

ГенИИ*: второй пилот со странностями

2000: Software is eating the world
2010: AI is eating the software
2023: GPT is eating AI

*Генеративный искусственный интеллект

Предмет исследования

Способности ChatGPT вызвали волну экспериментов применения искусственного интеллекта практически в каждой области, где есть когнитивная деятельность. **Цель настоящего исследования — проанализировать статус и потенциал применения новых моделей AI.**

ОПРЕДЕЛЕНИЯ

- **Искусственный интеллект (Artificial Intelligence, AI*).** Сводный термин для систем, способных имитировать когнитивную деятельность человека: принятие решений в условиях неопределенности, распознавание естественного языка, звука и изображений.
- **Машинное обучение (Machine Learning, ML*).** Один из подходов к реализации AI, при котором результат определяется не алгоритмом, а статистическими моделями, которые используют закономерности, сформированные на тренировочных данных. Включает классические статистические модели и нейронные сети. С середины 2000-х годов ML почти вытеснило другие подходы и стало синонимом AI.
- **Нейронные сети (Neural Network / Deep Neural Networks, DNN).** Модели машинного обучения, которые имитируют работу человеческого мозга: входные данные проходят через нейроны (линейные выражения), объединенные в слои. Нейроны одного слоя передают результат и активируют нейроны следующего слоя.
- **Генеративный AI (Generative AI).** Нейронные сети, способные не только распознавать, но и создавать контент.
 - **Большие языковые модели (Large Language Model, LLM).** Нейронные сети, обученные на большом количестве текста, за счет чего могут абстрагировать знания и решать новые задачи.
 - **Архитектура «Трансформер».** Наиболее используемая архитектура ML в основе больших языковых моделей, в том числе ChatGPT (Generative Pretrained Transformer). Предсказывают следующий элемент (слово), параллельно анализируя взаимное влияние до 50 тысяч слов*. Это позволяет учитывать сложный контекст из множества источников одновременно.

*на июнь 2023

Сокращения

AI / ML*	Artificial Intelligence and Machine Learning (искусственный интеллект и машинное обучение)
DNN	Deep Neural Networks (глубинные нейронные сети)
GAI	Generative AI (генеративный AI; глубинная нейронная сеть (DNN), способная создавать текстовый и мультимедийный контент)
LLM	Large Language Model (большая языковая модель)
GPT	Generative Pretrained Transformer (наиболее используемая архитектура в моделях LLM, включая ChatGPT)
NLP	Natural Language Processing (обработка естественного языка)
CNN	Convolutional Neural Network (сверточная нейронная сеть)
CV	Computer Vision (компьютерное зрение)
NN	Neural Networks (нейронные сети)
RB	Rule-based systems (If-then, heuristics) (системы, построенные на правилах)
RL	Reinforcement Learning (обучение с подкреплением)
RNN	Recurrent Neural Network (рекуррентная нейронная сеть)
SSL	Self-supervised Learning (самообучение на данных без разметки и исходя из их внутренней структуры)
TF	Transformer (архитектура машинного обучения в основе GPT)
UL	Unsupervised Learning (машинное обучение без учителя)

*Далее в исследовании AI и ML используются в контексте Generative AI и Large Language Models (LLM).

1. **Основные наблюдения**
2. Инвестиции
3. Области применения генеративного AI
4. Влияние на рынок труда
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

Статус AI*: основные наблюдения

- **Взрыв интереса к AI вызван качественным скачком генеративного AI* (особенно языковых моделей) и открытия доступа к нему широкого круга пользователей.** Вместе с этим AI используется более 20 лет и стал базовой технологией, как электричество и интернет.
- **Генеративный AI* эффективен во многих, но далеко не во всех областях.** Он увеличивает производительность, где есть большой объем неструктурированных данных, включая текст, изображения и медиа. Но там, где критичны скорость, точность и прозрачность, специализированные модели и алгоритмы могут быть более эффективны.
- **Компании наперегонки заявляют о планах внедрения генеративного AI,** но внедрение идет медленно из-за неприемлемых рисков конфиденциальности, низкой прозрачности и сложности соответствия законодательству.
- **AI повлиял на рынок труда в творческой сфере и образовании.** В большинстве других областей он увеличивает производительность и создает спрос на новые компетенции, в том числе для снижения рисков использования AI.
- **Регулирование AI затруднено из-за ограниченности данных по влиянию AI на бизнес, экономику и общество.** Предложения по странам варьируются от детальных требований к AI системам (ЕС, Китай, Канада) до «зонтичных» принципов в рамках существующего регулирования (Великобритания).
- **Снижение темпов роста технологического сектора и взрывной рост пользовательского интереса сделали AI основной инвестиционной идеей.** Капитализация Microsoft выросла на треть, стартапы без выручки получают многомиллиардную оценку. С марта на рынок вышли десятки сервисов от технологических лидеров в дополнение к 400 стартапам и десяткам open-source-моделей. Конкуренция затрудняет монетизацию сервисов, но увеличивает спрос и стоимость вычислительной инфраструктуры: капитализация Nvidia за шесть месяцев утроилась и превысила 1 трлн долларов США.

