Yandex Cloud

# Аспекты безопасности данных в Yandex Foundation Models

## О сервисе Yandex Foundation Models

Foundation Models

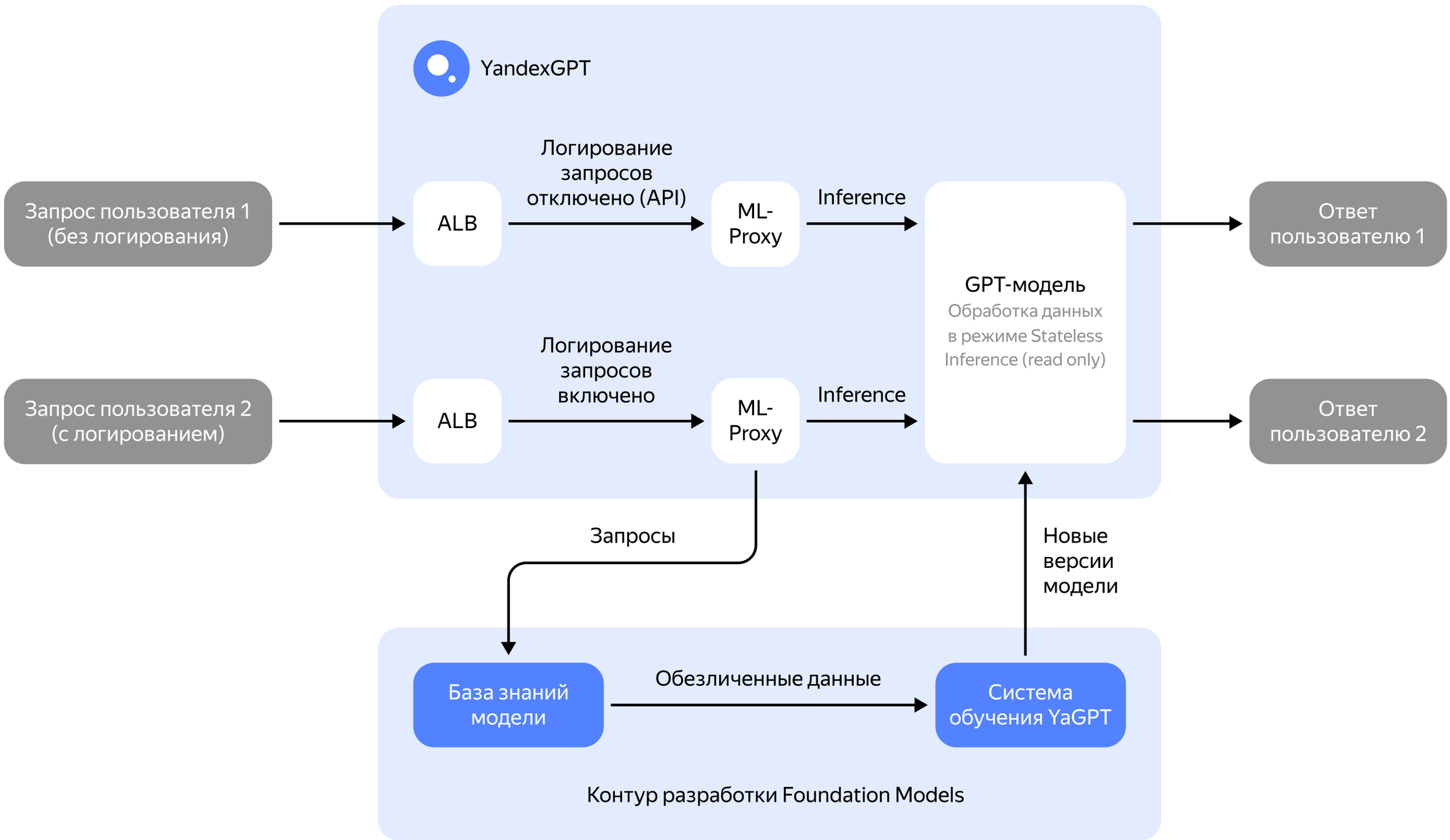
**Yandex Foundation Models** — это сервис фундаментальных моделей машинного обучения для бизнеса. Yandex Cloud предоставляет доступ к нейросетям YandexGPT, YandexART и других моделей, которые позволят вам использовать возможности генеративных моделей в ваших веб-сервисах и бизнес-приложениях. Сервис будет полезен всем, кто ищет способы ускорить развитие бизнеса с  с помощью технологий машинного обучения.

