Open in app ↗

Search

Sarvadnya

An open-source repository for All-Knowing Custom Chatbot

Yogesh Haribhau Kulkarni (PhD)

Published in **Technology Hits**

3 min read · 20 hours ago

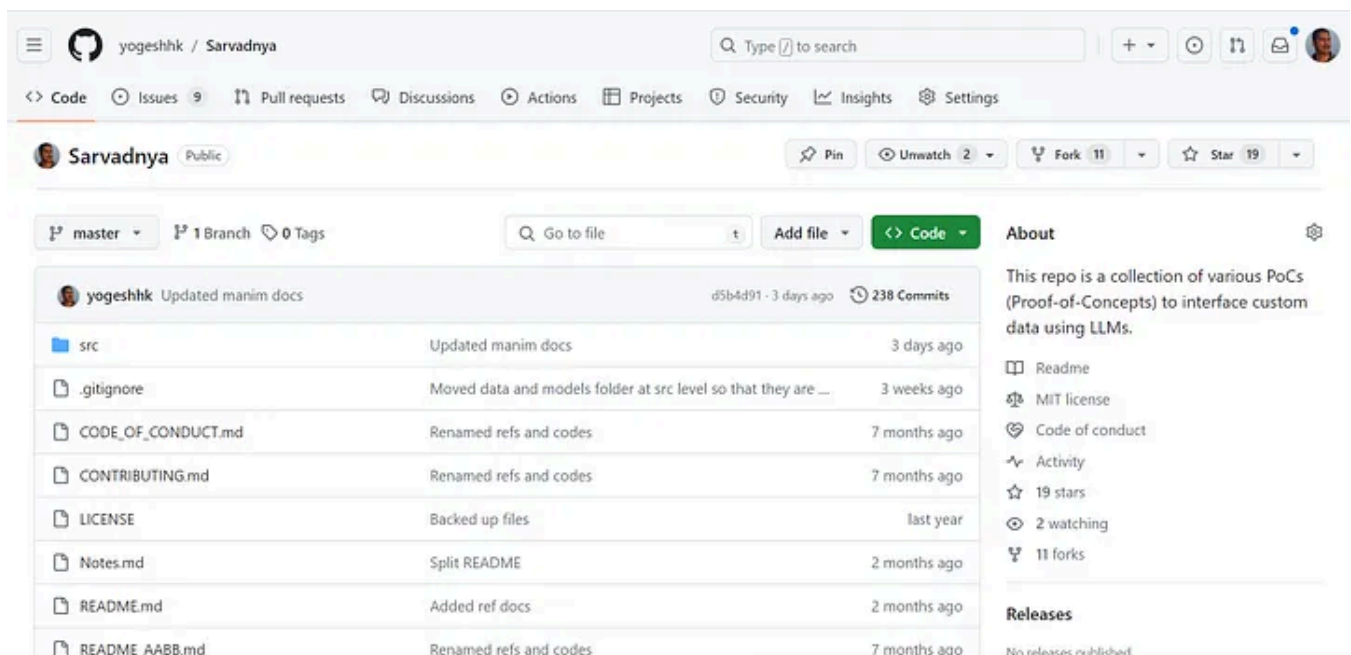▶ Listen   ⬆ Share   ••• More

Screenshot of the repository

Chatbots can be real WoW!! The recent evidence is: ChatGPT. Now that they are more human-like with the latest LLMs (Large Language Models). But these LLMs are Pretrained on their own (HUGE) data. Mere mortals don't have any ways ($$, time, expertise) to train own LLMs. RAG and/or Fine-tuning is the way out for Domain Adaptation ie. LLMs answering on your corpus.

This repo is a collection of various PoCs (Proof-of-Concepts) to interface custom data using LLMs. Some notes here.

What?

PoCs Projects

Chatbots of various modalities, use cases, and domains, diff datasets. Videos, write Medium Posts (GDE/TH), LinkedIn posts, Youtube channel.

