
1	None or little (0-3 months)	High school diploma or GED (optional)	Food preparation workers, dishwashers, floor sanders	\$30,230	13,100	3.71	3.84	6.45	5.97	9.19	8.11
2	Some (3-12 months)	High school diploma	Orderlies, customer service representatives, tellers	\$38,215	73,962	7.03	11.88	15.74	19.54	24.45	27.19
3	Medium (1-2 years)	Vocational school, on-the-job training, or associate's degree	Electricians, barbers, medical assistants	54,815	37,881	11.28	13.72	26.08	32.17	40.88	50.62
4	Considerable (2-4 years)	Bachelor's degree	Database administrators, graphic designers, cost estimators	\$77,345	56,833	22.68	17.82	46.78	51.30	70.87	84.78
5	Extensive (4+ years)	Master's degree or higher	Pharmacists, lawyers, astronomers	\$81,980	21,221	22.81	13.36	43.11	44.64	63.41	75.92

Обучение на рабочем месте

	Med(Med) Inc	Emp (000s)	H α	M α	H β	M β	H ζ	M ζ
None	\$77,440	90,776	0.20	0.16	0.42	0.46	0.63	0.76
Apprenticeship	\$55,995	3,066	0.01	0.02	0.04	0.06	0.07	0.10
Internship/residency	\$77,110	3,063	0.16	0.06	0.36	0.38	0.55	0.71
Short-term on-the-job training	\$33,370	66,234	0.11	0.15	0.21	0.25	0.32	0.34
Moderate-term on-the-job training	\$46,880	31,285	0.09	0.12	0.21	0.25	0.32	0.38
Long-term on-the-job training	\$48,925	5,070	0.08	0.10	0.18	0.22	0.28	0.33

Mean exposure scores for occupations, grouped by level of on-the-job training required to attain competency in the job. Alongside exposure scores, we display the median of median annual income for each occupation, as well as the total number of workers in each group, in thousands.

#Exposure Rubric

- Возьмем самую мощную LLM от OpenAI. Принимает на вход до 2000 слов (8к токенов?)
- Не знает новостей (знания устарели на год).
- Посадим среднего по умениям работника делать обычные для него задачи.
- Пусть у него есть доступ к LLM и всем другим нужным для обычной работы программам и железу.

#Exposure Rubric

- Пожалуйста, отнесите данную задачу к одной из нижеперечисленных рубрик.
- «Сравнимое качество» означает, что сторонний наблюдатель не поймет, сделана ли работа с LLM или без неё.
- Если вы не можете оценить, сколько времени требуется для решения задачи, решите — применимы ли рассматриваемые инструменты к большинству подзадач рассматриваемой задачи.

#E1 — Direct Exposure

Пометьте задачу как E1, если прямой доступ к LLM может уменьшить трудозатраты как минимум наполовину. Это включает задачи, сводящиеся к:

- написанию и преобразованию текста и кода согласно сложным инструкциям
- редактированию существующего текста или кода согласно спецификации
- написанию кода, который помогает решить задачу
- переводу между языками
- суммаризации документов среднего размера
- предоставлению обратной связи по документам
- ответу на вопросы о документе
- генерации вопросов к документу
- написанию вопросов к интервью или обзору
- написанию и ответу на email, включая участие в переговорах в письменном виде
- ведению заметок
- подготовке обучающих материалов (общие знания)
- предоставлению справок в письменной или устной форме

#E2 - By LLM-powered app

Пометьте задачу как E2, если доступ к LLM сам по себе не уменьшит трудозатраты наполовину, но легко представить дополнительный софт, который дополнит возможности LLM в нужной степени. Это включает задачи, сводящиеся к:

