

# Рынок ИИ-ускорителей

Обзор российского рынка GPU для задач в области ИИ



# Таксономия рынка ИИ и терминология



## ИИ-ускоритель

Специализированный аппаратный компонент, оптимизированный для высокопроизводительного выполнения матрично-тензорных операций, характерных для обучения и инференса моделей ИИ.

## GPU (Graphics Processing Unit)

Графический процессор с массово-параллельной архитектурой, который благодаря высокой пропускной способности и эффективности в параллельных вычислениях широко применяется как базовый ускоритель для задач ИИ и высокопроизводительных вычислений (High-Performance Computing).

## Сервер с GPU

Серверная платформа, оснащенная одним или несколькими GPU/ИИ-ускорителями, а также соответствующей подсистемой питания, охлаждения, памяти и ввода-вывода данных для выполнения ресурсоемких задач.

## ПАК для ИИ

Готовое прединтегрированное решение, включающее GPU-серверы (и/или кластеры), системное ПО, драйверы, библиотеки ускорения и средства управления, предоставляемое как единый продукт для ускоренного развертывания и эксплуатации ИИ-контуров.

## Инференс

Применение обученной модели в продуктивной среде, при которой модель обрабатывает новые входные данные и формирует результат (прогноз, классификацию, генерацию) без изменения своих параметров.

## Обучение

Процесс оптимизации параметров модели по обучающим данным, в ходе которого веса обновляются для минимизации функции потерь и повышения качества решения задачи.

## GPUaaS (GPU-as-a-Service)

Сервисная модель предоставления GPU-вычислений по требованию, при которой клиент арендует GPU-ресурсы на нужный срок, оплачивая потребление вместо покупки и обслуживания оборудования.

## AI PaaS (AI Platform-as-a-Service)

Платформа для разработки и эксплуатации ИИ, предоставляющая управляемые инструменты и сервисы (среда интерактивных вычислений, пайплайны, инструменты MLOps для управления жизненным циклом моделей, развертывание моделей и мониторинг), при этом инфраструктура с GPU скрыта за интерфейсами платформы и управляется провайдером.

## Bare Metal

Выделенный физический сервер без слоя виртуализации, предоставляющий прямой доступ к аппаратным ресурсам и обеспечивающий максимальную предсказуемость производительности для высоконагруженных вычислений.

## LLM (Large Language Model)

Большая языковая модель — нейросетевая модель с большим числом параметров, обученная на массивных текстовых корпусах для понимания и генерации естественного языка и решения связанных задач.

# Рост глобальных расходов на ИИ



- + ИИ перешел из категории экспериментальных технологий в класс критически важных бизнес-активов. На этом фоне растут инвестиции в ИИ — от данных и платформ разработки до специализированной инфраструктуры.
- + Ключевым вычислительным ресурсом для современных ИИ-нагрузок стали GPU-ускорители. Спрос на GPU поддерживается переходом организаций к промышленной эксплуатации ИИ: от обучения и донастройки больших языковых моделей до инференса в продуктивных контурах, включая интеллектуальных ассистентов, рекомендательные механизмы, компьютерное зрение и обработку естественного языка.
- + Экономика ИИ-инфраструктуры в текущем цикле развития характеризуется высокой концентрацией капитальных затрат в контуре ИИ-ускорителей (GPU и др. решения). На уровне сервера, по оценке UBS, свыше 70% стоимости приходится на GPU, тогда как доля памяти составляет 15%.<sup>1</sup>
- + Динамика рынка указывает на структурный разрыв между темпами роста потребления вычислительных мощностей и скоростью расширения производственных возможностей и цепочек поставок. В среднесрочной перспективе это формирует устойчивое давление на рынок аппаратной инфраструктуры для ИИ и усиливает конкуренцию за доступ к ускорителям, системам памяти и компонентам уровня ЦОД.
- + На стороне предложения основными драйверами остаются крупнейшие ИТ-компании и операторы облачных платформ, которые ускоряют модернизацию существующих мощностей и ввод новых объектов.
- + Глобальный рынок ИИ-ускорителей остается сверхконцентрированным. Американская компания Nvidia контролирует около 80% рынка ИИ-чипов<sup>3</sup>, а, по данным IDC, в сегменте серверов со встроенными GPU ее доля поставок превышала 90%.<sup>6</sup>

<sup>1</sup>UBS. Artificial intelligence: Sizing and seizing the investment opportunity. 10.06.2024. URL: <https://clck.ru/3S6oQD>

<sup>2</sup>Gartner. January 2026.