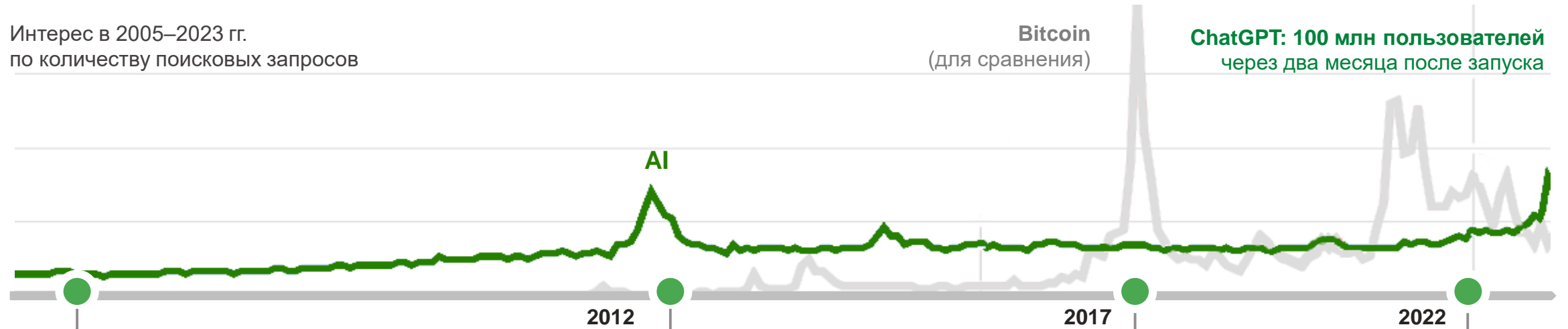
*Для целей настоящего исследования AI используется в контексте Generative AI и Large Language Models (LLM).

1. Основные наблюдения
2. **Инвестиции**
3. Области применения генеративного AI
4. Влияние на рынок труда
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

AI используется более 20 лет, и за это время он стал базовой технологией, как интернет и электричество. Взрыв интереса вызван качественным скачком технологий языковых моделей и доступа к ним широкого круга пользователей

*Список неисчерпывающий

Интерес в 2005–2023 гг.
по количеству поисковых запросов



2000-е: статистические модели

- ▶ Фильтрация спама
- ▶ Сегментация клиентов
- ▶ Рекомендации
- ▶ Выявление трендов
- ▶ Прогнозирование
- ▶ Скоринг (банки)
- ▶ Выявление аномалий

2006: нейронные сети

- ▶ Распознавание изображений, объектов
- ▶ Биометрия (лицо, голос)
- ▶ Распознавание текста
- ▶ Управление сложными системами

2014: генеративный AI: языковые модели и создание изображений

- ▶ Создание текста
- ▶ Создание изображений
- ▶ Чат-боты
- ▶ Перевод

2017: архитектура «Трансформер» (в основе GPT)

- ▶ Рост качества создаваемого контента
- ▶ Создание видео и аудио
- ▶ Создание и проверка кода

2022: ChatGPT

- ▶ Доступ к GPT-3.5 через веб-интерфейс

Примеры использования*

Источники: arXiv.org, MathWorks, GTrends, СМИ

Генеративный AI стал основной инвестиционной идеей

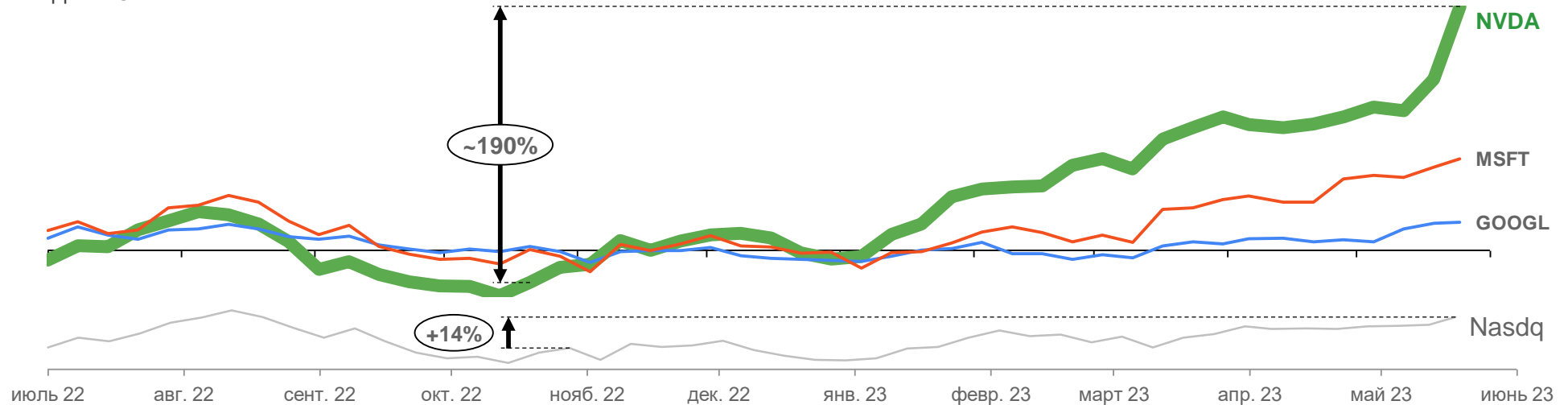
Стартапы без продукта привлекают сотни миллионов долларов. Стоимость MSFT выросла на треть, NVIDIA утроилась. Вместе с этим растет конкуренция: более 400 AI стартапов, десятки конкурентов ChatGPT от технологических компаний.

Инвесторы ищут новые точки роста. Взрыв интереса к генеративному AI в ноябре 2022 года стал почти единственной инвестиционной идеей.

