Георгий (Гоша) Мареев

AI-инженер

RAG · AI-агенты · Корпоративные LLM-платформы

Москва (GMT+3) · hello@mareevapps.com · GitHub · mareevapps.com

ПРОФИЛЬ

AI-инженер с коммерческим опытом в AI/ML с июня 2025 года в команде корпоративной платформы. Разворачиваю и сопровождаю LLM-системы в защищённом контуре, разрабатываю RAG-пайплайны и внутренних ассистентов. Работаю с контролем доступа, оценкой качества, наблюдаемостью и аудитом запросов. Английский B2.

ОПЫТ РАБОТЫ

Международная FMCG-компания | AI Engineer (AI/ML Solutions Engineer)

Июнь 2025 - настоящее время · Москва

- Развернул и сопровождаю корпоративную AI-платформу: единый шлюз LiteLLM, Open WebUI, локальные и OpenAI-совместимые модели, изоляция сервисов и SSO через Microsoft Entra ID. Платформой пользуются 120 пользователей в 10 командах.
- Проектирую и развиваю двух RAG-ассистентов на общем ядре: загрузка документов, разбиение, эмбеддинги, поиск с учётом метаданных, ранжирование, ответы со ссылками и обоснованные отказы. Синхронизация SharePoint через Microsoft Graph.
- Настроил доставку эксплуатационных логов через Fluent Bit в Loki и отдельный транспорт аудита запросов в SIEM, маскирование учётных данных и метрики запросов и задержек в Prometheus.
- Реализовал настраиваемую JSON-систему сценариев: приоритеты, диагностика и правила меняются без изменения кода и выпуска приложения.
- Поддерживаю разграничение доступа, изоляцию развёртывания и регулярную оценку качества ответов. Политика guardrails и правила детектирования SIEM находятся в разработке.

ИЗБРАННЫЕ ПРОЕКТЫ

Secure LLM Platform Reference | 2026

Публичная запускаемая реализация на синтетических данных: Open WebUI, LiteLLM, pgvector, RAG с разграничением доступа, проверки приватности EN/RU и устойчивый аудит. Базовые RAG-проверки: 32/32 в каждом из трёх прогонов; расширенные: 125/128. Методика и ограничения.

Women’s Clothing Reviews | NLP-анализ | 2025

Анализ тональности и тематическое моделирование: TextBlob, BERTopic с ускорением на GPU, интерактивный дашборд Streamlit и PDF-отчёты.

Data Science Salaries | ETL и аналитика | 2025

Загрузка данных Kaggle, оркестрация Airflow, витрина PostgreSQL, визуализации Superset; Docker Compose, Python и SQL.

Личные лаборатории | mareevapps.com

Neural Bonsai: обучение и исследование MLP в браузере на TypeScript. ApsisDB: встраиваемая база данных на Go для циклической телеметрии. Генеративная типографика и лаборатория шейдеров.

НАВЫКИ

Стек: Python · FastAPI · Docker · Postgres/pgvector · LiteLLM · Open WebUI · Loki · nginx · Linux

LLM и RAG: разбиение документов, эмбеддинги, поиск, reranking, ответы с источниками, оценка качества, ChromaDB, FAISS, Ollama, OpenAI-совместимые API; проверки приватности и инъекций в публичной реализации.

Backend и эксплуатация: Flask, Pydantic, REST API, Bash, Git, Prometheus, Fluent Bit, SSO / Entra ID, транспорт аудита в SIEM.

ML и данные: scikit-learn, PyTorch, BERTopic, pandas, NumPy, SQL, Airflow, Streamlit, Superset.

ОБРАЗОВАНИЕ И ДОПОЛНИТЕЛЬНО

МГПУ: магистратура «Бизнес-аналитика и большие данные» (обучение продолжается); бакалавриат, педагогическое образование, профиль «Информатика».

По совместительству: учитель информатики в школе № 1533 «ЛИТ», Москва, с сентября 2023 года; программирование, алгоритмы и подготовка к олимпиадам.

Достижение: 1-е место, олимпиада «Технологии успеха», направление «Фотоника», руководитель команды.

Языки: русский - родной; английский - B2.