Analytics Vidhya | DATAHACK SUMMIT

Workshop

NLP using Generative Models : What's Next?

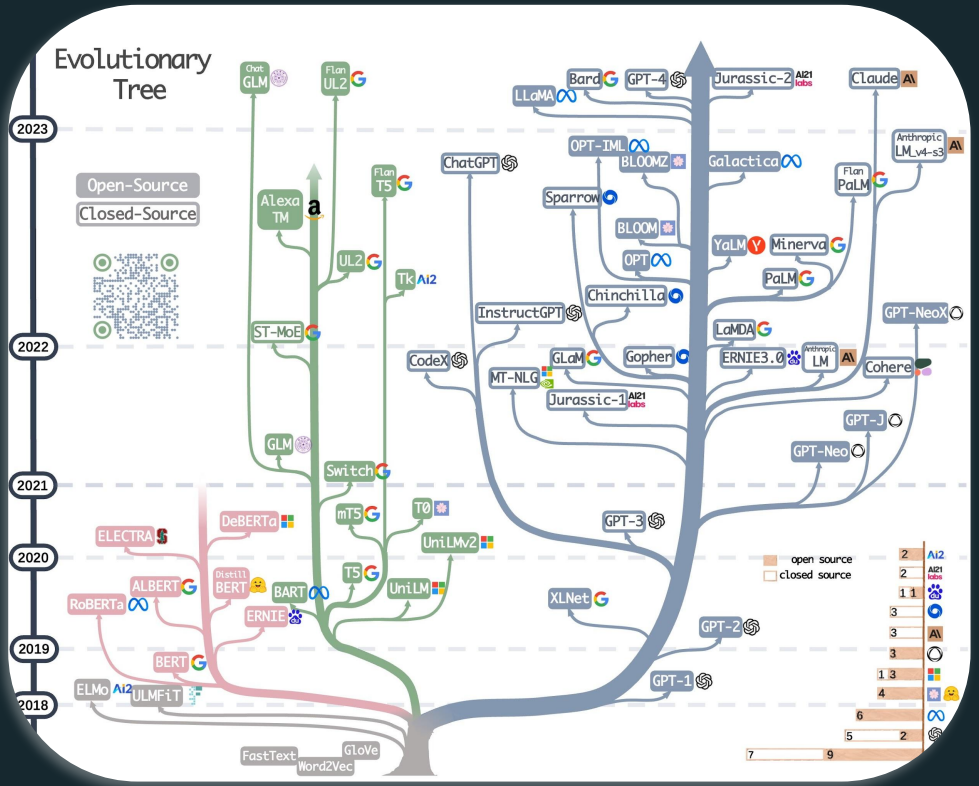
Speaker

Amar Lalwani

Building a startup (stealth mode)
Ex - Head of AI and R&D

Current State : Model Universe

- Huge improvements from the days of *word2vec* to current *LLMs*
- Extremely expensive to train and research (few key players)
- Far more accessible today than a few months ago
- Breakneck speed of innovation



Source: https://github.com/Mooler0410/LLMsPracticalGuide

Exercise

- What is 2 + 2 ?
- Which object is smaller ?

- Complete the phrase “This Saturday I am ….”

Language Models

Pretrained Language Models

- Huge Datasets
- Very Large Compute
- Predict Next token

GPT, LLama, PaLM

Supervised Finetuned Models

- Small Dataset
- Medium Compute
- Predict Next token

Vicuna, MPT, ChatGLM

RLHF Models/Agents

- Smaller Dataset
- Medium Compute
- Predict Next token & Maximize Reward

chatGPT, Bard, Claude

- Extremely good at predicting the next token
- Prompts help structure the entropy for specific tasks