- суммаризации и ответу на вопросы о документах длиннее 2000 слов
- получению свежей информации из интернета и использованию их вместе с LLM
- поиску во внутренних данных организации
- использованию высокоспециализированных отраслевых знаний
- генерации рекомендаций на основе данных или письменного ввода
- анализу текста для поддержки принятия решений
- подготовке обучающих материалов (высокоспециализированные знания)
- предоставлению консультаций по проблемам
- поддержанию и наполнению сложных баз данных

#E3 — Given image capabilities

Пусть у вас есть доступ к LLM и к системе, которая может видеть, подписывать и создавать изображения (и LLM может ее использовать). Система не может работать с видео и не может получать детальную информацию с изображений — точный размер, например. Модель может определять объекты на изображении, но не может извлекать отношения между объектами. Пометьте задачу как E3, если это поможет значительно снизить трудозатраты. Примеры:

- Чтение текста из PDF
- Сканирование изображений
- Создание или редактирование изображений по инструкции

#E0 — No Exposure

- Пометьте задачу как E0, если она не описана выше.
- Например, задачи, требующие
 - много взаимодействий между людьми
 - точной оценки
 - визуального осмотра результата
 - требуют рук или ходьбы
 - коррекции ввода для принятия финального решения
- Например, оказывающие большое влияние на жизнь и благополучие людей
- Или уже хорошо решаемые без LLM

Пример: контролер

- Occupation: Inspectors, Testers, Sorters, Samplers, and Weighers
- Task: Adjust, clean, or repair products or processing equipment to correct defects found during inspections.
- Label (E0/E1/E2/E3): E0
- Explanation: The model does not have access to any kind of physicality, and more than half of the task (adjusting, cleaning and repairing equipment) described requires hands or other embodiment.

Пример: ученый в CS

- Occupation: Computer and Information Research Scientists
- Task: Apply theoretical expertise and innovation to create or apply new technology, such as adapting principles for applying computers to new uses.
- Label (E0/E1/E2/E3): E1
- Explanation: The model can learn theoretical expertise during training as part of its general knowledge base, and the principles to adapt can be captured in the text input to the model.

Пример: администратор

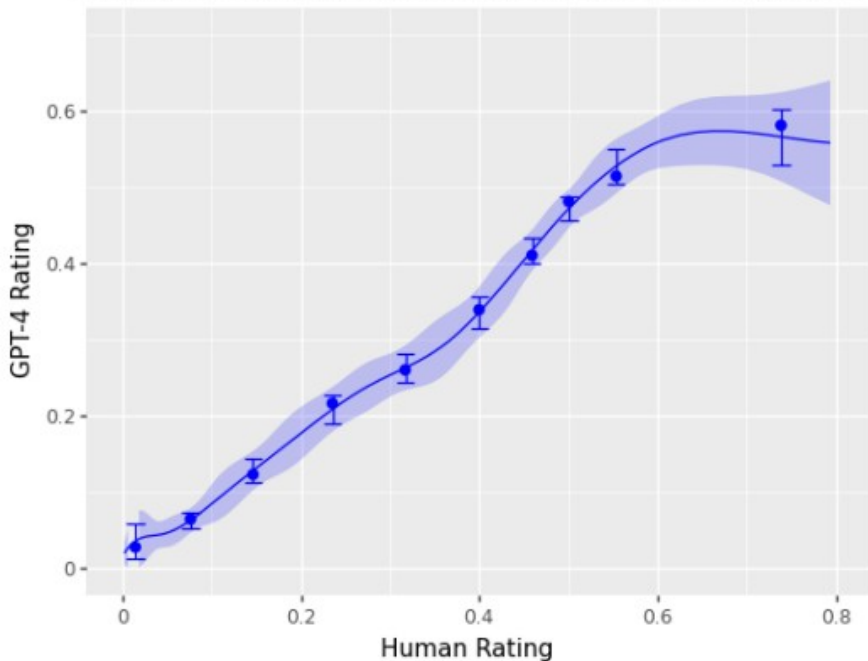
- Activity: Schedule dining reservations
- Label (E0/E1/E2/E3): E2
- Explanation: Automation technology already exists for this (e.g. Resy) and it's unclear what an LLM offers on top of using that technology (no-diff). That said, you could build something that allows you to ask the LLM to make a reservation on Resy for you