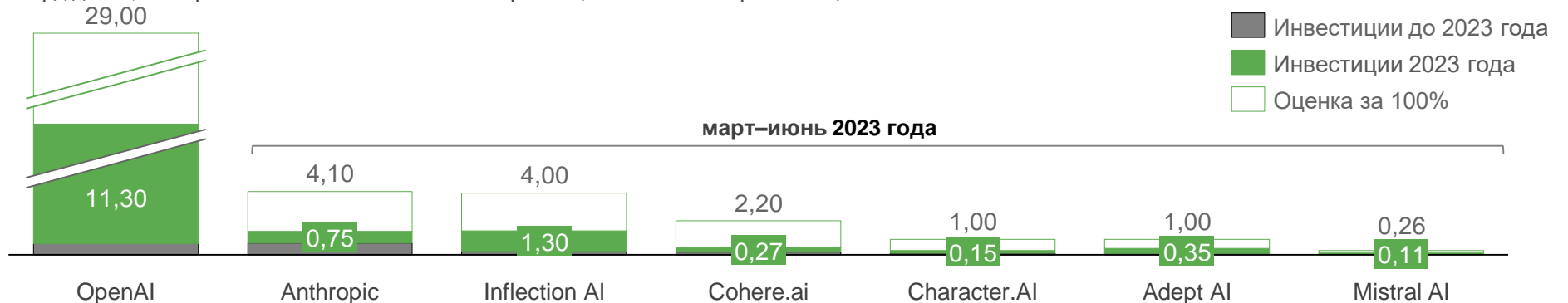
В наибольшем выигрыше игроки AI-инфраструктуры: капитализация Microsoft за шесть месяцев выросла на 30%; Nvidia – более чем на 200% и превысила 1 трлн долларов США.

Крупнейшее финансирование привлекают выходцы из ведущих лабораторий AI, включая OpenAI и Google. Основные направления их работы – решение проблем корпоративного использования генеративного AI и улучшение качества создаваемого контента.

Относительная динамика стоимости акций к ноябрю 2022 года (выход ChatGPT)
Долл. США



Крупнейшие инвестиции в стартапы генеративного AI
млрд долл., выборка на основе анализа 450 стартапов, список неисчерпывающий



Источники: CrunchBase, Pitchbook, Nasdaq



1. Основные наблюдения
2. Инвестиции
- 3. Области применения генеративного AI**
4. Влияние на рынок труда
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

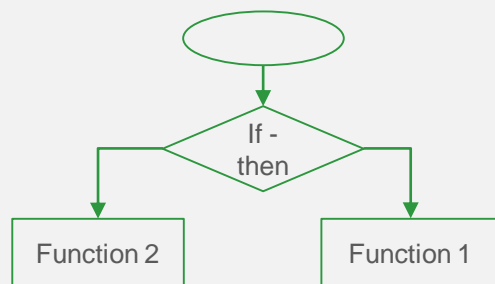
Машинное обучение, нейронные сети и генеративный AI

Машинное обучение: логика имитируется статистикой и математикой. Используется там, где нет четко выраженных правил и / или наблюдается большая вариативность. С середины 2000-х годов этот термин стал синонимом AI.

Список моделей и алгоритмов неисчерпывающий

ТРАДИЦИОННЫЕ АЛГОРИТМЫ

Результат определяют алгоритмы, основанные на логических правилах или эвристиках.



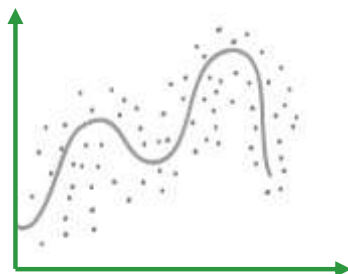
Предпочтительны, когда:

- задача имеет ограниченную сложность и вариативность;
- критичны интерпретируемость, скорость и отклик;
- нет данных для машинного обучения.

МАШИННОЕ ОБУЧЕНИЕ (ML)

Результат определяется статистическими моделями на закономерностях, сформированных на тренировочных данных. Могут работать со сложными, неявными и меняющимися правилами.

СТАТИСТИЧЕСКИЕ МОДЕЛИ



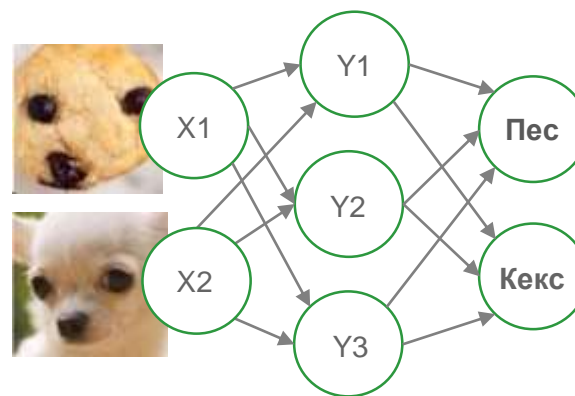
Наиболее используемые группы:

- регрессии
- классификаторы
- кластеризация
- деревья решений

НЕЙРОННЫЕ СЕТИ (НС, NN)

Входные данные проходят через нейроны, объединенные в слои. Каждый слой определяет, какие нейроны будут активированы на следующем слое.

Слои NN (Layers)



Нейрон — это линейное выражение:

$$Y_3 = F(x_1 * w_1 + x_2 * w_2 + b).$$

Выход нейронов предыдущего слоя (X) взвешивается (w, weight) и корректируется (b, bias). Функция активации (F) определяет, что будет передано на нейроны следующего слоя.

ГЕНЕРАТИВНЫЙ AI

Нейронные сети, создающие контент

Большие языковые модели (Large Language Models, LLM)

Самообучаются на больших массивах текста, создают новый текст, решают незнакомые задачи.