System 1 Characteristics

Error Prone

Fast

Automatic

Source: https://lmsys.org/blog/2023-05-25-leaderboard/

Systems 2 Thinking

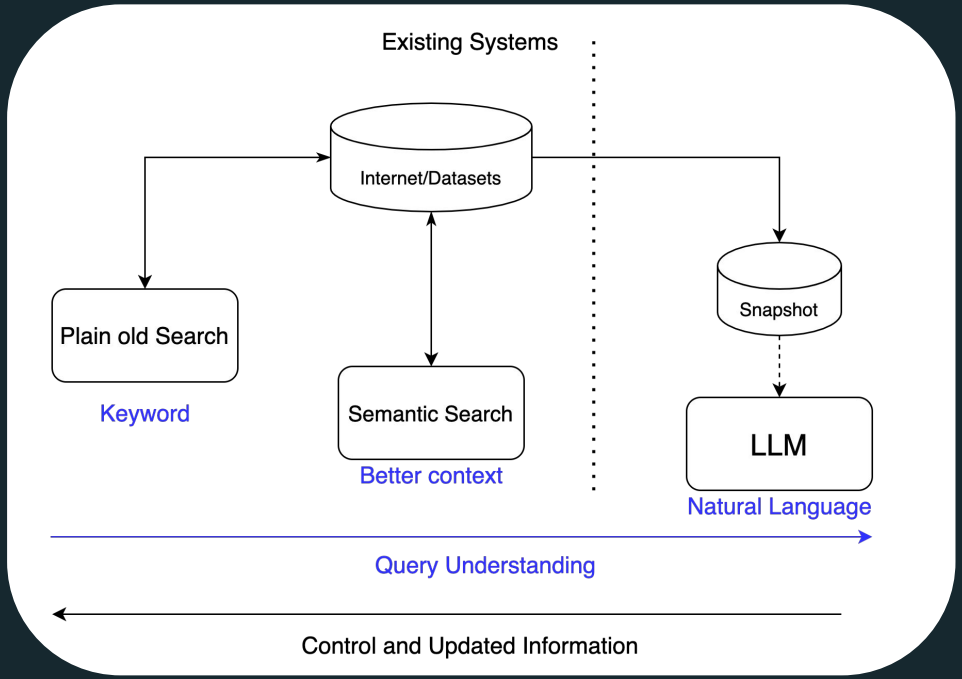
- Write a blog post about LLM Landscape
  - Research and list key areas in the LLM landscape
  - Create an outline with key areas as headings
  - Research and read papers and write understanding of each topic
  - Rephrase paragraphs to make them understandable
  - Re-read entire draft and rephrase to make it cohesive
  - Proofread for any errors
  - Add references
  - Add illustrations and images
  - Post

System 2 Characteristics

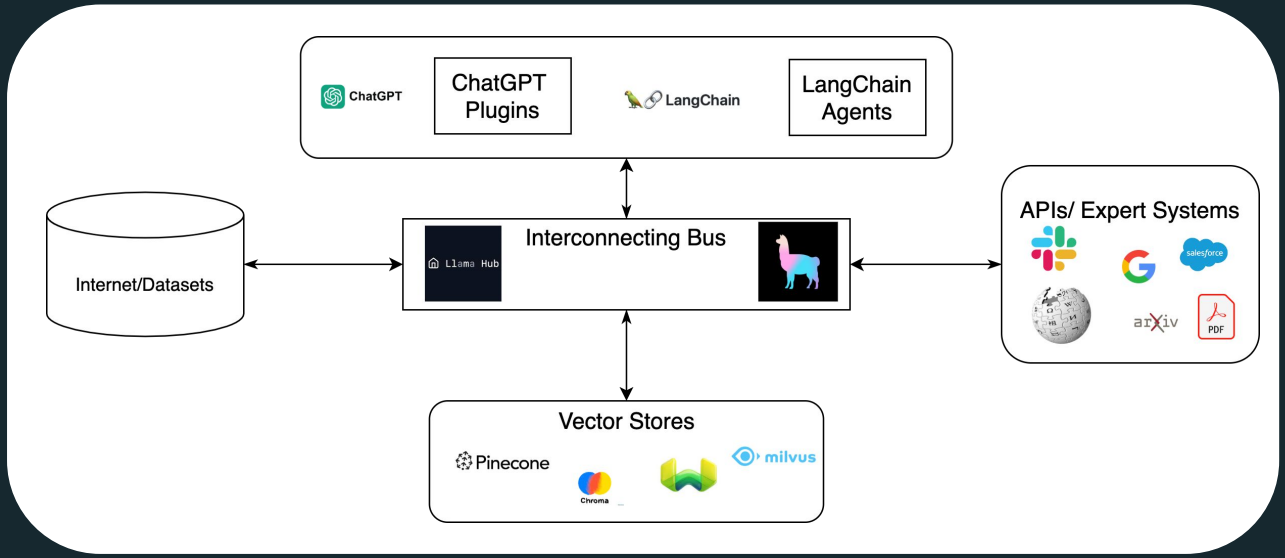
Conscious Effortful Slow

Agents of Change

- Existing systems leverage keyword based or semantic search
- Current LLMs are far better at understanding natural language but not good at expert tasks and recent information