URL: <https://www.statista.com/statistics/1624686/ai-spending-across-it-markets-worldwide/>

<sup>3</sup>Statista. AI chips. 2025. URL: <https://www.statista.com/study/71650/artificial-intelligence-ai-chips/>

<sup>4</sup>Рост в 15 раз вызван мощным бумом вложений в 2025 г. и эффектом низкой базы на начальных этапах

<sup>5</sup>Reuters. Nvidia says its next-generation AI chip platform to be rolled out in 2026. 2.06.2024.

URL: <https://www.reuters.com/technology/nvidia-says-its-next-generation-ai-chip-platform-be-rolled-out-2026-2024-06-02/>

<sup>6</sup>IDC. Worldwide Quarterly Server Tracker. 18.03.2025. URL: <https://my.idc.com/getdoc.jsp?containerId=prUS53264225>

## Рост глобальных расходов на ИИ %

Источник: расчеты основаны на данных Gartner<sup>2</sup>

2025	2026	2027	
69%	34%	29%	Услуги ИИ
	98%	67%	Кибербезопасность ИИ
238%	60%	41%	Программное обеспечение ИИ
152%	83%	65%	Модели ИИ
	42%	43%	Платформы для обработки данных и машинного обучения
	28%	30%	Платформы для разработки ИИ
	277%	106%	Данные для ИИ
x15 <sup>4</sup>	42%	28%	Инфраструктура для ИИ
78%	44%	32%	Совокупные расходы на ИИ

## Объем глобального рынка ИИ-чипов \$ млрд

Источник: Statista Market Insights<sup>3</sup>



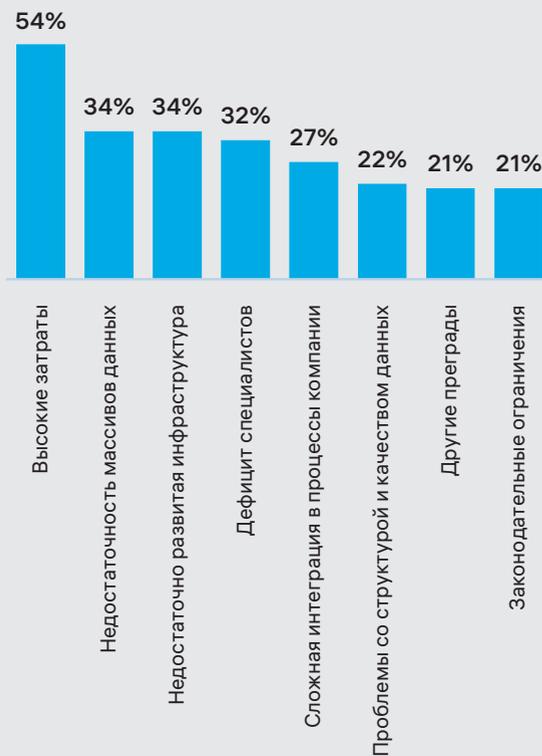
# Текущая потребность в вычислительной инфраструктуре для ИИ в РФ



- + Ключевыми барьерами для внедрения ИИ являются высокие затраты, необходимость развития инфраструктуры, а также недостаточность массивов данных.
- + Высокие затраты включают не только стоимость лицензий и разработки, но прежде всего расходы на вычислительную инфраструктуру: GPU-ускорители, серверные мощности, системы хранения данных, сетевую связность и эксплуатацию.
- + На практике для многих компаний именно доступ к высокопроизводительным вычислительным ресурсам становится критическим фактором, сдерживающим переход от пилотных проектов к промышленному внедрению ИИ.
- + Особую роль играет ограниченная доступность физических GPU и высокая стоимость построения собственной инфраструктуры. По данным НЦРИИ, только 9% организаций в РФ можно считать полностью обеспеченными вычислительной инфраструктурой для задач ИИ.<sup>1</sup>
- + Спрос на вычислительные ресурсы для ИИ в наибольшей степени формируется крупными компаниями и экосистемными игроками, которые масштабируют ИИ-решения в клиентских и операционных процессах. При этом постепенно растет и спрос со стороны небольших ИИ-разработчиков и средних компаний.
- + Отечественный рынок ИИ-инфраструктуры характеризуется высокой зависимостью от западных GPU. При этом, ограничения в поставках остаются одним из основных факторов, сдерживающих развитие рынка. На ситуацию одновременно влияют дефицит GPU на глобальном рынке и усиление экспортного контроля США.