Modes

- Retrieval Augmented Generation (RAG) on own data

- Fine-tuning LLMs with own data using LoRA etc

RAG

Why ?: World needs, ready to pay, I am good at, and I like it making wow chatbots

Domain:

- On knowledge graphs, more grounding

- Tabular financial data, representation and similarity

- MidcurveNN Geometric serialization and retrieval

- Active loop idea of fine-tuning your data

- Langchain and Llamaindex with any new LLM

Fine-Tuning

When?: Sufficient curated data is available, not a whole lot though, in a batch (not running) state

What?: Instead of unstructured text (input prompts) to unstructured text (output response), more value is in prompt to structured output, such as:

- text2json: many enterprises such as financial companies.

- text2cypher: for graph databases, from Neo4j, like Langchain implementation by Tomaz Bratanic

- text2SQL: classical case, many pro solutions available, study them, follow them, for other QLs

- text2Manim: Maths Animation, dataset available, see if generated video can be shown in the same Streamlit page

- text23DJS: Good for 3D+LLM+Agents like Metamorph from Nvidia, Geometry or shape representation as text, is the key

- textGraph2textGraph: MidcurveNN if we get Graph representation as text, right.

Here, the key would be robust post-processing and evaluation as the response needs to be near perfect, no scope of relaxation even in syntax or format.

Tech Stacks

- Enterprise: Google Doc AI, Vertex AI, Microsoft Azure Language AI Services

- Open Source: Langchain (Serve/Smith/Graph), HuggingFace, Streamlit for UI

Product Specs

- {KG + LLM} chatbot Building LLM based Chatbot on Knowledge Graph

- Knowledge as a Service (KaaS) Building Knowledge Graph from Text and serving it as a Service

- LLM models for Indic (especially Sanskrit) languages. Here is a collection of similar efforts going on Awesome AI By Bharat (AABB)

Disclaimer

- The author (firstnamelastname at yahoo dot com) provides no guarantee of the correctness of the content.

- Notes have been built using lots of publicly available material.

- Although care has been taken to cite the original sources, there could be some missing ones. Point them out, and I will update wherever possible.

- Many improvements are still to be made. Don't fully depend on it.

- Send in your suggestions, comments, corrections, and pull requests.

Click pic below or visit LinkedIn to know more about the author

Artificial Intelligence   Chatbots   Large Language Models   Future

Generative Ai Tools

Written by Yogesh Haribhau Kulkarni (PhD)

1.3K Followers   ·   Editor for Technology Hits

PhD in Geometric Modeling | Google Developer Expert (Machine Learning) | Top Writer 3x (Medium) | More at https://www.linkedin.com/in/yogeshkulkarni/

More from Yogesh Haribhau Kulkarni (PhD) and Technology Hits

Yogesh Haribhau Kulkarni (PhD) in Technology Hits

Unveiling Manifold learning

What a neural network is really doing?

5 min read · Feb 23, 2024

275   1

Hammad Hassan in Technology Hits

ChatGPT Vs Gemini: Which One I Like As A Writer

As a writer, I know the struggle of digging through mountains of information.

✦ · 4 min read · Feb 8, 2024

👏 1.8K   💬 44   🔖   ⋯

Nuno Campos in Technology Hits

Obsidian Plugins Review — 44

Elevate Your Obsidian Experience: Discover the Best New Plugins for Enhanced Note-Taking and Boost Productivity