Примеры: GPT, Palm (Bard)

Модели создания изображений

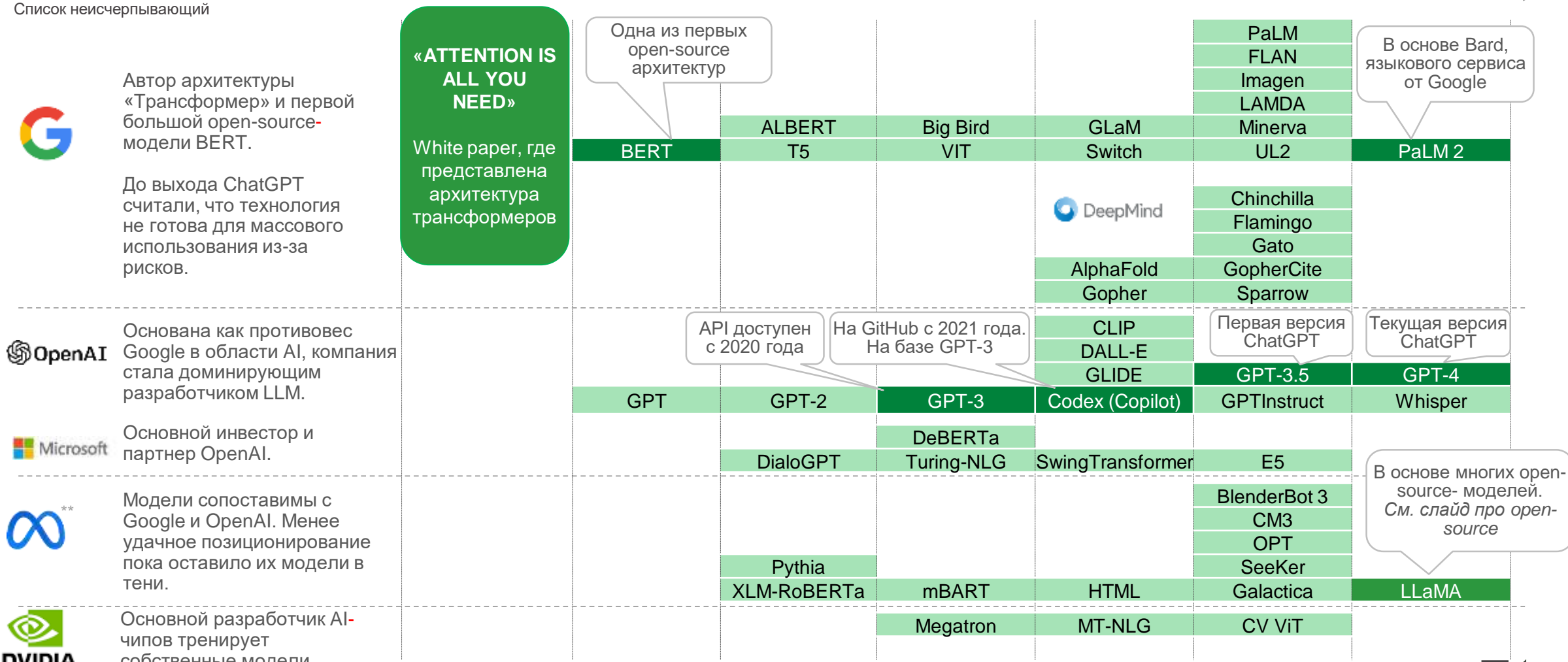
Наиболее популярная модель – Diffusion. Примеры: Dall-E 2, Stability AI, Midjourney

Архитектура нейронных сетей "Трансформер" вывела качество языковых моделей на новый уровень. С 2017 года созданы десятки базовых моделей по всему миру, включая Россию*

Обучение базовых моделей требует месяцы и тысячи специализированных чипов, ресурсы доступные лишь некоторым компаниям. Объем этих ресурсов можно оценить по объему инвестиций в стартапы генеративного ИИ. (См. Инвестиции)

Разнообразие «трансформеров»

Список неисчерпывающий



Внедрение генеративного AI часто планируют там, где «модно», а не там, где наибольший потенциал

Генеративный AI как гениальный второй пилот со странностями, которого надо держать под присмотром. Особенно при общении с внешним миром. Основной потенциал GAI — где есть большие объемы неструктурированных (текстовых) данных. Там, где нужны точность, прозрачность и обработка в реальном времени, лучше подходят специализированные модели и алгоритмы.

Потенциал генеративного AI

МАКСИМАЛЬНЫЙ

- Большой объем неструктурированных текстовых данных из несвязанных источников.
- Обработка в реальном времени, и неточности не критичны.

ВЫСОКИЙ

- Работа с внешними контрагентами.
- Работа с вычислениями и фактами.

ОГРАНИЧЕННЫЙ

- Требуются точность, интерпретируемость и обработка в режиме реального времени.
- Ограниченный объем данных и вычислительных ресурсов.

Пример бизнес-функции

Управление знаниями, аналитика

Оптимизация процессов

Дизайн, контент (например, сценарии)