## Барьеры при внедрении ИИ

Доля российских организаций, ощутивших барьеры для внедрения ИИ  
Источник: оценка ИТ-холдинга Т1 по данным Росстата 2025 г. по более 260 000 компаний



## Обеспеченность вычислительной инфраструктурой для ИИ

Источник: НЦРИИ, 2025

# 9%

Полностью обеспечены вычислительной инфраструктурой



# 40%

Частично обеспечены вычислительной инфраструктурой



# 51%

Имеют потребность в вычислительной инфраструктуре



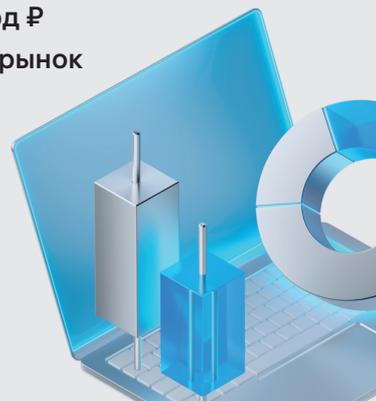
<sup>1</sup>НЦРИИ. Индекс готовности приоритетных отраслей экономики Российской Федерации к внедрению ИИ. 2025. URL: <https://ict.moscow/projects/ai/research/index-gotovnosti-prioritetnykh-otraslei-ekonomiki-rossiiskoi-federatsii-k-vnedreniiu-iskusstvennogo-intellekta/>

# Отечественный рынок ИИ-ускорителей



- + По текущей оценке ИТ-холдинга Т1, объем российского рынка ИИ-ускорителей в 2025 году составил порядка 63 млрд руб., при этом значительная часть поставок приходится на крупнейших экосистемных и корпоративных игроков, а также на государственный сектор.
- + С точки зрения фактического потребления, рынок развивается в двух основных контурах: коммерческие дата-центры и локальные развертывания. В эквиваленте Nvidia A100 в дата-центрах используется более 10 000 карт GPU, а в локальной инфраструктуре — порядка 8 000 GPU.
- + Прогнозная динамика до 2030 года отражает ускоренное наращивание вычислительной базы. Траектория обусловлена не только запуском новых ИИ-проектов, но и повторными закупками, дооснащением действующей инфраструктуры и расширением использования ускорителей в продуктивных сценариях — от обучения моделей до промышленного инференса.
- + Развитие рынка сдерживается рядом факторов. Ключевые ограничения — монопольное положение Nvidia и связанной технологической экосистемы, нестабильность поставок, высокая стоимость ускорителей и сопутствующей инфраструктуры.
- + Конкуренция на рынке ИИ-инфраструктуры разворачивается сразу на трех уровнях: ускоритель, серверная платформа, ПАК для ИИ.
- + Отечественная отрасль активно ищет альтернативы: компании рассматривают китайские карты и отечественные нейроускорители как часть стратегии снижения зависимости от западных вендоров.