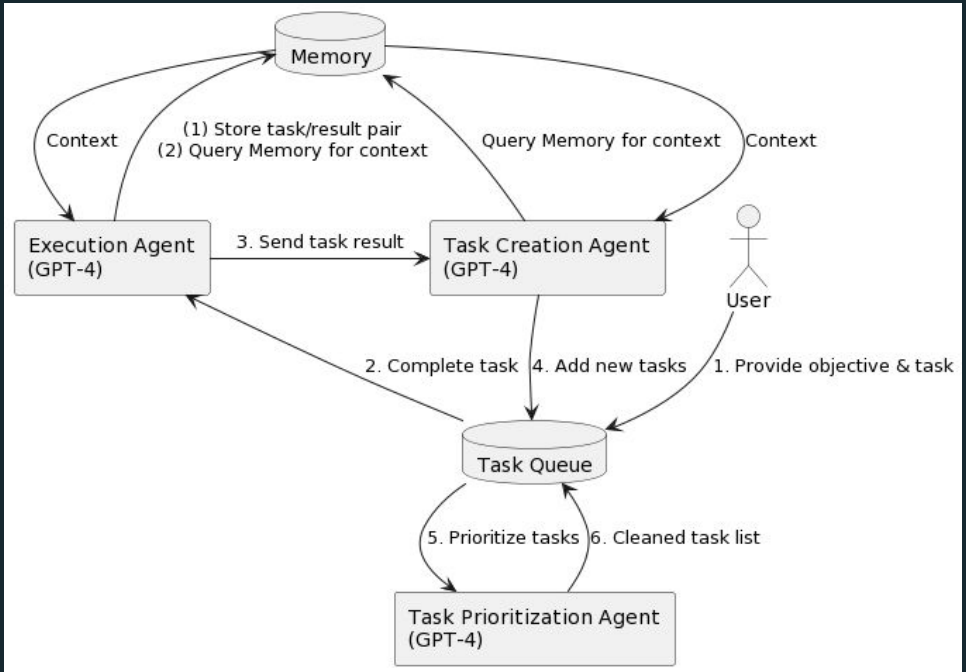
Agents of Change



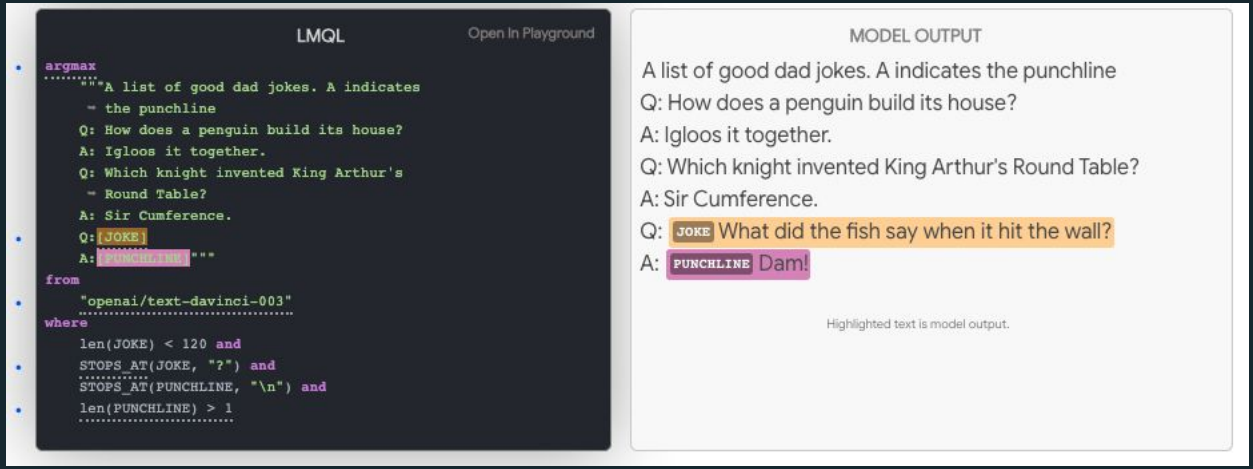
A *Retrieval Augmented System* leveraging LLMs for understanding and expert systems for specific tasks