Все операции проверяются в сервисе Yandex Identity and Access Management (IAM). Если у субъекта нет необходимых ролей и прав, сервис выдаст ошибку.

Подробнее о сервисе читайте в документации.

Сервис **Yandex Foundation Models** прошёл независимый аудит безопасности. Это заключение подтверждает, что при работе с компонентами сервиса обеспечено соответствие требованиям федерального закона от 27 июля 2006 года № 152-ФЗ «О персональных данных» согласно приказу ФСТЭК России от 18 февраля 2013 года № 21.

Подробнее




# Как обрабатываются данные в YandexGPT

При работе с данными пользователя существует несколько **принципов безопасности,** которых придерживается сервис:

1. **Изоляция пользовательских данных**

   Данные различных пользователей изолированы на логическом уровне во всех компонентах сервиса от Application Load Balancer и до баз данных, где обрабатываются данные клиента.

2. **Шифрование и защита данных**

   Обработка данных в сервисе обеспечивается с учётом требований безопасности, включающих обязательное шифрование данных, резервирование данных, управление уязвимостями и реагирования на инциденты. Все процессы обеспечиваются в соответствии с требованиями 21-го приказа ФСТЭК и в соответствии с регламентами безопасности Yandex Cloud.

3. **Анонимизация данных при обучении модели**

   Обучение модели происходит не в реальном времени, а после тщательной подготовки данных. При обработке запросов и их подготовке к обучению модели, данные проходят обязательную анонимизацию, форматирование и очистку и получают отдельную разметку для корректного обучения и улучшения больших моделей.

4. **Работа только в режиме Stateless Inference**

   В сервисе YandexGPT существует два режима работы
   с пользовательскими данными:

   **Режим обучения** — процесс создания и оптимизации модели с использованием обучающих данных. Для подготовки обучающих данных, необходимо провести серию преобразований и классификаций, которые проводятся выделенной командой обучения модели.

   **Режим запроса (Inference)** — это процесс, в котором обученная модель применяется к живым данным для вывода результата. Система принимает входные данные от конечных пользователей, обрабатывает их в режиме Stateless Inference, передаёт в модель и возвращает выходные данные пользователям, не сохраняя данные запроса (режим аквариумной рыбки).

   Пользовательские данные по умолчанию логируются для дальнейшей обработки командой обучения YandexGPT. Эту опцию можно отключить при работе через API.

   При этом процесс обучения не зависит от обработки запросов (промтов) пользователей и логически разделён, а при отключении опции логирования запросов пользователей пользовательские данные нигде не сохраняются.
   Такой подход позволяет упрощать поддержку и реализацию модели и требует, чтобы входные данные были автономными.

## Что такое Stateless Inference?

В переводе с английского заключение и логический вывод без изменения состояния в машинном обучении означает, что каждый ответ модели выполняется независимо, без учёта каких-либо предыдущих входных или выходных данных.
Ключевыми аспектами вывода  без состояния являются:

- **Отсутствие долгосрочного хранения данных в памяти:** каждый ответ выполняется изолированно, основываясь исключительно на входных характеристиках.