✦ · 5 min read · Feb 10, 2024

👏 434   💬 4   🔖   ⋯

Yogesh Haribhau Kulkarni (PhD) in Google Developer Experts

A CAMEL ride

A Story of AI Role-Playing using CAMEL, Langchain and VertexAI
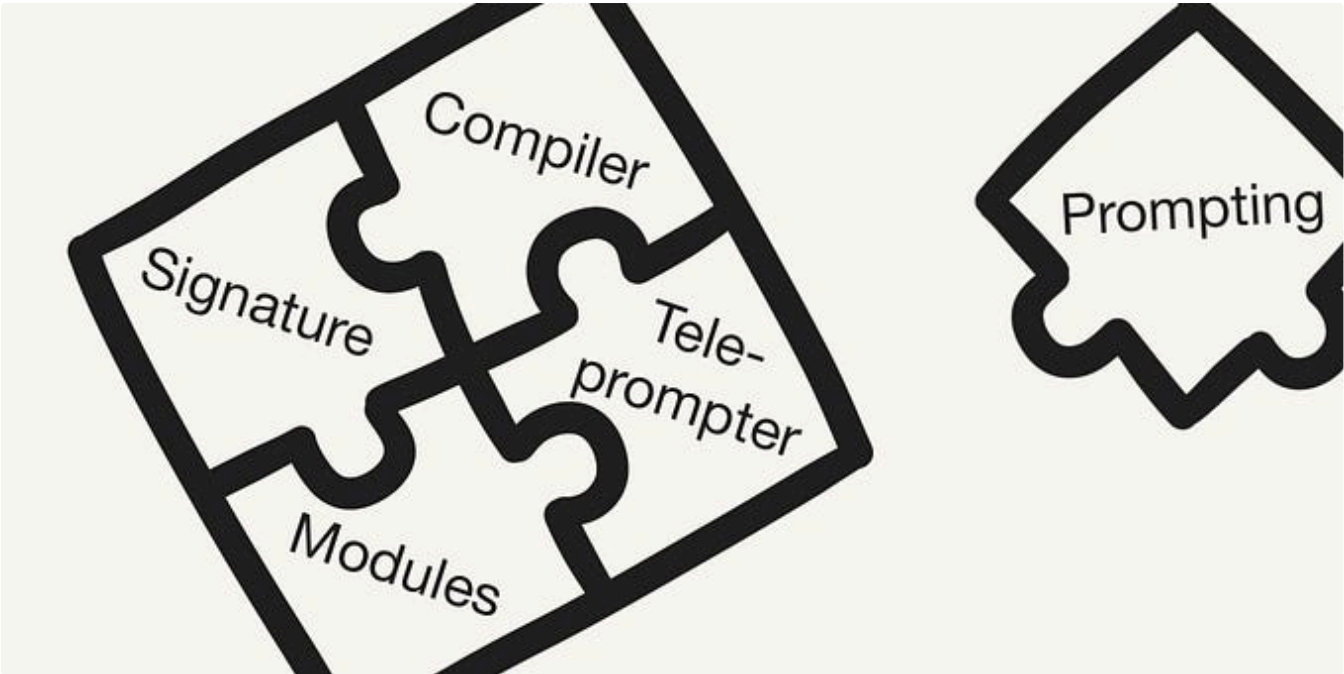
8 min read · Oct 2, 2023

See all from Yogesh Haribhau Kulkarni (PhD)

