

<https://cloud.ru/blog/cloud-ru-vypustit-resheniye-dlya-raboty-s-iskusstvennym-intellektom-v-oblake>

Автор статьи (при наличии)

10.04.2025

Cloud.ru выпустит решение полного цикла для работы с искусственным интеллектом в облаке

Провайдер облачных и AI-технологий Cloud.ru анонсировал цифровую среду для разработки приложений и агентов на основе искусственного интеллекта (AI) в облаке — Cloud.ru Evolution AI Factory. Решение с необходимым набором инструментов и сервисов будет доступно пользователям уже летом этого года. Об этом рассказал Евгений Колбин, генеральный директор Cloud.ru на ежегодной конференции GoCloud.

Cloud.ru Evolution AI Factory – это цифровая фабрика готовых сервисов. Она объединит облачные AI- и ML-инструменты для работы с генеративными нейросетями и управления жизненным циклом приложений на базе AI. Универсальная среда предоставит широкий спектр возможностей: от обучения моделей до запуска мультиагентных систем. С ее помощью компаниям любого масштаба будет проще разрабатывать AI-агентов и внедрять их в бизнес.

Решение позволит создавать собственные AI-приложения по готовым шаблонам или работать с кастомизированными AI-моделями. Благодаря удобному набору технологий и встроенных ассистентов воспользоваться сервисами смогут не только профессиональные разработчики, но и пользователи без специальных навыков. Cloud.ru Evolution AI Factory поможет значительно минимизировать затраты на вычислительную инфраструктуру, повысить эффективность работы команд и сократить время вывода новых продуктов.

Cloud.ru Evolution AI Factory будет включать полностью готовую инфраструктуру, сервисы и продукты для простого и быстрого создания AI-приложений:

- **Foundation Models** – быстрый и удобный доступ в российском облаке к таким популярным моделям open source AI, как GigaChat, DeepSeek, Qwen и Gemma. Сервис создан для разработчиков, исследователей и компаний, которым нужно адаптировать AI-модели под свои задачи, проверять гипотезы и проводить эксперименты — без необходимости разворачивать и оптимизировать сложную инфраструктуру. В составе сервиса также доступна AI Playground – удобная среда для быстрого прототипирования, проверки гипотез и тестирования моделей на практике;
- **Evolution ML Inference** – сервис для запуска ML-моделей из библиотеки Hugging Face на облачных мощностях с GPU, в том числе в пользовательских Docker-образах. Он поддерживает среды исполнения vLLM, TGI, Ollama, Diffusers, Transformers, а также модели-трансформеры, диффузионные модели и большие языковые модели (LLM). Инференс как сервис позволяет разделять GPU-ресурсы между задачами на уровне production-ready решения. Благодаря этому пользователи смогут гибко и максимально эффективно управлять ресурсами в облаке, снижая затраты и ускоряя работу AI-моделей;

- **AI Assistants Studio** – для создания агентных и мультиагентных систем с подключением к любым внешним системам с MCP (Model Context Protocol). Он поможет настроить взаимодействие между ИИ-агентами и источниками данных с возможностью использования продвинутых моделей и собственных решений. Сервис обеспечивает быстрый доступ к информации, автоматизацию рутинных задач и поддержку принятия решений с такими инструментами, как RAG (retrieval augmented generation);
- **Cloud.ru ML Space** – платформа для задач глубокого обучения приложений с использованием суперкомпьютеров и ML на базе российского решения. Подходит для полного цикла ML-разработки в облаке и on-premise.

Cloud.ru Evolution AI Factory — это среда с универсальными инструментами для сценариев, где идеи превращаются в умных агентов быстро, просто и без лишних затрат для бизнеса. С запуском Cloud.ru AI Factory мы даём пользователям готовые сервисы для полного цикла работы с AI: от обучения и инференса до создания собственных агентов и ассистентов — без необходимости глубоких технических знаний. Платформа позволяет быстро проверять гипотезы, создавать ПОС и MVP, используя все необходимые инструменты в облаке Cloud.ru Evolution, ориентированном на скорость, гибкость и доступность GenAI-инструментов. Поэтому инфраструктура и технологии собраны в одном месте – это упрощает доступ к ним и ускоряет запуск AI-приложений. Представители AI-сообщества могут предложить свои идеи и гипотезы для быстрой проверки на нашей платформе и получить доступ к запуску моделей на новой технологической базе. Наша цель – объединить в Cloud.ru не только инструменты для создания AI-приложений, но и независимых разработчиков ПО, чтобы упростить работу с облаком и бизнесу в целом.

Евгений Колбин генеральный директор провайдера облачных и AI-технологий Cloud.ru