Разработка ПО, исследования

Управление персоналом

Юридическая служба

Финансы и отчетность

Поддержка клиентов

Продажи и маркетинг

Поставки и логистика

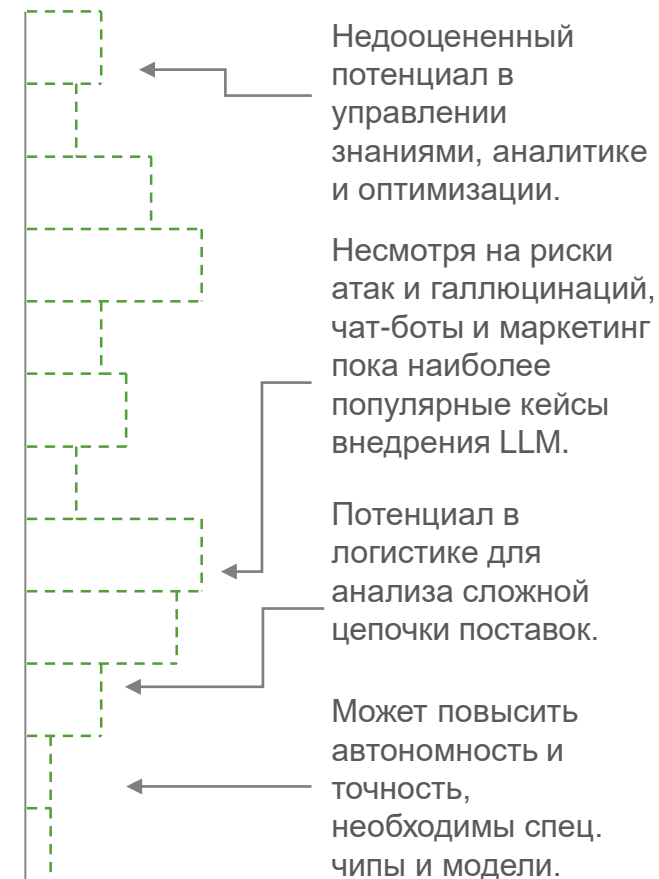
Управление энергетикой

Управление оборудованием

+	+	+
+	+	+
+	+	+
+	+	~
+	+	~
+	+	x
+	+	x
+	~	x
+	~	x
~	~	x
~	x	x
~	x	x

Большой объем текстовых данных
Скорость некритична
Неточности некритичны

Текущие внедрения*



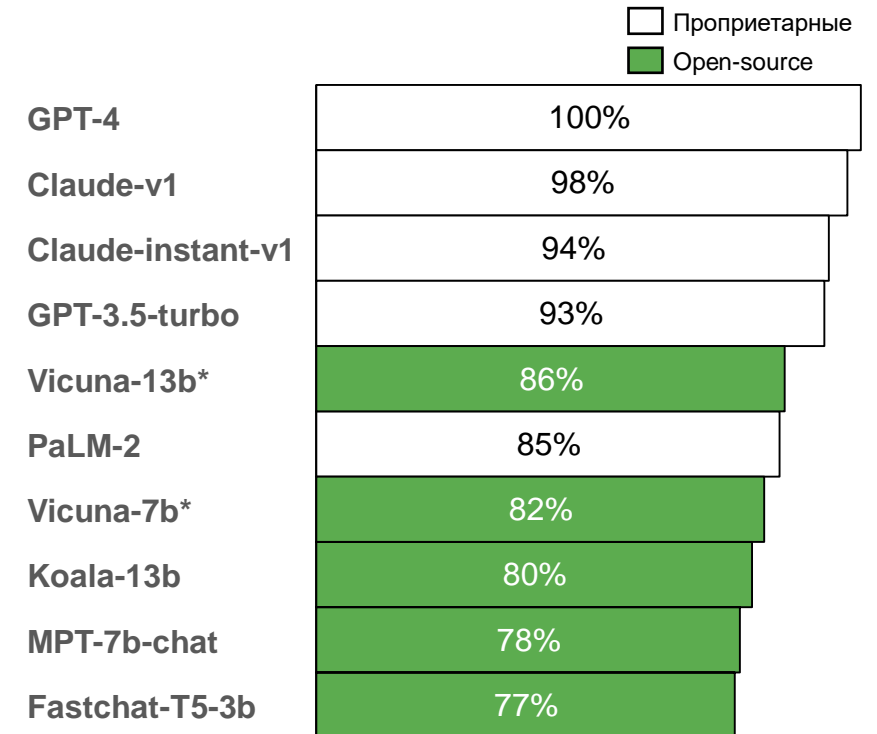
*На основе количества упоминаний в СМИ, пресс-релизах, форумах, анализ за март-май 2023 года

Open-source модели генеративного AI становятся сопоставимы по качеству с проприетарными моделями, снижают барьер для корпоративного использования

За счет использования готовых весов и тонкой донастройки open-source модели выдают сопоставимый результат с GPT и Palm (чат-бот Bard Google) и могут работать на простом пользовательском оборудовании.

- **Качество open-source моделей существенно уступало закрытым проприетарным сервисам.** Стоимость обучения собственной языковой модели, сопоставимой с GPT-3, составляет до 10 млн долларов США, занимает месяцы и требует доступа к десяткам тысяч GPU и тренировочным данным.
- **В марте 2023 года параметры (веса) большой языковой модели LLaMA стали доступны* в open-source.** На базе нее появилось множество OS моделей, результат которых пользователи не могут отличить от ChatGPT.
- **Эти модели могут работать даже на домашнем оборудовании: LLaMA модель запустили на ноутбуке, iPhone и даже на тостере.** При этом OS не требуют подписки, позволяют тонкую донастройку на данных компании без риска доступа к конфиденциальным данным.
- **Венчурные инвестиции поддерживают рост и развитие open-source.** Портал по разработке open-source LLM Huggingface привлек 160 млн долларов США, его текущая оценка – в 2 млрд долларов США. В июне 2023 года стартап Mistral AI по созданию OS моделей для корпоративных пользователей привлек 113 млн долларов США через четыре недели после основания.
- **OS-модели усложняют регулирование AI.** Их сложно отслеживать и в случае необходимости блокировать и накладывать штрафы. Они уже используются для имитации доступа к моделям OpenAI, их веса и настройки могут быть скорректированы для выдачи заведомо предвзятых или опасных результатов.

Сравнение качества моделей относительно GPT-4
Тестирование lmsys.org



* На базе LLaMA

1. Основные наблюдения
2. Инвестиции
3. Области применения генеративного AI
4. **Влияние на рынок труда**
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

Что говорят ведущие идеологи развития AI про рынок труда:

- **«Мы не понимаем, как они (LLM) работают, как повлияют на политику, экономику и рабочие места».**
Eric Schmidt, ex Google CEO, апрель 2023 года, слушания в конгрессе США
- **«AI по всему миру может заменить до 25% или 300 млн рабочих мест... Все технологические революции приводили к созданию большего количества рабочих мест, чем замещали».**
Goldman Sachs, март 2023 года
- **«История является плохим предсказателем будущих событий, в этот раз может быть по-другому».**
Gary Marcus, профессор нейрологии, NYU, май 2023 года, слушания в конгрессе США
- **«Языковые модели — это инструменты, а не существа. Они хорошо выполняют задачи, но не работу».**
Sam Altman, CEO Open AI, май 2023 года, слушания в конгрессе США



Автор рекламы на фото: агентство по подбору персонала IMPACT (Бельгия)

Развитие AI может повлиять на рынок труда в творческой сфере и образовании

В большинстве других областей AI увеличивает производительность когнитивных задач и создает спрос на новые компетенции, но не замещает человека. Рост занятости может быть сопоставим с ростом эффективности.