**63** млрд ₽  
российский рынок  
GPU в 2025

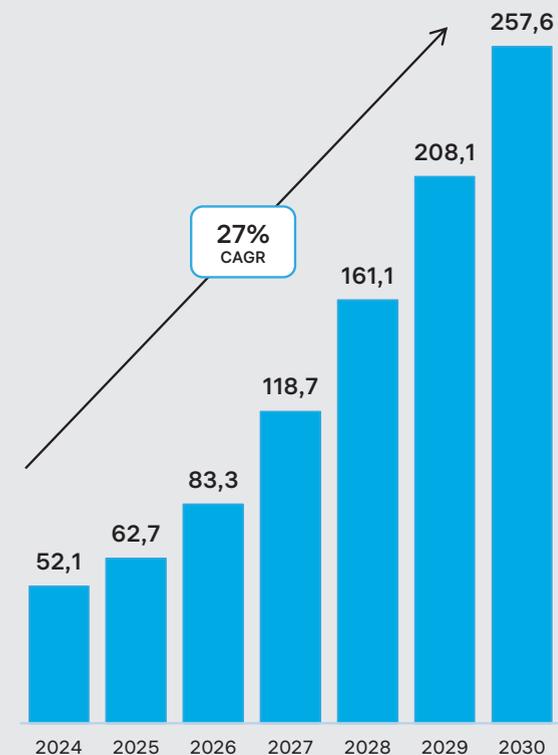


**80%**  
поставки крупнейшим  
экосистемным игрокам



Объем российского рынка  
ИИ-ускорителей,  
₽ млрд

Источник: собственная оценка ИТ-холдинга Т1



# Российский рынок облачных сервисов с GPU



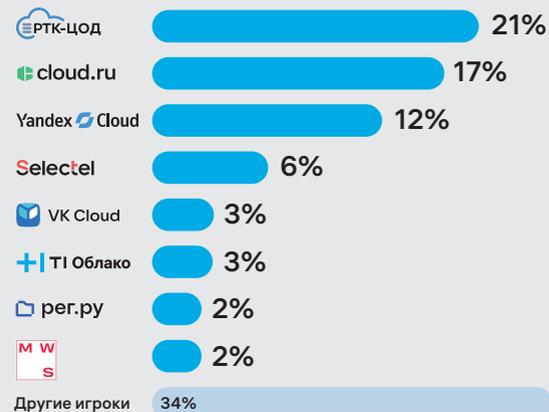
- + Широкое внедрение ИИ-сервисов, логистические сложности импорта ИИ-ускорителей и другие факторы повысили спрос на облачные вычисления с GPU и платформы для ИИ как сервис (AI PaaS).
- + В рамках российского рынка облачных GPU, согласно исследованию Apple Hills Digital, быстрее всего растет сегмент «Инфраструктура как сервис» (IaaS) — 45% за 2025 год, тогда как «Платформы для ИИ как сервис» вырос на 35%.<sup>1</sup>
- + В ближайшие годы, по собственной оценке ИТ-холдинга Т1, российский рынок облачных сервисов с GPU будет расти с совокупным среднегодовым темпом роста в 31% с потенциалом достижения 85,5 млрд руб. в 2030 г.
- + Ключевыми арендаторами выделенных физических серверов на рынке РФ являются ИТ-компании, а также крупнейшие игроки из отраслей с высоким уровнем проникновения ИИ. Согласно данным ИТ-холдинга Т1, в их число входят банки, ритейл, страхование, наука.
- + Несмотря на наличие у некоторых крупных компаний собственных GPU-ускорителей, часть из них тестируют новые GPU на базе облачных сервисов или проверяют их эффективность по отношению к отдельным задачам.
- + Компании, которым важно сохранить контроль над физической инфраструктурой и соответствовать повышенным требованиям к безопасности, арендуют инфраструктуру по модели «Оборудование как услуга». По ней провайдер передает клиенту оборудование в аренду для размещения на площадке заказчика и отвечает за его обслуживание и обновление.
- + Практически все игроки предоставляют сервисы на базе GPU Nvidia, а именно модели A100, V100. У некоторых облаков появляются и альтернативные ускорители.

**17** млрд ₽  
российский рынок GPUaaS в 2025



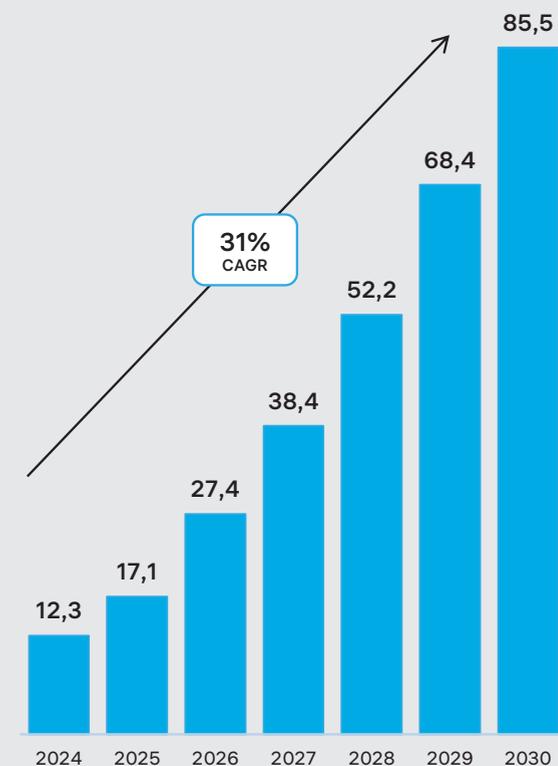
## Распределение игроков рынка облачных сервисов с GPU Выручка от GPU-сервисов в 2024 г.

