

Summer Internship Report (Summer Term 22-23)

Name- **Rohan Saha** | RollNo.- **20085086** | Electrical

Student Info: **Rohan Saha** | **20085086** | **Electrical Part-IV**

Company: **EXL Services, India**

Role: **Data Analyst Intern**

Location & Duration: **Gurgaon, 18th May 2023 – 14th July 2023**

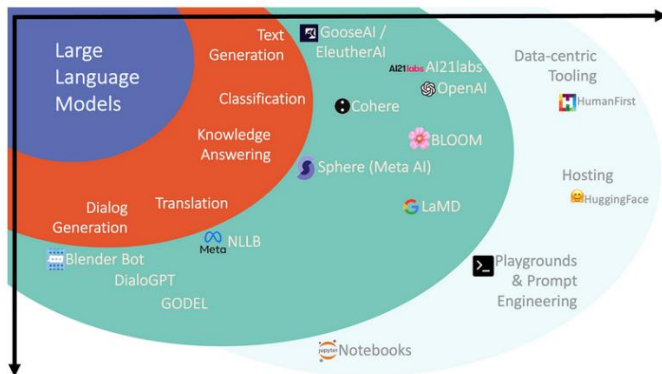
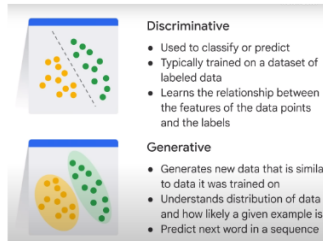
Project Title: **Servicing Insights | Generative AI – LLM Playbook**

EXL Analytics

Generative AI - LLM Playbook

GENERATIVE AI

- Generative AI is a type of AI System capable of **Generating Text, Images & other Media** in response to Prompts
- GenAI Models learn the **Patterns & Structures** of their Input Training Data, then generate new data with similar Characteristics
- Eg. **ChatGPT**, a ChatBot built by OpenAI using their GPT-3 & GPT-4 foundational LLMs; **Bard**, a ChatBot built by Google using their LaMDA foundational Model



LARGE LANGUAGE MODELS (LLM)

- LLMs are Deep Learning Models that **Understand & Generate Text** in a Human-like fashion, have become Indispensable for NLP
- Traditionally LLMs are **Pre-Trained** by Academic Institutions & Tech Giants like OpenAI, Google & NVIDIA, then made Available for **Public/Commercial Use**
- Businesses then focus on **Fine-Tuning** existing LLMs for Specific Use-Cases

LLM Playbook

WHAT IS AN LLM PLAYBOOK?

- The LLM Playbook Is Site/Portal that hosts a **Collection of Deployed LLMs** catered to **Specific Client Use-Cases**, all in One Place
- Easily **Accessible & User-Friendly**, allows Clients to use Powerful LLMs without requiring extensive technical knowledge

HOW TO BUILD AN LLM PLAYBOOK?

- Pick up a Use-Case, research LLMs for that Specific Use-Case
- LLMs must check all Boxes before Integration : Commercially Licenced, Computationally Convenient, Reliable, etc.
- After Checking the Feasibility & Use-Cases, deploy it Locally, explore it, understand its variables & parameters, Fine-Tune it with Open-Sourced Data
- Finally Integrate the LLM into the Playbook

WHY DO WE NEED AN LLM PLAYBOOK?

- Simplifies **Accessibility**, Reduces Technical Barriers makes advanced LLMs easily accessible to Individuals and Organizations without requiring extensive technical knowledge
- Collection of deployed LLMs tailored to **Specific Use-Cases**, ensuring that users can find models that directly address their needs without the need for customization or complex setup
- Pre-Trained and deployed LLMs eliminate the need for users to spend **Time & Effort** on Training and Deployment processes
- Quality Assurance**, LLMs undergo thorough Curation, Testing, & Deployment, ensuring they meet the Required Quality and Performance standards
- Democratizes Access**, Centralized platform that bridges the gap between LLMs and Users, enabling a wider range of individuals and organizations to leverage their capabilities

Use-Cases	Text Generation/ Q&A	Text Summarization	Named Entity Recognition (NER)	Document Q&A	Code Generation
Models	BLOOM, StarCoder, Dolly, FLAN, Falcon, Pythia, BERT, OPT, GPT-3, Open LLaMA, Alpaca, Vicuna	Pegasus, BART, DistilBART, GPT-3, T5	Falcon, h2oGPT, spaCy, BERT, RoBERTa	FLAN, RoBERTa, TAPAS, TAPEX, GPT-3, DocBERT	StarCoder, CodeT5, FalCoder, CodeBERT

Summer Internship Report (Summer Term 22-23)

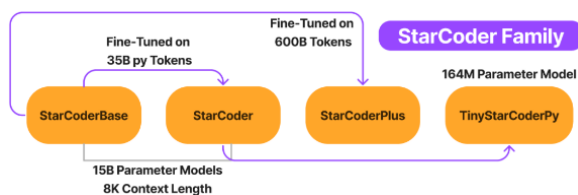
Name- **Rohan Saha** | RollNo.- **20085086** | Electrical

HOW TO BUILD AN LLM PLAYBOOK?	WHAT DID I DO?
Pick up a Use-Case , Research LLMs for that Specific Use-Case	I picked up Code-Generation , Researched & Experimented on StarCoder , CodeT5 , FalCoder [Code-Generation LLMs]
LLMs must Check all the Boxes before Integration- Commercially Licenced , Computationally Convinient , Reliable , etc.	✓
After Checking the Use-Cases & Feasibility, deploy it Locally, Explore it, Understand its Variables & Parameters, Fine-Tune it with Open-Sourced Data	We fixed the Hyper-Parameters according to our Needs, but couldn't Fine-Tune them due to Data & Architecture Inadequacy
Finally Integrate the LLM into the Playbook	Integrated the LLMs into the Playbook locally Deployed on Streamlit

For Eg. Research & Experimentation on **StarCoder** & **CodeT5**-

salesforce / **CodeT5**

StarCoder



Model	Parameters	Tested?	Description
tiny_StarCoder_py	164M	✓	15.5B >> Works really well but needs Colab Pro
StarCoder	15.5B	✓	fine-tuned from StarCoder to act as Coding Assistants
StarChat_alpha	15.5B	✓	

Model	Parameters	Tested?	Remarks
CodeT5 base	220M	✓	770M > 220M as more Parameters
CodeT5 large	770M	✓	
CodeT5+ 2B, 6B, 16B	2B, 6B, 16B	✗	Couldn't use due to Licensing

Use-Cases	StarCoder 15.5B	CodeT5 770M	Remarks
Text-to-Code	✓	✓	StarCoder > CodeT5 as more Parameters
Code-to-Code	✓	✓	
Code-to-Text	✓	✗	CodeT5 can't act as a Technical Assistant
Text-to-Text	✓	✗	

Streamlit Integration

Finally we **Integrate** the **Code-Generation LLMs** [StarCoder, CodeT5 & FalCoder] into the **LLM Playbook**, **Locally Deployed** on **Streamlit**. Streamlit is an **Open-Source App Framework** that turns **Data Scripts** into **Interactive & Sharable Web-Apps**, all in **Python**. Its **Advantages** are-

- Simplicity & Ease-of-Use, Beginner Friendly, Minimal Code
- Interactive & Dynamic Visualizations, Rapid Prototyping
- Seamless Integration with **Python Ecosystems**