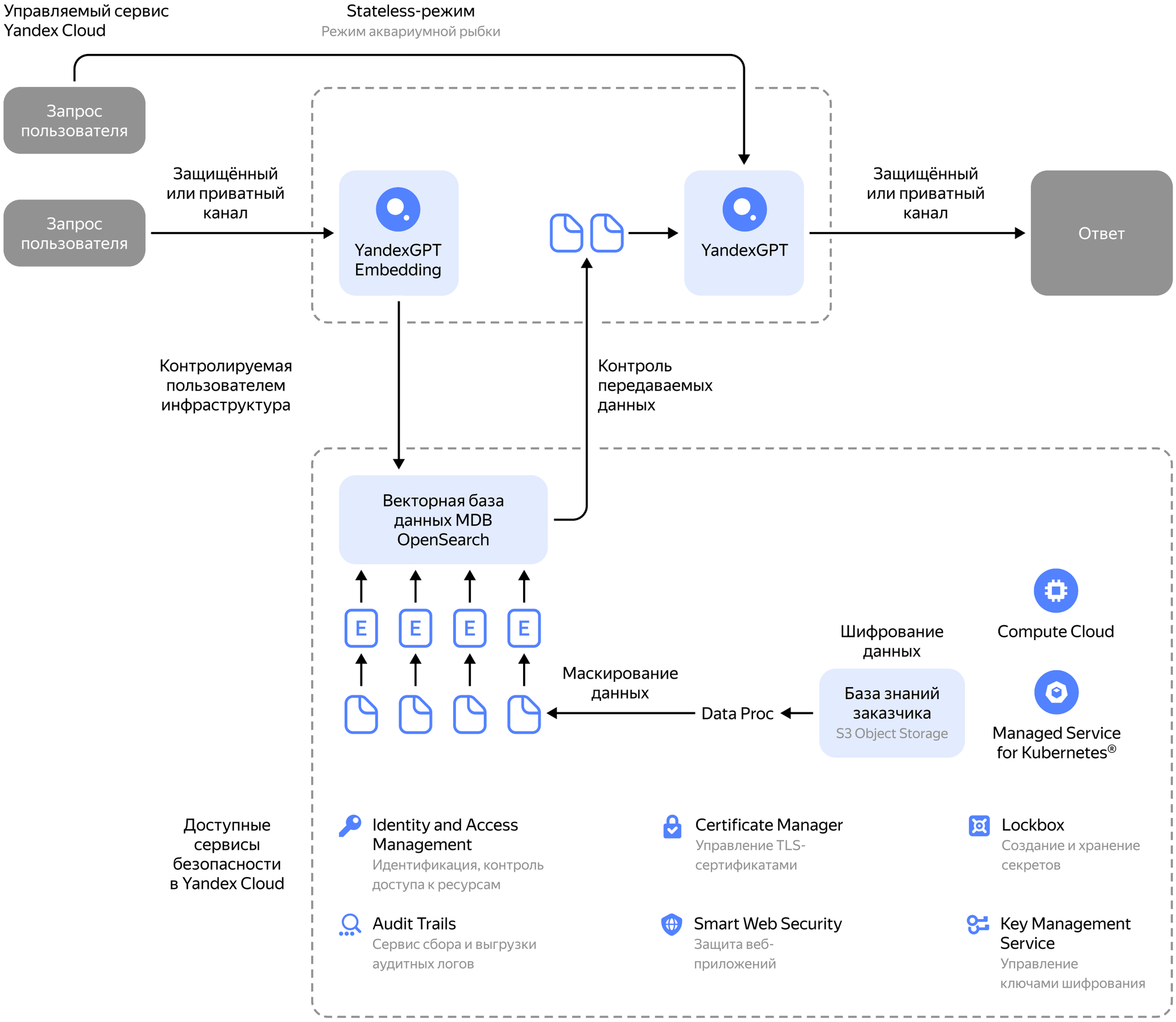
- **Независимые ответы:** каждый ввод обрабатывается независимо, как если бы это было первое взаимодействие. Модель не знает ни о каких предыдущих вводах или результатах и не полагается на них.

- **Упрощённая реализация:** отдельный механизм управления состоянием модели, упрощает работу с независимыми входными данными и позволяет их эффективно изолировать.

# Основные компоненты для создания архитектуры вопросно-ответных систем (RAG) с использованием YandexGPT

При создании архитектуры вопросно-ответных систем (Retrieval Augmented Generation, RAG) на базе YandexGPT используются следующие основные компоненты:

- **Первичная база знаний**. Хранится в S3-объектном хранилище.
- **Векторное представление первичной базы знаний**. Хранится в MDB OpenSearch или ClickHouse®.
- **YandexGPT Embeddings** — преобразует запросы пользователей и фрагменты текстовой базы знаний (чанки) в вектора-эмбеддинги.
- **YandexGPT** — большая языковая модель, которая, получая на вход установку на решение задачи (системный промт) дополнительно к запросу пользователя и фрагментам из базы знаний, отвечает на запрос пользователя.