See all from Technology Hits

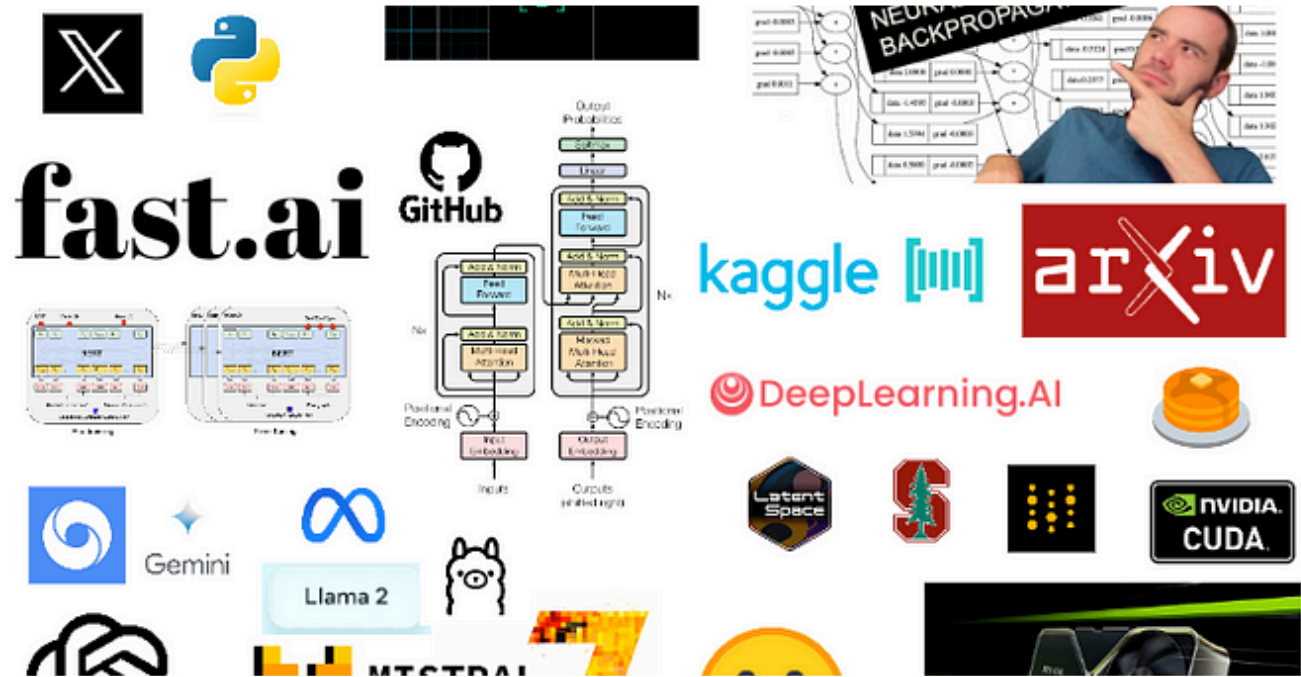## Recommended from Medium

Leonie Monigatti in Towards Data Science

Intro to DSPy: Goodbye Prompting, Hello Programming!

How the DSPy framework solves the fragility problem in LLM-based applications by replacing prompting with programming and compiling