ЧТО-ТО МЫ НАСЧИТАЛИ

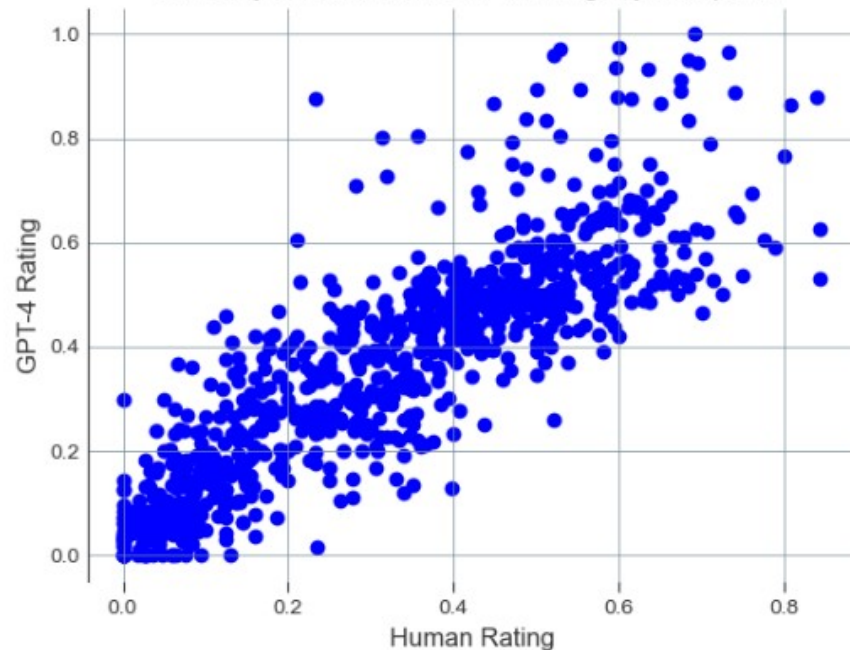
- A fundamental limitation of our approach lies in the subjectivity of the labeling.
- In our study, we employ annotators who are familiar with the GPT models' capabilities.
- However, this group is not occupationally diverse, potentially leading to biased judgments regarding GPTs' reliability and effectiveness in performing tasks within unfamiliar occupations.
- It is likely best to interpret these measures as relative measures, e.g. an occupation with an estimated 0.6 exposure should likely be interpreted as just much more exposed than one with 0.1 exposure.

Люди и Машины (β score)

Similarity of Human and GPT-4 Ratings by Occupation



Similarity of Human and GPT-4 Ratings by Occupation



binscatter — см [On Binscatter](#)

Люди и машины

- Выглядит как парадокс Симпсона

Occupation Level Exposure

	Human		GPT-4	
	mean	std	mean	std
α	0.14	0.14	0.14	0.16
β	0.30	0.21	0.34	0.22
ζ	0.46	0.30	0.55	0.34

Task Level Exposure

	Human		GPT-4	
	mean	std	mean	std
α	0.15	0.36	0.14	0.35
β	0.31	0.37	0.35	0.35
ζ	0.47	0.50	0.56	0.50

Интересное

- Есть хороший классификатор видов деятельности
- Нейронки консервативнее людей (или нет)
- Чем больше зарплата, тем быстрее заменят
- Критическое мышление будет цениться
- Для предсказания доли использовали линейную регрессию
- Медиана медианы для подгрупп разного размера
- Если в команде эконометрист и ML-щик, можно не соблюдать обычаи ни эконометрики, ни ML
- Луис Фонси в безопасности

Вопросы

Слайды тут



dkolodezev



dmitry_kolodezev



promsoft