## Типовой процесс обработки данных при создании архитектуры вопросно-ответных систем (RAG) с использованием YandexGPT

**На первом этапе** происходит преобразование первичной базы знаний в векторную форму для реализации алгоритма векторного поиска.

**На втором этапе** происходит регулярное общение пользователя с вопросно-ответной системой:

- Запрос пользователя преобразуется в вектор-эмбеддинг.
- В векторной базе данных происходит поиск близких к вектору запроса пользователя векторов фрагментов базы знаний.
- Преобразованные в текстовую форму вектора фрагменты базы знаний вместе с оригинальным запросом пользователя и установкой на решение задачи в виде системного промта подаются на вход YandexGPT.
- Модель YandexGPT обрабатывает запрос и выдаёт ответ со ссылками на источники информации.

## Совместная ответственность при создании архитектуры вопросно-ответных систем (RAG) с использованием YandexGPT

Обеспечение безопасности в публичном облаке — совместная задача пользователя и самой платформы. Она включает и работу с компонентами Retrieval Augmented Generation на базе YandexGPT, где с точки зрения ответственности за обеспечение безопасности можно выделить:

- **Управляемые компоненты**, где безопасность обеспечивается специалистами Yandex Cloud.
- **Инфраструктуру заказчика**, в которой права доступа и функции обеспечения безопасности полностью управляются и обеспечиваются пользователем.

Ваша база данных — это отдельный компонент всей системы. Данные в ней выделены в отдельную инфраструктуру, доступ к которой клиенты контролируют самостоятельно с применением необходимых им инструментов безопасности (см. схему на стр. 2). Размещаемые в этой инфраструктуре данные не становятся частью основной модели YandexGPT и не используются в ответах внешних (публичных) обращений к YandexGPT.

Подробнее о концепции разделения ответственности можно прочитать на сайте.

Разделение ответственности подразумевает, что физическая безопасность, безопасность сервисов и доступность всей инфраструктуры — это ответственность специалистов облака, которую они не могут переложить на клиентов. Клиент должен контролировать права доступа к ресурсам, например к виртуальным машинам. Как в локальных, так и в облачных моделях компании сами несут ответственность.

**При этом при взаимодействии с YandexGPT API специалисты Yandex Cloud обеспечивают:**

- ✅ Stateless-режим — возможность исключения хранения данных клиента.
- ✅ Изоляцию данных на нескольких уровнях.
- ✅ Обязательное шифрование данных при хранении и передаче.
- ✅ Строгий контроль доступа сотрудников к управляемым сервисам.
- ✅ Построение процессов сбора и обработки событий безопасности на базе SOC Yandex Cloud.

Подробнее о безопасности платформы Yandex Cloud можно прочитать здесь.

Подробнее об изоляции данных — здесь.

## Рекомендации по обеспечению безопасности архитектуры вопросно-ответных систем (RAG) с использованием YandexGPT

- Для исключения хранения данных из запросов пользователей отключите логирование обращений к YandexGPT с помощью этой инструкции.
- Используйте методы деперсонализации/маскирования чувствительных данных перед обращением в YandexGPT. Например, с помощью сервиса Data Transfer.
- Создайте приватное подключение между вашей инфраструктурой и Yandex Cloud с помощью Interconnect или используйте VPN.
- При обработке и хранении данных в виртуальной среде следуйте рекомендациям по безопасной конфигурации.
- Для защиты критичных данных в Yandex Object Storage используйте шифрование бакета на стороне сервера с помощью ключей Yandex Key Management Service (server-side encryption). Подробнее смотрите в разделе Шифрование в Object Storage.
- Для критичных виртуальных машин настройте шифрование диска с помощью KMS и удостоверьтесь, что оно включено по инструкции Шифрование диска виртуальных машин с помощью KMS.
- При использовании Kubernetes воспользуйтесь чек-листом безопасности.

## Стандарт по защите облачной инфраструктуры

Собрали рекомендации по техническим мерам защиты, которые помогут обеспечить безопасность при развёртывании информационных систем. Здесь вы найдёте инструкции и решения по настройке конфигураций ресурсов с помощью средств Yandex Cloud.