СОКРАЩЕНИЕ ЗАНЯТОСТИ

РОСТ ЗАНЯТОСТИ

Сокращение времени на выполнение когнитивных задач.
LLM – это калькулятор «когнитивной деятельности»

Рост времени на доработку и устранение последствий применения AI, обеспечение работы AI: приложения, инфраструктура, данные



% времени, экспертные оценки Stanford, Goldman Sachs, Б1

Профессии с наибольшим потенциальным влиянием от GAI* по количеству упоминаний в СМИ

Создание контента: художники, дизайнеры, писатели и музыканты

Влияние GAI на творческие профессии, которые традиционно считались защищенными от автоматизации, оказалось сильнейшим.

- **Спрос на художников в разработке и дизайне игр упал на 71%** (Китай, апрель 2023 года). Компании-разработчики игр полагаются на генеративные AI-модели для создания изображений и привлекают дизайнеров лишь для финальной доработки.
- **Сценаристы Голливуда ушли на забастовку с требованиями гарантий по защите рабочих мест и авторских прав на сценарии**, использовавшиеся при обучении больших языковых моделей.
- **На Amazon лавинообразное количество книг, написанных ChatGPT** (более 200 за февраль), и отзывов, написанных AI.
- **Spotify удалил «сотни тысяч» треков, созданных через сервис Voomy**. Эти треки проигрывались другими ботами, которые получали авторские отчисления.
- **Voomy утверждает, что его пользователи уже создали 14 млн треков**, или 13%* всей мировой музыки. Качество и популярность этих записей сервис не раскрывает.



*На май 2023 года

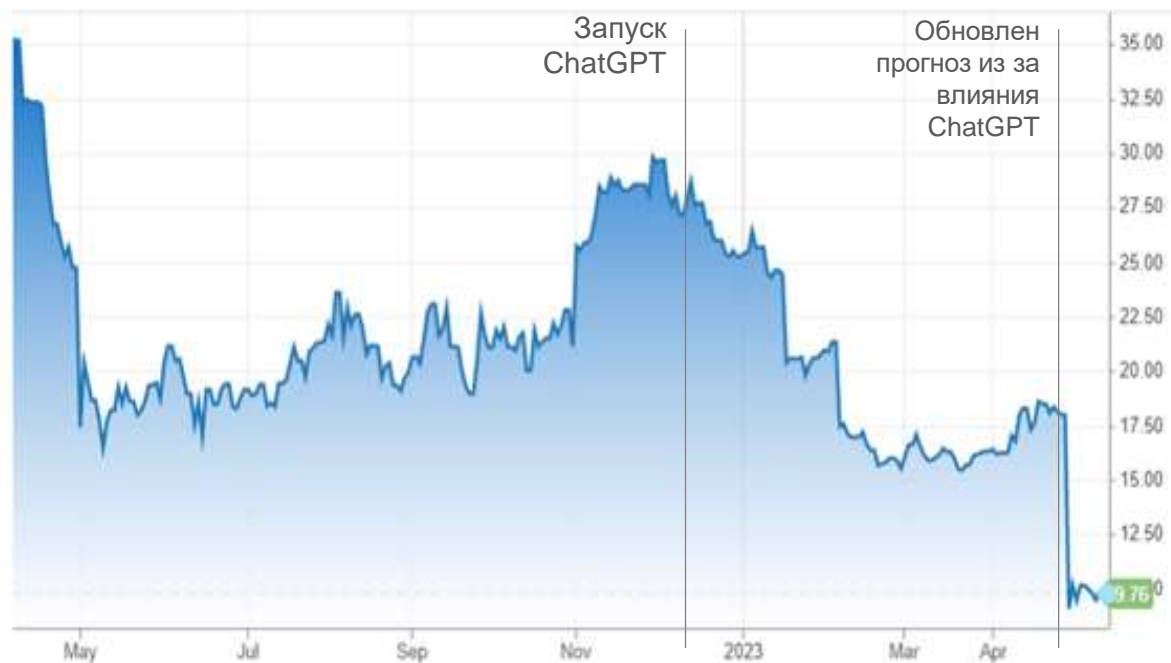
Преподаватели: из оратора в модератора и экзаменатора

Домашние задания стало сложнее использовать как инструмент проверки знаний, потому что ChatGPT применяется при их подготовке. Запретить или выявить его использование технически трудно. Акции компании Chegg Inc., специализирующейся на помощи в выполнении домашних заданий, упали в три раза после запуска ChatGPT.

- ▶ **Формат современного обучения, появившийся в XVIII веке, устарел с появлением интернета.** GAI подчеркнул это, став персонализированным репетитором по любому предмету. Более 50% учителей используют ChatGPT ежедневно*; 88% отметили позитивный эффект и экспериментируют с форматом: предлагают студентам использовать ChatGPT для генерации текста по теме, просят указать «галлюцинации» и предложить правильную версию.
- ▶ **Домашние задания почти потеряли смысл как инструмент проверки знаний — по всему миру их выполняет ChatGPT.** Запретить или определить использование технически сложно. Акции Chegg Inc., специализирующейся на помощи в выполнении домашних заданий, упали в три раза после запуска ChatGPT.
- ▶ **Роль преподавателей меняется.** Сложность вопросов учеников, а вместе с ней и требования к квалификации будут расти, а проверка знаний теперь требует большего персонального общения. Преподаватели становятся модераторами дискуссий, репетиторами, персональными экзаменаторами.

