



AI Repetitor in Tech & AI

Core Expertise

Tech Programming

QA

Product Management

ML

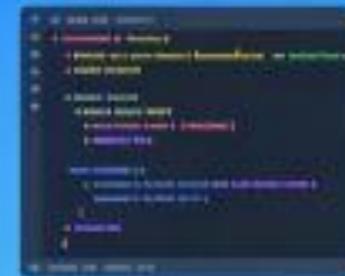
DL

GENAI

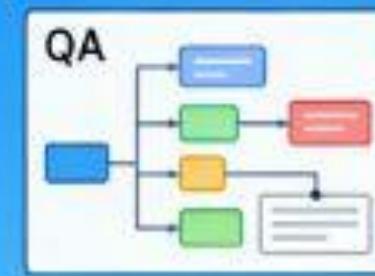
LLM

Diffusion Models

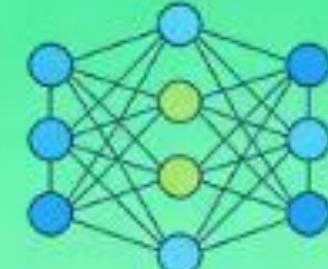
Tech Programming



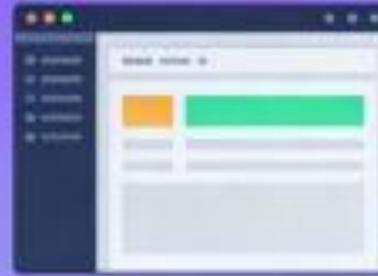
Test Case



Product Management



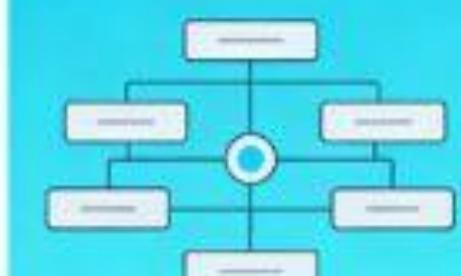
ML & DL



Generate AI



Diffusion
Models





Хакатон МИФИ и Cloud.ru, команда Lab_Story



Перова
Анна

Tech Lead/
Team Leader

Data Engineer
ML Researcher
Full-Stack

ML researcher
Data Engineer
Backend

ML researcher
Data Engineer
MLOps
Backend

ML researcher
Data Engineer
Backend

ML researcher
Data Engineer
Backend

ML researcher
Data Engineer
QA



1. Цель проекта

Целью проекта является разработка прототипа AI-репетитора для обучения IT-дисциплинам (Python, Machine Learning) на основе подхода Retrieval-Augmented Generation (RAG). Решение предназначено для помощи студентам в освоении учебных материалов с опорой на проверенные источники.

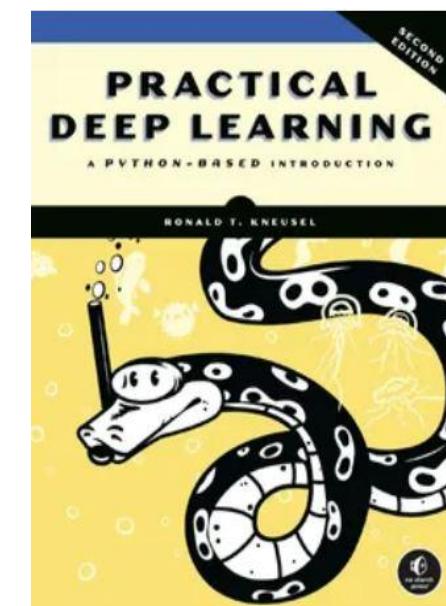
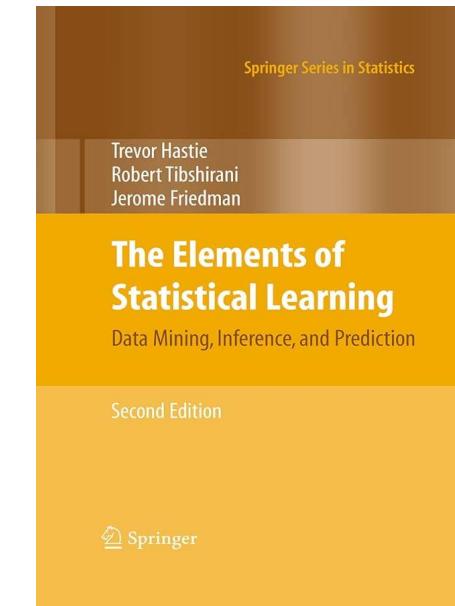
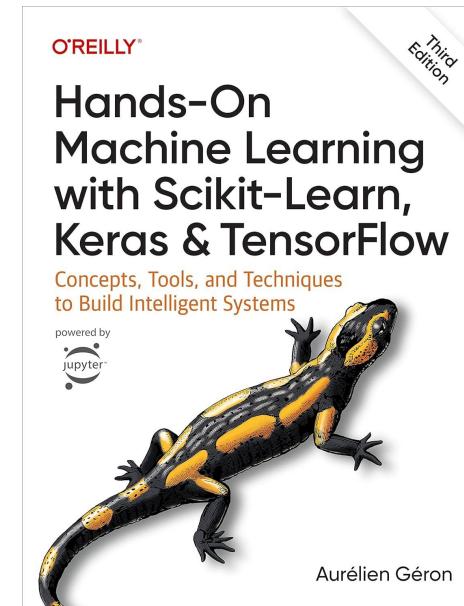
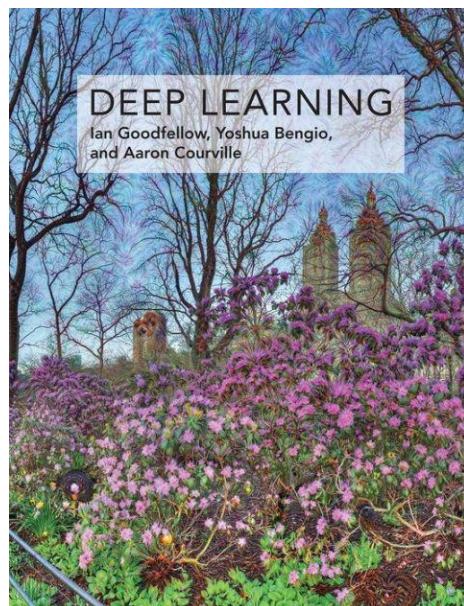
Разработать **AI-репетитора**, который:

- помогает изучать технические дисциплины (ML, DL, CV, NLP, Diffusion Models, LLM)
- объясняет темы **по учебной логике**,
- опирается **только на проверенные источники**
- снижает риск галлюцинаций за счёт RAG-подхода,
- адаптирует объяснение под уровень пользователя.



2. Сбор и подготовка корпуса знаний

- В рамках проекта был сформирован корпус учебных материалов,
- Включающий открытые образовательные источники и учебные тексты по тематике Python и Machine Learning.
- Материалы приведены к текстовому формату и подготовлены для дальнейшей обработки.





3. Архитектура решения

Решение реализовано по классической архитектуре RAG и включает следующие этапы:

1. Разбиение текстов на смысловые фрагменты (чанкинг).
2. Построение эмбеддингов для текстовых фрагментов.
3. Индексация эмбеддингов в векторном индексе.
4. Поиск релевантных фрагментов по запросу пользователя.
5. Формирование промпта на основе найденного контекста.
6. Генерация ответа языковой моделью.

Логика RAG-пайплайна

Запрос пользователя → векторизация → поиск контекста → формирование промпта → генерация ответа

Поток обработки запроса выглядит следующим образом:

1. Пользователь вводит вопрос или тему через простой UI.
2. Запрос преобразуется в векторное представление.
3. Выполняется semantic search по корпусу учебных материалов.
4. Отбираются релевантные фрагменты текста (retrieved chunks).
5. На основе найденного контекста формируется prompt.
6. Языковая модель генерирует ответ **строго на основе этого контекста**.
7. Пользователь получает структурированное объяснение темы и контрольный вопрос.



5. Точность и ссылки

Ответы AI-репетитора формируются на основе загруженного корпуса знаний. Использование **Retrieval-подхода** обеспечивает прозрачность и снижает риск генерации недостоверных ответов.

При необходимости возможно расширение механизма явного отображения источников.

В основе решения лежит Retrieval-подход, который:

- обеспечивает **прозрачность происхождения информации**,
- снижает риск галлюцинаций и недостоверных утверждений,
- позволяет контролировать, **на какие источники опирается модель**.

Роль Retrieval в повышении доверия

В процессе генерации ответа:

- выполняется semantic search по корпусу знаний,
- отбираются только релевантные фрагменты текста,
- языковая модель получает **жёстко ограниченный контекст**,
- ответ формируется **исключительно на основе найденных фрагментов**.

Это принципиально отличает решение от обычных LLM-чатов.



6. Пользовательский интерфейс

Для демонстрации работы решения предусмотрен простой пользовательский интерфейс, позволяющий студенту вводить вопросы и получать текстовые ответы. Интерфейс ориентирован на минимализм и удобство взаимодействия.

RAG-репетитор (Gradio, прямо в Colab)

Ни каких Telegram и туннелей. Работает сразу.

Направление

Machine Learning Computer Vision NLP

Уровень

Beginner Middle Senior

Тема / вопрос

Overfitting

Найти материалы

Overfitting (Machine Learning, Middle)

Найдено в источниках:

- Goodfellow ML the training data differs from the challenge of finding patterns that generalize to new data. Most machine learning algorithms have settings called hyperparameters, which must be determined outside the learning algorithm itself; we discuss how to set these using additional data. Machine le... ↗ <https://www.deeplearningbook.org/contents/ml.html>

Как установить библиотеки Python? Загрузи данные с Google Диска Покажи поименованные переменные

Спасибо за внимание!