## Пользовательские возможности для обеспечения безопасности вопросно-ответных систем (RAG) с использованием YandexGPT

Безопасность таких компонентов, как первичная база знаний, ответы и запросы клиентов, а также среда разработки и запуска программного кода приложения, управляется пользователем.

- **Шифрование данных и управление ключами и секретами**

  Yandex Cloud предоставляет встроенные функции шифрования при использовании ряда сервисов. В зоне ответственности клиента — включение шифрования в этих сервисах, а также самостоятельная реализация шифрования в других компонентах обработки критичных данных. Для шифрования данных и управления ключами шифрования предназначен сервис Key Management Service (KMS).

- **Аутентификация и управление доступом**

  В Yandex Cloud управление идентификацией, аутентификацией и контролем доступа выполняется сервисами Yandex Identity and Access Management (IAM) и Yandex Cloud Organization. В инфраструктуре Yandex Cloud взаимодействуют различные категории ресурсов, ролей и пользователей. Контроль доступа к ресурсам выполняется сервисом IAM, который проверяет каждый запрос, чтобы все операции с ресурсами выполнялись только пользователями с необходимыми правами.

- **Изоляция и безопасность сети**

  Чтобы изолировать приложения друг от друга, поместите ресурсы в разные группы безопасности, а если требуется наиболее строгая изоляция — в разные сети. Трафик внутри сети по умолчанию разрешён, а между сетями — нет. Трафик между сетями можно передавать только через виртуальную машину с двумя сетевыми интерфейсами в разных сетях, VPN или сервис Yandex Cloud Interconnect.

- **Безопасность приложений**

  Сервис Smart Web Security позволяет защитить вашу инфраструктуру от DDoS-атак и ботов на прикладном уровне L7 модели OSI. Для защиты ваших веб-приложений от внешних угроз в Smart Web Security также реализован Web Application Firewall.

- **Сбор аудитных логов и событий безопасности**

  Основным инструментом сбора логов уровня Yandex Cloud является сервис Yandex Audit Trails, который позволяет собирать аудитные логи о происходящих с ресурсами Yandex Cloud событиях и загружать эти логи в бакет Yandex Object Storage или лог-группу Cloud Logging для дальнейшего анализа или экспорта. Смотрите инструкцию, как запустить сбор логов, а также формат записей и справочник событий.

## Часто задаваемые вопросы

1. **Другие пользователи увидят наши данные?**

   Модель YandexGPT, используемая при взаимодействии с API, работает в stateless-режиме и не хранит ваши данные.

2. **При создании запросов передаются персональные данные. Как защитить их или соответствовать требованиям 152-ФЗ?**

   Сервис Yandex Foundation Models прошёл независимый аудит безопасности. Это заключение подтверждает, что при работе с компонентами сервиса обеспечено соответствие требованиям Федерального закона от 27 июля 2006 года № 152-ФЗ «О персональных данных» согласно приказу ФСТЭК России от 18 февраля 2013 года № 21. Также рекомендуется провести мероприятия по обфускации передаваемой информации (алгоритм защиты конфиденциальной информации путём замены исходных данных фиктивными).

3. **Зачем логируются запросы пользователей? Можно ли это отключить?**

   Логирование проводится в целях улучшения качества работы сервиса. Для исключения хранения данных из запросов пользователей отключите логирование обращений к YandexGPT с помощью инструкции.

4. **Мы хотим работать с документами, в которых есть коммерческая тайна. Как нам быть?**

   Инфраструктура Yandex Cloud аттестована по требованиям законодательства России по защите коммерческой тайны. Запросить этот аттестат или любые другие внутренние регламенты можно через портал соответствия требованиям ИБ.

5. **Вы дообучаете свою модель на наших данных?**

   По умолчанию данные запросов пользователей используются для мониторинга и улучшения качества работы сервиса. Логирование запросов можно отключить.

## Команда на связи

Если вам необходима дополнительная консультация и у вас появились вопросы, дайте нам знать: cloud-trust@yandex-team.ru

Техническая поддержка
Отдел продаж
yandex.cloud