AutoGPT

- Agent driven system
- Ability to leverage external systems/APIs
- Ability to reference additional memory
- Ability to self-prompt/meta-prompt



Source :https://github.com/Significant-Gravitas/Nexus/wiki
Source: https://www.lesswrong.com/posts/566kBoPi76t8KAkoD/on-autogpt

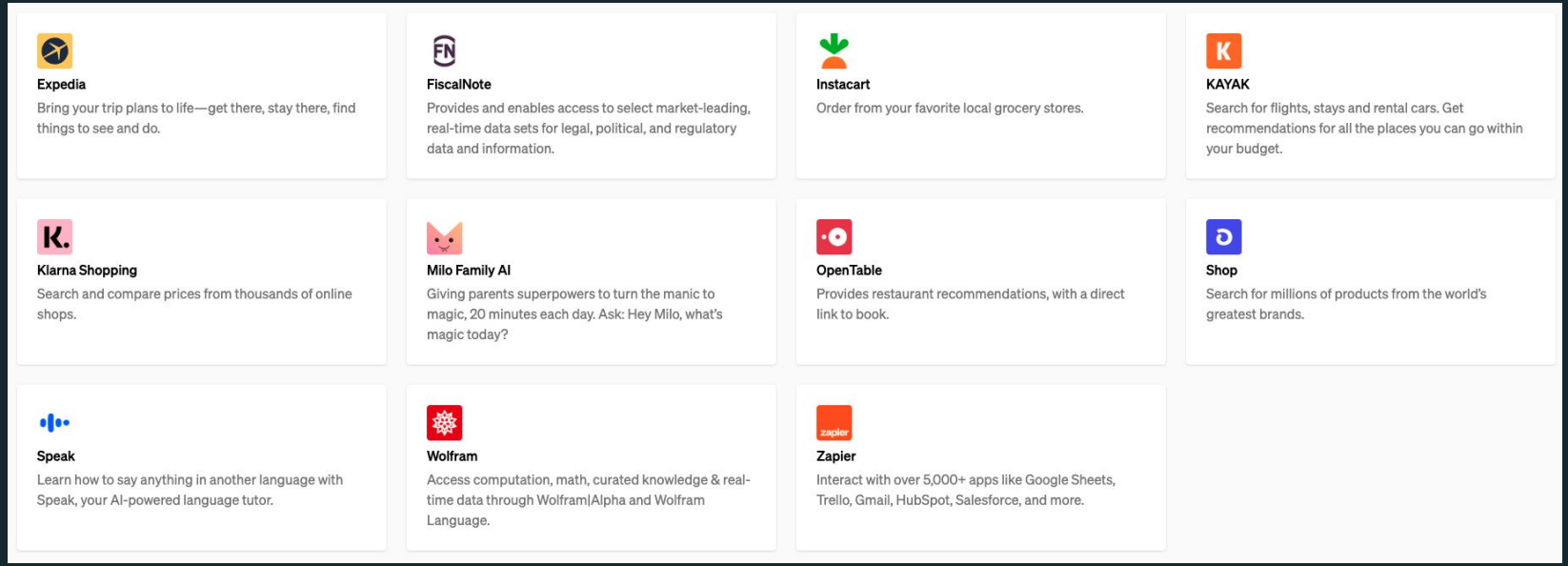


- Language Model Query Language is a simplified interface for interacting with LLMs
- Expressiveness of Python combined with ease of Natural Language *(COBOL you there?)*
- Enables multi-part prompting, tool augmentation, scripting, constraint guided decoding, etc.