Источник: собственная оценка ИТ-холдинга Т1 на основе анализа финансовых показателей компаний и открытой информации



## Объем российского рынка облачных сервисов с GPU, ₽ млрд

Источник: собственная оценка ИТ-холдинга Т1



<sup>1</sup> Apple Hills Digital. Октябрь 2025. URL: <https://apple-hills.com/ru>

# Факторы влияния на рынок



	Сильное влияние	Среднее влияние	Низкое влияние
 Регулирование	Ужесточение политики импортозамещения оборудования в госзакупках	Ужесточение требований к ЦОД как к объектам КИИ Развитие законодательства в области ИИ	Изменения в регулировании закупок санкционного оборудования Законы о локализации как стимул развития ЦОД
 Геополитика	Санкции и экспортный контроль GPU-поставок	Ускоренные программы финансирования ЦОД Увеличение объема госзакупок GPU	ИИ-суверенитет как основа гос. суверенитета Усиление глобальной конкуренции за технологии
 Технологии	Появление и развитие альтернативных GPU	Использование устаревших GPU из-за санкций Зависимость от CUDA	Развитие децентрализованных вычислительных сетей для ИИ
 Экономика	Серый импорт западных GPU	Макроэкономические риски: валютная волатильность, монетарная политика и др. Развитие рынка облачных сервисов	Государственные инвестиции в ИИ/ЦОД Инвестиции в ИИ/ЦОД крупных корпоративных игроков
 Социум	Дефицит высококвалифицированных ИИ-специалистов	Развитие программ подготовки и переподготовки специалистов Растущий спрос на ИИ-приложения на русском и языках народов РФ	Рост эффективности использования GPU через развитие ПО с открытым исходным кодом Общественный скептицизм к технологиям ИИ
 Энергия		Влияние энерготарифов и инфраструктурных затрат на совокупную стоимость владения (ТСО)	Локализация ЦОД в регионах с дешевой и стабильной энергией

# Государственная политика в области ИИ и ЦОД



## Национальная стратегия развития искусственного интеллекта до 2030 г.

Ключевой документ, определяющий:

- + Основные принципы использования и развития технологий ИИ
- + Основные задачи развития ИИ и показатели по достижению целей
- + Приоритетные направления развития и использования технологий ИИ

## Федеральный проект «Искусственный интеллект»

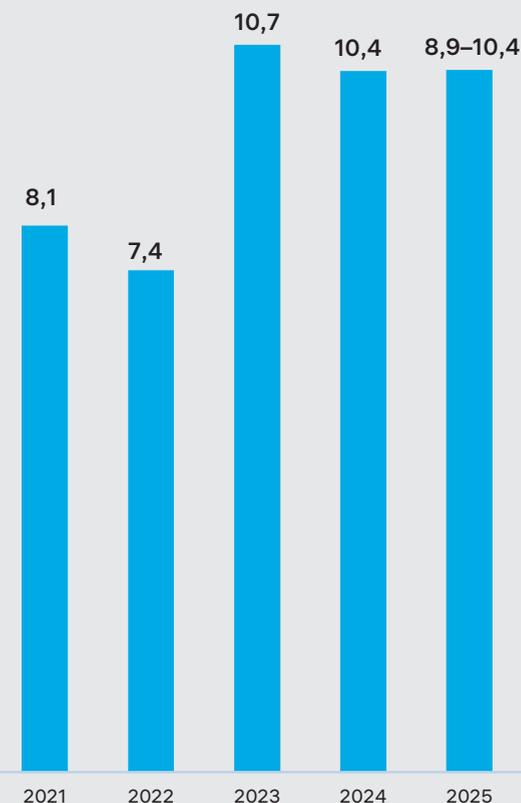
Часть нацпроекта «Экономика данных и цифровая трансформация государства»:

Ожидаемые результаты:

- + Стимулирование научных исследований в области ИИ
- + Развитие ИИ-решений в госуправлении
- + Совершенствование системы подготовки кадров и поддержки разработчиков

## Российские государственные инвестиции в ИИ, ₽ млрд

Источник: МТС<sup>2</sup>



**>70** млрд ₽

планируется выделить на развитие ИИ в России до 2030 г.