✦ · 13 min read · 6 days ago

👏 1.8K   💬 8   🔖⁺   •••

Benedict Neo in bitgrit Data Science Publication

Roadmap to Learn AI in 2024

A free curriculum for hackers and programmers to learn AI

11 min read · Feb 21, 2024

👏 3.9K   💬 45   🔖⁺   •••

Lists

AI Regulation
6 stories · 349 saves

Natural Language Processing
1253 stories · 733 saves

Generative AI Recommended Reading
52 stories · 783 saves

ChatGPT
21 stories · 499 saves
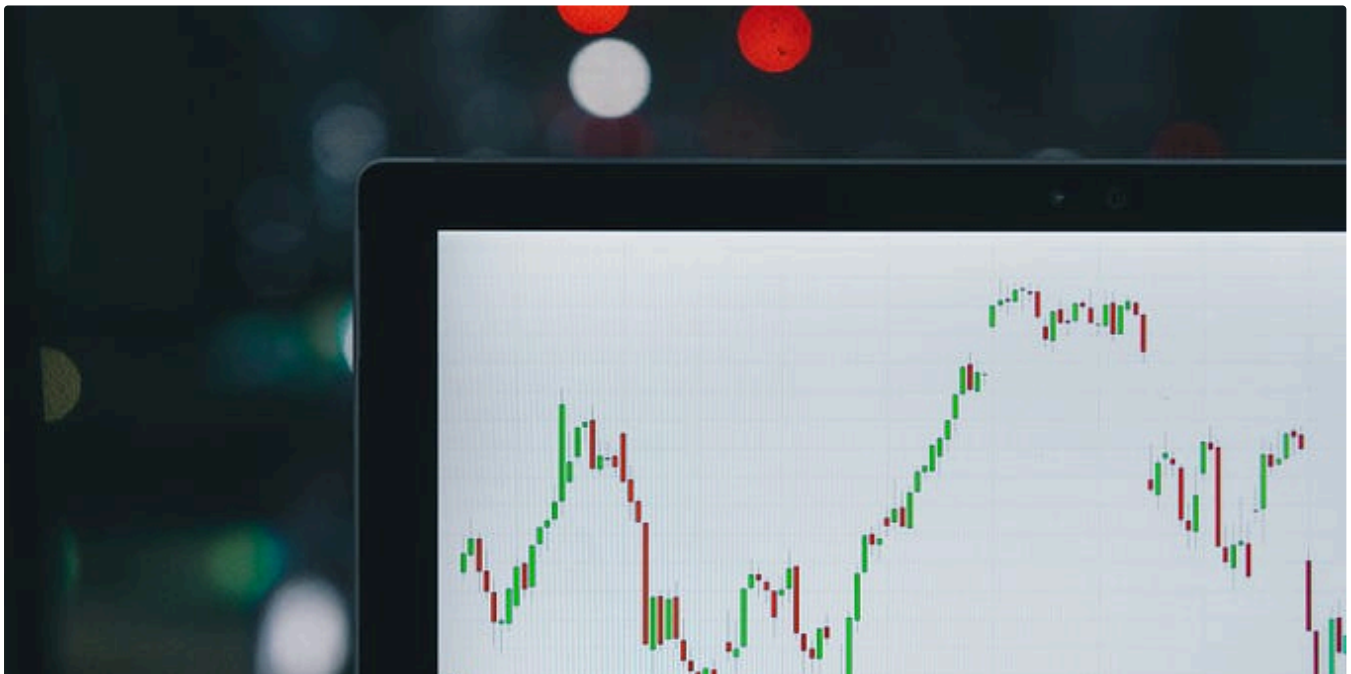
👤 Yanli Liu in Level Up Coding

The Future of Financial AI: Stock Sentiment Analysis with SLIM Models

How to perform complex multi-step analytics using LLMWare on CPU

✨ · 8 min read · 6 days ago
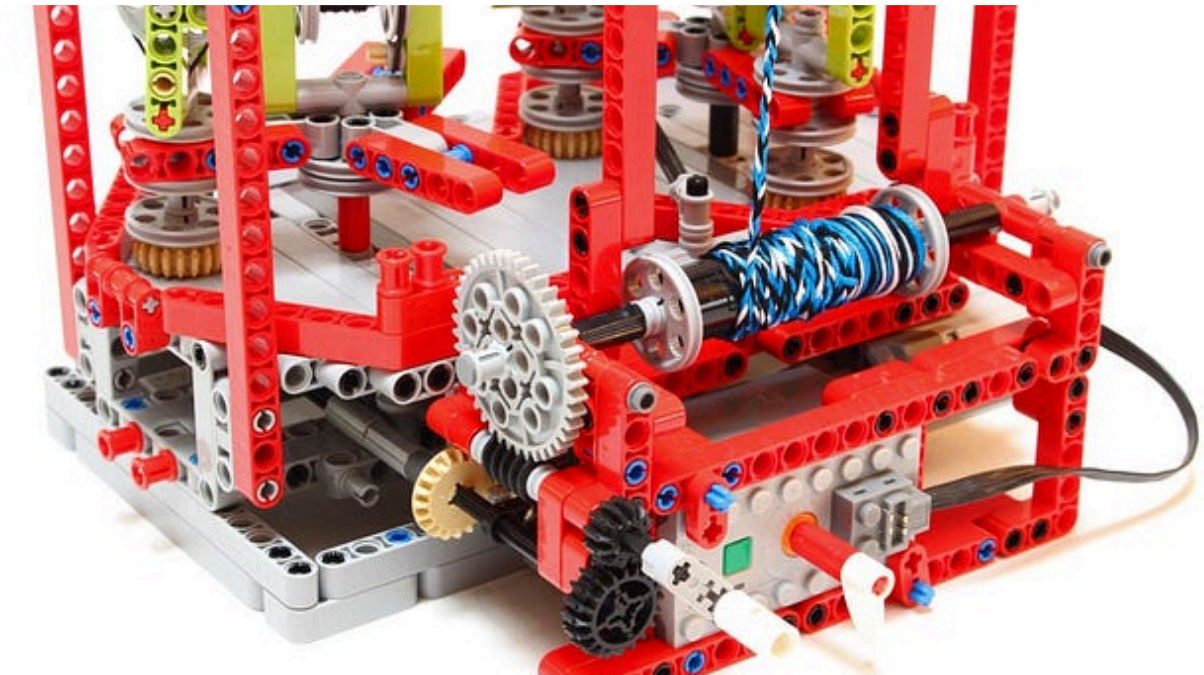
Fabio Matricardi