*Опрос преподавателей старшей школы в США

Стоимость акций Chegg Inc.

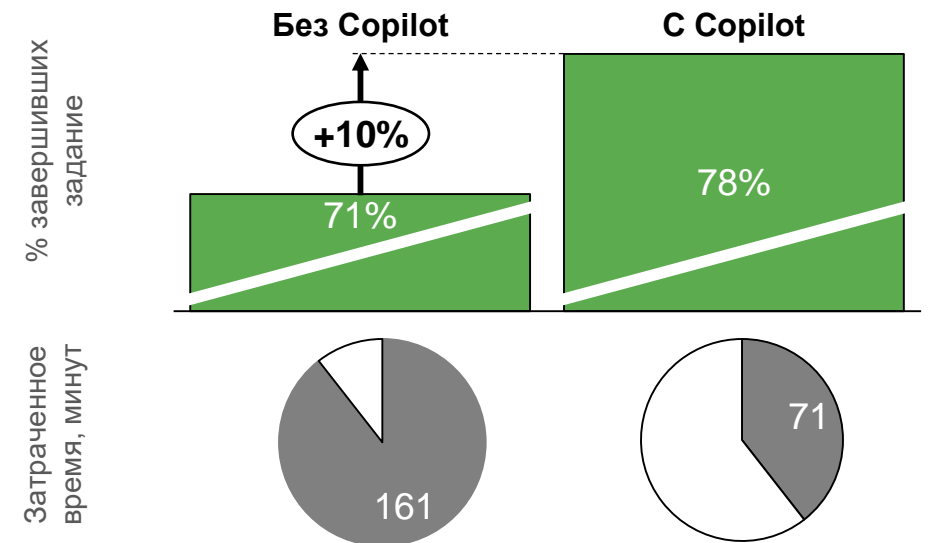


Разработка программного кода: «Co-pilot, not an Autopilot»

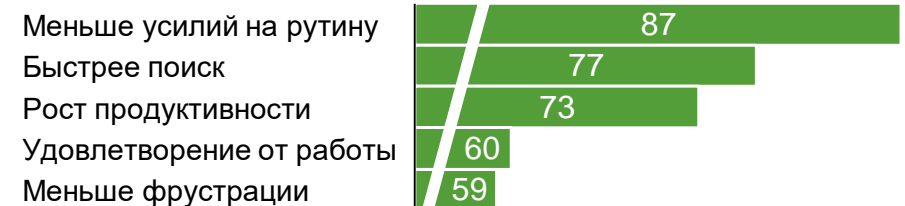
Разработчики ПО первыми получили доступ к GPT (в основе Copilot) на GitHub в 2021 году. Производительность и удовлетворенность выросли; к падению спроса на разработчиков не привело.

- ▶ **Большие языковые модели помогают разработчикам с 2021 года.** GitHub запустил тестовый доступ к Copilot на базе GPT-3 от OpenAI еще в 2021 году. Система предлагает код по уже написанной части или по комментариям. Amazon представил свою версию CodeWhisperer в июне 2022 года. Google заявил, что использует аналогичный инструмент для внутренних разработок.
- ▶ **Скорость разработки при использовании AI Copilot на GitHub выросла более чем в два раза.** AI также ускоряет поиск ошибок и автоматизирует тестирование, подготавливает проекты технической документации.
- ▶ **Запуск Copilot и других AI-помощников не уменьшил спрос на разработчиков.** AI плохо справляются с длинным кодом или сложными задачами: рефакторинг функций выдавал неработающий код в 70% случаев. В то же время AI хорошо справляются с «джуниор» задачами.
- ▶ **Распространение AI увеличит спрос на разработчиков,** особенно в области разработки и обучения нейронных сетей. А рост «любительского» кода потребует времени разработчиков на оптимизацию, интеграцию и доработки, в том числе для обеспечения безопасности.

Тест GitHub: написание JavaScript с AI (Copilot) и без



Отзывы от использования Copilot



*В 2022 году поданы иски против GitHub за нарушение авторских прав при обучении Copilot на коде других разработчиков (см. раздел «Регулирование»).

Источники: GitHub, Stateof.ai

1. Основные наблюдения
2. Инвестиции
3. Области применения генеративного AI
4. Влияние на рынок труда
5. **Новые «старые» риски**
6. Регулирование
7. Прогнозы

Генеративный AI создает новые типы рисков для человека, бизнеса и общества

Спектр рисков от быстрого развития и доступности генеративного AI растет. Наиболее острые связаны с дезинформацией, конфиденциальностью, достоверностью, нарушением прав на интеллектуальную собственность.

Риски по количеству упоминаний (март-июнь 2023 года)

Выборочные примеры

СОЦИАЛЬНЫЕ

- ▶ Дезинформация
- ▶ Влияние на рынок труда

Eric Schmidt* считает дезинформацию основным риском GAI, так как с его помощью можно легко подорвать социальные и финансовые институты.

ИНФОРМАЦИОННЫЕ и ОПЕРАЦИОННЫЕ

- ▶ «Галлюцинации»: создание вымышленных фактов
- ▶ Доступ к конфиденциальным данным запросов пользователей
- ▶ Рост спама и низкокачественного кода
- ▶ «Jailbreak»: взлом чат-ботов через серию запросов (промтлов)

Citigroup, Deutsche Bank, Goldman Sachs, Apple, Google запрещают сотрудникам использовать чат-боты, включая собственные, из-за риска ошибок и конфиденциальности данных.