**60** млрд ₽

затраты гос. организаций РФ в 2024 году на строительство ЦОД

**32,1** млрд ₽

планируется выделить в 2025–2030 гг. на льготное кредитование строительства ЦОД

## В 2024 году в эксплуатацию введены ЦОД:<sup>1</sup>

- + Росатом
- + МВД России
- + Минприроды России
- + ФНС России
- + ФМС России
- + Правительство Москвы

## К 2036 году в России должно появиться 8 новых дата-центров, согласно плану Правительства РФ<sup>3</sup>

- + 2025 г. «Удомля-3»
- + 2026 г. «Санкт-Петербург»
- + 2026 г. «Кей Поинт Екатеринбург»
- + 2026 г. «Кей Поинт Санкт-Петербург»
- + 2026 г. «Кей Поинт Ростов-на-Дону»
- + 2029 г. «Юрловский»
- + 2030 г. «Иннополис-2»
- + 2031 г. «Кселент-2»

<sup>1</sup>Ведомости. Госсектор занял половину рынка по строительству дата-центров. 23.12.2024.

URL: <https://www.vedomosti.ru/technology/articles/2024/12/23/1083022-gossektor-zanyal-pолоvinu-rinka-po-stroitelstvu-data-tsentrov>

<sup>2</sup>МТС. ИИ-агенты. Обзор мирового и российского рынков. URL: [d-economy.ru/wp-content/uploads/2026/02/mts-research-insights-ai-agents-2025.pdf](https://d-economy.ru/wp-content/uploads/2026/02/mts-research-insights-ai-agents-2025.pdf)

<sup>3</sup>Правительство РФ. Распоряжение от 29 августа 2025 г. № 2365-п. URL: <http://static.government.ru/media/files/rPcu24jtAayWAV30XBiqHdM5qherWYKe.pdf>



## Рынок ИИ-ускорителей

## Рынок облачных сервисов с GPU

### Периметр

Под рынком ИИ-ускорителей понимаются поставки ускорителей и решений на их основе, используемых для задач ИИ. Оценка включает поставки ускорителей как отдельных изделий для установки в инфраструктуру, а также ускорителей в составе серверов и ПАК для ИИ. Денежные показатели приводятся в текущих ценах соответствующего года.

Рынок облачных сервисов с GPU включает услуги предоставления GPU-ресурсов в облаке (GPU-инфраструктура как услуга: виртуальные машины с GPU, выделенные сервера с GPU, управляемые GPU-кластеры), а также связанные сервисы по потреблению GPU-мощностей в составе облачной инфраструктуры платформы для ИИ.

### Объем рынка

Оценка объема рынка ИИ-ускорителей в 2023–2024 гг. выполнена на основе статистики Росстата по затратам организаций на цифровые технологии (форма № 3-информ). Оценка калибровалась экспертно с учетом структуры спроса на ускорители по ключевым сценариям.

Оценка объема рынка в 2025 году уточнялась с использованием статистической информации по инвестициям в программное обеспечение и ИКТ-оборудование за 3 квартала, макроэкономических индикаторов (динамика инвестиций, ценовая конъюнктура и др.) и оценки уровня проникновения ИИ-технологий в отраслях.

При расчете учитывались сдвиги в структуре спроса: переход от пилотных проектов к промышленной эксплуатации, рост доли инференса и увеличение потребности в инфраструктуре для корпоративных контуров.

Оценка рынка облачных сервисов с GPU рассчитана методом «снизу-вверх» на основе анализа выручки ключевых игроков и оценки доли выручки от сегмента облачных GPU-услуг.

### Прогноз

Прогнозные значения формируются как оценка емкости рынка, включая отложенный спрос. Модель учитывает ожидаемую динамику цен и стоимость компонентов, рост потребности в вычислительных мощностях вследствие расширения применения ИИ, а также ограничения и инерцию масштабирования (поставка оборудования, ввод мощностей, эксплуатационные ограничения). Прогноз отражает потенциальную потребность в инфраструктуре при сохранении тренда промышленного внедрения ИИ.

Прогноз рынка GPU-облаков опирается на ожидаемую динамику проникновения ИИ и облачных технологий в организациях, прогноз уровня цен, а также прогноз количества организаций и потенциальной базы потребителей.

В расчете учитывается переход части спроса к сервисной модели потребления и рост спроса со стороны сегмента СМБ. Прогнозные значения интерпретируются как оценка емкости рынка с учетом роста потребления ИИ-нагрузок и коммерциализации корпоративных сценариев.





[info@t1.ru](mailto:info@t1.ru)

+7 (495) 727-09-85

[t1.ru](http://t1.ru)