Source: https://github.com/microsoft/guidance

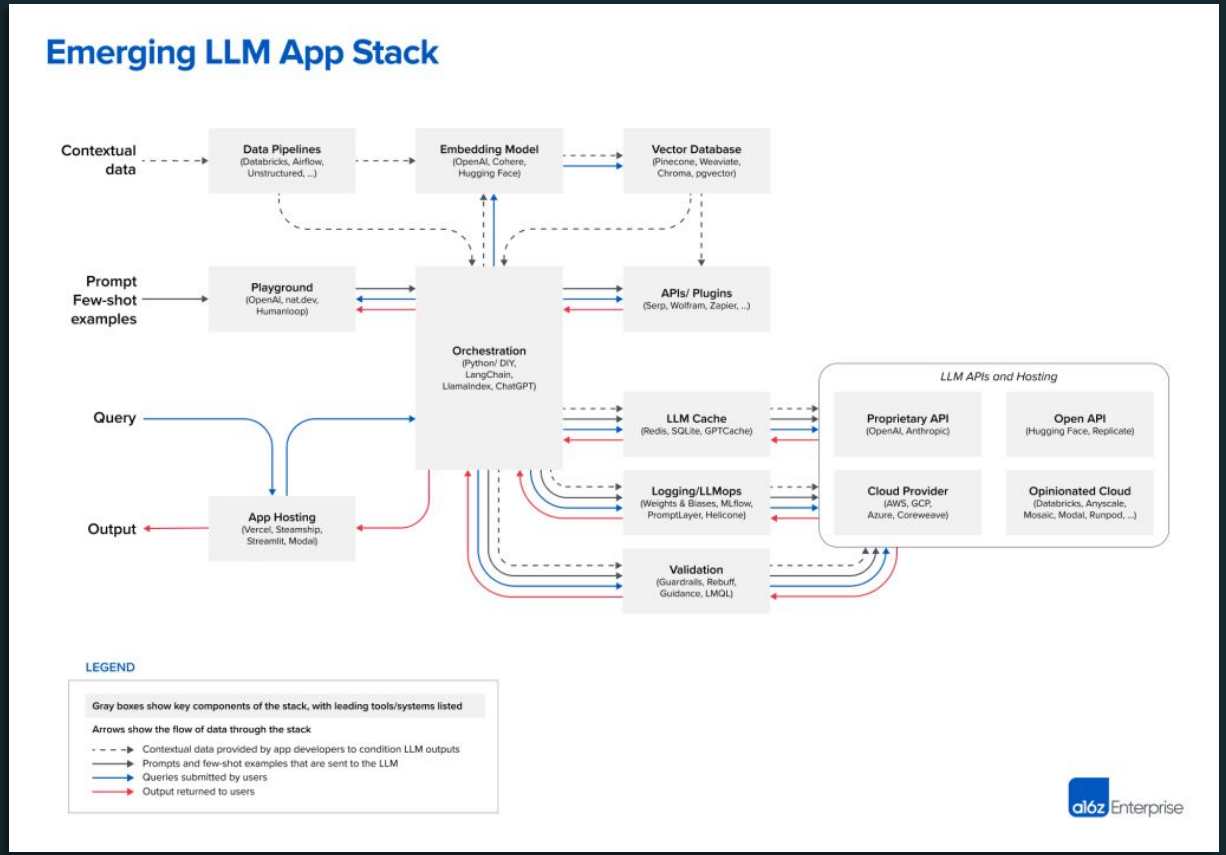
Source: https://lmql.ai/

chatGPT Plugins



Source: https://openai.com/blog/chatgpt-plugins

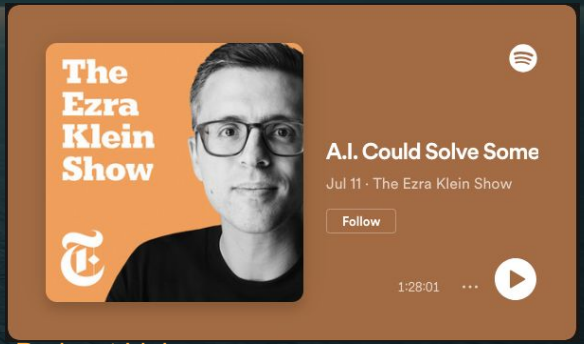
LLM App Stack



Source: https://a16z.com/2023/06/20/emerging-architectures-for-llm-applications/

Beyond the Horizon

- Multi-modal Models (successors of MUM, GPT-4, VIMA, PaLM-E)
- Even Longer Context Windows
- Efficient and Faster Training/ Smaller Robust Pretrained Models
- Improvements and Better access to RLHF
- Improvements in LLOps
- Improvements in handling attacks
- Improvements in controlling hallucinations



Podcast Link

Source: https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models

Thank You!