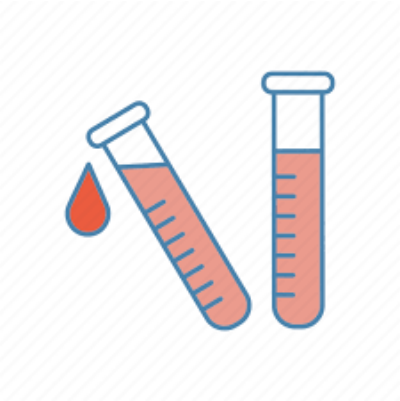
LLMsLab Infrastructure

An experimentation playground for exploring and harnessing the power of large language models



LLMsLab: Our Tools and Methodologies

LLMsLab: Our Tools and Methodologies

Version Control: We leverage **GitHub**, a robust platform for collaborative coding and version control.

IDEs: Our toolkit includes **Visual Studio Code** for a powerful local development environment, **GitHub Codespaces** for seamless online development, and **Google Colab** for convenient, web-based Jupyter notebooks.

Build & Automation: GitHub Actions serves as our core CI/CD tool, automating testing, building, and deployment tasks.

AI Model Interfaces: We use **Gradio** and **Hugging Face Spaces** to create interactive, user-friendly interfaces for our AI models.

LLMsLab: Our Tools and Methodologies

Database Systems: When necessary, we incorporate **Vector Databases** to manage our data effectively.

Project Management: Jira Software helps us track tasks, bugs, and progress, ensuring our projects are well-organized and on track.

Communication: We use **Slack** for instant communication, fostering a dynamic and collaborative work environment.

Project Management Philosophy: We adopt an **Agile** approach with **Kanban**, promoting continuous delivery and flexible response to change.

LLMsLab Repositories

LLMsLab Repositories

- **chat-gpt-api-lab**: A sandbox for testing, playing, and experimenting with the various components and capabilities of the ChatGPT API
- **qa-app-lab**: A sandbox for developing question answering applications using real insurance chat data
- **model-lab**: An experimentation playground for exploring and harnessing the power of open-source large language models
- **langchain-lab Public**: A sandbox for testing, playing, and experimenting with the various components and capabilities of the LangChain framework

MLOps with Hugging Face Spaces, Gradio & GitHub Actions

Hugging Face

gradio

GitHub

How to a deploy Hugging Face Spaces App using Continous Delivery - a true MLOps Workflow

MLOps and CI/CD

- MLOps is the application of DevOps practices to ML and AI, enabling faster experimentation and reliable deployment of models.
- CI/CD automates app development stages, crucial for fast, reliable delivery in AI/ML projects.

Key Tools

- **Hugging Face Spaces:** A platform to host and share AI models as interactive web spaces.
- **GitHub Actions:** Automates tasks like testing, building, and deploying applications in a CI/CD pipeline.
- **Gradio:** An open-source library for creating user-friendly interfaces for ML models, simplifying testing and demonstration.

MLOps Architecture

In this slide, we will explore the architecture of our MLOps approach. This diagram illustrates how different components interact with each other in our CI/CD pipeline, incorporating Hugging Face Spaces, GitHub Actions, and Gradio.