ЮРИДИЧЕСКИЕ

- ▶ Нарушение прав интеллектуальной собственности
- ▶ Ответственность за ущерб от «галлюцинаций»
- ▶ Ответственность за несоблюдение требований при работе с персональными данными

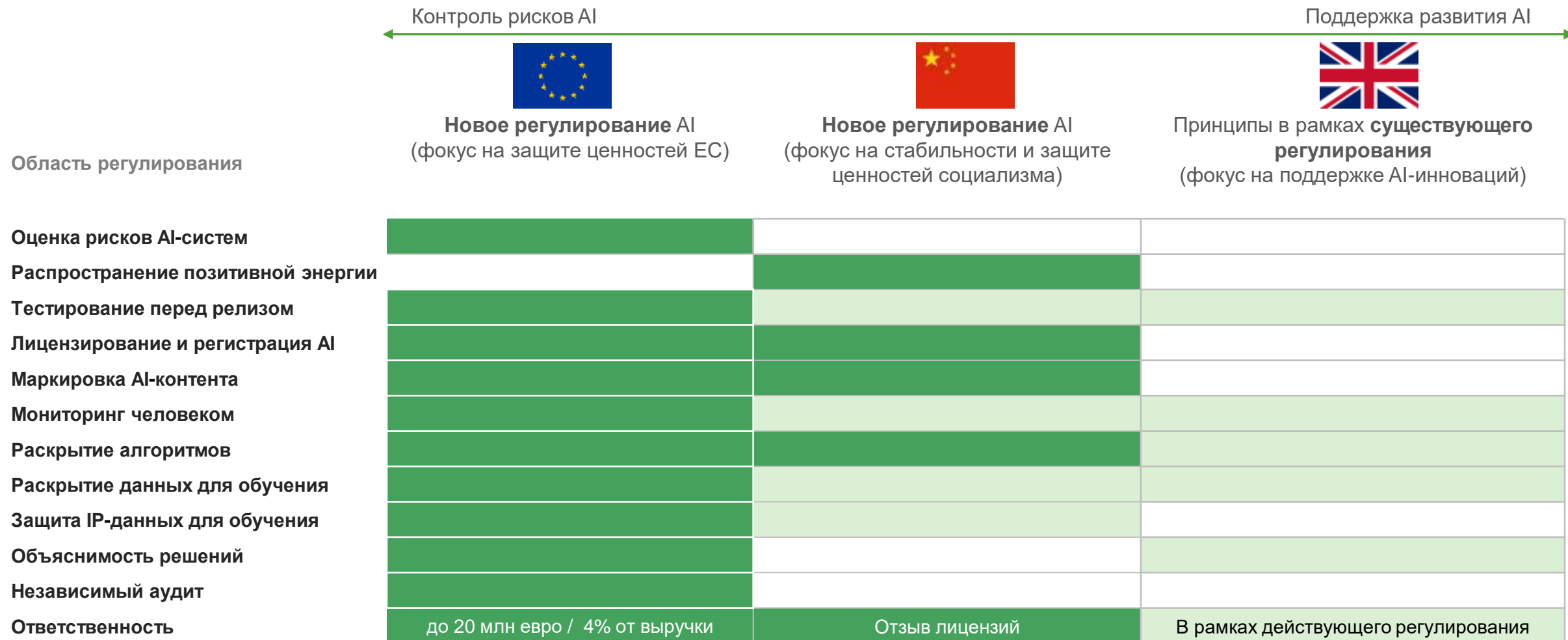
Рост исков к OpenAI от Google из-за генерации недостоверных данных в отношении конкретных людей и компаний.

* Бывший CEO Google, один из основных идеологов развития AI

1. Основные наблюдения
2. Инвестиции
3. Области применения генеративного AI
4. Влияние на рынок труда
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

Регуляторы пытаются снизить риски генеративного AI в условиях ограниченных данных

Рост использования GAI и связанных с ним рисков вызвал волну предложений по введению регулирования: от детальных требований к AI (ЕС, Китай, Канада) до дополнений к существующему секторальному регулированию с фокусом на поддержку развития AI (Великобритания). Все пока на стадии обсуждений.



Источники: europa.eu, oecd.ai, Stanford, lexology, Б1

1. Основные наблюдения
2. Инвестиции
3. Области применения генеративного AI
4. Влияние на рынок труда
5. Новые «старые» риски
6. Регулирование
7. Прогнозы

Прогнозы по трендам в области генеративного AI

- **Рост инвестиций в технологии, облегчающие корпоративное использование генеративного AI.** Эти технологии должны обеспечить конфиденциальность, прозрачность работы и соблюдение требований быстроразвивающегося регулирования.
- **Рост конкуренции сервисов генеративного AI и поиск подходов к их монетизации.** Десятки централизованных сервисов, сотни стартапов и тысячи независимых разработчиков, работающих над open-source моделями, увеличивают конкуренцию и давление на цены. Это происходит одновременно с ростом затрат на «обучение» моделей и неопределенностью в регулировании.
- **Нехватка и рост стоимости вычислительной инфраструктуры.** Количество моделей и сервисов вызывает спрос, который может превышать доступные мощности до 10 раз*. Это вызывает рост стоимости GPU, специализированных чипов, облачных AI сервисов и технологий оптимизации вычислений.
- **Рост инвестиций в технологии верификации контента.** Это необходимо для борьбы с дезинформацией, мошенническим контентом и спамом, стоимость создания которых благодаря AI стремится к нулю.
- **Рост количества инициатив по регулированию AI с целью снижения растущих рисков.** Учитывая экономический потенциал генеративного AI, регуляторы будут искать баланс между интересами пользователей, бизнеса, государства и поддержкой развития AI.

*По данным Andreessen Horowitz