Smarter, not Bigger — 1 Million token context is not all you need!

Open source at the test bench to verify long context information extraction. Are they really that good?

✨ · 10 min read · 5 days ago

Build & Run
LangChain Applications
Locally & Offline

LangChain   Titan^ML
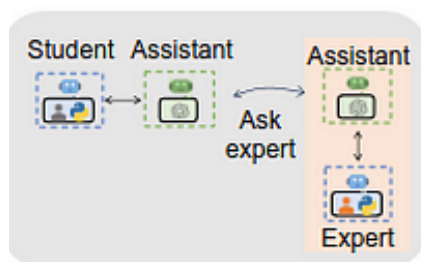
Jupyter   TinyLlama

👤 Cobus Greyling

How To Create A LangChain Application That Runs Locally & Offline

In tightly regulated sectors such as Telecommunications or Banking, the utilisation of a commercially available API from an LLM provider…

9 min read · 5 days ago
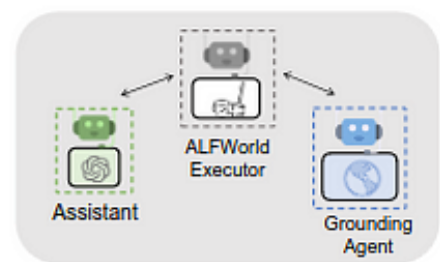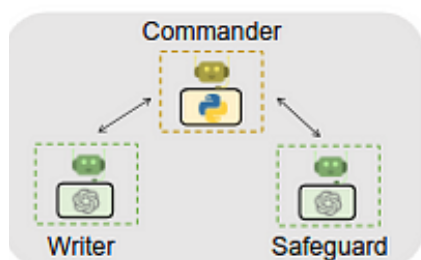
👏 97   💬
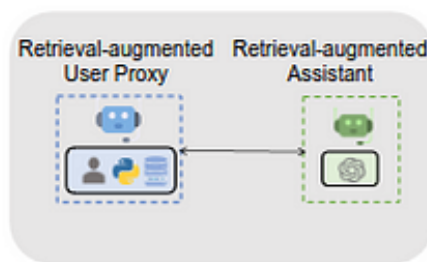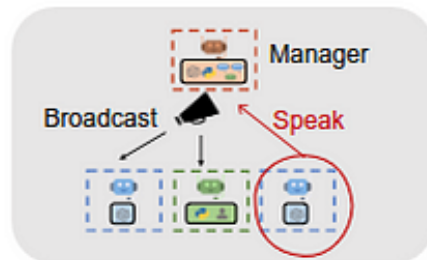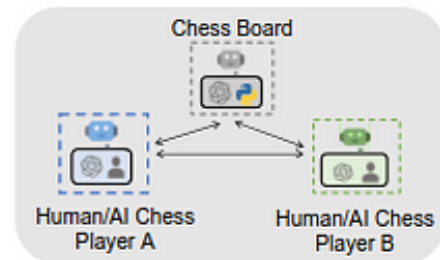
👤 Rubens Zimbres

Cost-Efficient Multi-Agent Collaboration with LangGraph + Gemma for Code Generation

In the end of January, 2024, I received an email about LangGraph and LangSmith. Some features were not in GA, so I signed up for the…

11 min read · Feb 24, 2024

See more recommendations