Analytics Vidhya

# Skills for Success

A Comprehensive and Industry-Aligned Curriculum

# Table of Contents

# Curriculum Overview

The GenAI Pinnacle program is meticulously crafted in response to the burgeoning demand in Generative AI technology, featuring elite instructors renowned for their profound expertise in the field. Our comprehensive curriculum delves into Large Language Models, Stable Diffusion Models, Reinforcement Learning, and the foundational elements of Data Science & Machine Learning, preparing you to excel in the rapidly evolving landscape of Generative AI.

# Large Language Models

## 1: Getting Starting with LLMs

- **Evolution of NLP**
  - Evolution of NLP: Pre Deep Learning
  - Evolution of NLP: Deep Learning
  - Transformers

- **Introduction to LLMs**
  - What is a language model?
  - What is a large language model?
  - Understanding Foundational Model
  - Different types of LLMs: Continuing the text
  - Different types of LLMs: Instruction Tuned
  - Encoder LLMs, Decoder LLMs and Encoder Decoder LLMs

- **The Current State of the Art in LLMs**
  - The Current State of the Art in LLMs
  - What Next?

# 2: Building LLM applications using Prompt Engineering

- **How to build LLM Applications?**
  - Overview of different methods to build LLM applications

- **Introduction to Prompt Engineering**
  - Why do we need Prompt Engineering?
  - What is Prompt Engineering?

- **Prompt Engineering using ChatGPT API and Open Source LLM**
  - Introduction to ChatGPT API
  - Prompt Engineering with Open Source LLM

- **Understanding Different Prompt Engineering Techniques**
  - Zero-shot Prompting
  - Few-shot Prompting
  - CoT Prompting
  - Self Consistency
  - ToT
  - Advanced Prompt Engineering Techniques

- **Building LLM App using Prompt Engineering**

# 3: Building Production Ready RAG systems using LlamaIndex

- **Introduction to RAG systems**
  - Why RAG systems?
  - What is a RAG system?
  - A brief overview of steps involved in a RAG Framework
  - Quiz

- **Introduction to LlamaIndex**
  - What is a LlamaIndex?
  - Why LlamaIndex?
  - Quiz

- **Components of LlamaIndex**
  - Overview of Different Components of LlamaIndex
  - Data Sources
  - Indexing
  - Embeddings
  - VectorDBs
  - Quiz
  - LLMs
  - Retrieval
  - Decoupling chunks used for retrieval vs. chunks used for synthesis
  - Adding Metadata
  - Quiz

- **Code: Build your first QA system on private data**

- **Evaluation of RAG systems**
  - Necessity of Evaluation
  - Retrieval Evaluation
  - Response Evaluation
  - Quiz

- **Code: RAG + Evaluation**

- **Advanced approaches for powerful RAG system**
  - Router Query Engine, SubQuestionQuery Engine
  - Structured Retrieval for Larger Document Sets: Metadata filters + Auto Retrieval
  - Document Hierarchies + Recursive Retrieval
  - Finetuning Embeddings
  - Finetuning with Retrieval Augmentation (RADIT)

- **Agents**

- **Building a powerful RAG system using LlamaIndex**

# 4: Finetuning LLMs

- **Introduction to Finetuning LLMs**
  - What is Finetuning LLMs?
  - Why Finetuning LLMs?
  - Research Papers and blogs for additional details
  - Quiz

- **Understanding different finetuning techniques**
  - Feature Extraction
  - Retraining all parameters
  - The problem with traditional methods
  - Quiz
  - Setting up the runpod instance
  - Code Part1: Finetuning NLP usecase using feature extraction and full finetuning
  - Code Part2: Finetuning NLP usecase using feature extraction and full finetuning
  - Research Papers and blogs for additional details

- **Introduction to Parameter Efficient Finetuning Techniques**
  - What is Parameter Efficient Finetuning Techniques?
  - How PEFT addresses the issues with Full Fine-tuning
  - Different types of PEFT methods
  - Quiz

- **Prompt Tuning Techniques**
  - Prompt Tuning
  - Prefix Tuning
  - Code:Finetuning NLP usecase using Soft Prompting Techniques
  - Quiz

- **Adaptor Modules**
  - (IA)3
  - LoRA
  - Quantization
  - QLoRA
  - AdaLoRA
  - Scaling and combining multiple loras
  - Code:Finetuning NLP usecase using PEFT LoRA
  - Code: Finetuning instruction following LLM using PEFT

# 5: Training LLMs from Scratch

- **Training LLMs from Scratch**
  - Scaling laws
  - Parallel and Distributed computing strategies
  - Challenges involved in training LLMs
  - Best practices to train LLMs
  - How do you train LLMs from scratch?
  - Project: Continuing text
  - Project: Instruct optimized + RLHF

- **Evaluating LLMs**

# Stable Diffusion Models

## 0: Embark on an Exciting Journey with Generative AI

- Revisit the Prerequisites: DL, ML, Optimization fundamentals needed.
- What is Generative AI?: Introduction, Landscape, Domains, Sub-Fields.
- Why Generative AI?: Why it matters, How to augment it, How to Capitalize on it.
- History of Generative AI: Short but Significant history, Milestones, Ahha moments.
- Why should you care about the history of Stable Diffusion?: Professional Impact, Societal Implications, Technology Reshaping Information Exchange.
- Applications of Generative AI: An Umbrella look at the Concrete applied Use Cases.
- Gauging the State of Art in Stable Diffusion: What is the forefront and Who is at the forefront.

# 1: Fundamentals of Diffusion Models

- Explanation of Diffusion Models and their purpose.
- Acquire the Intuition behind Stable Diffusion Model.
- Paper Review: Fergus & Zeiler :Visualizing and Understanding CNNs Gradients
- Paper Review: CLIP (Contrastive Language–Image Pre-training)
- Understanding Text and Image Embeddings and Their Mutual Relation.
- Tokens as Embedding: Understanding Nuances.
- Detailed analysis of the inner workings of Diffusion Models. Understand Math and how it looks into the code.
- Setup the Stable Diffusion Development Environment and attaining GPU/vCPU stage.
- Understand the training Paradigm of Stable Diffusions.
- Hugging Face’s Diffusers library, Setup of Hugging Face Spaces and API Key.
- Quick overview of Google Colab for Stable Diffusion.
- How to Generate Images using State of Art Stable DIffusion Model. Code and Exercise.

## 2: Understanding the Deepest Levels Needed for Diffusion Models

- **Paper Review: The VAE (variational autoencoder)**
- **What it means by Denoising Diffusion and Reverse Diffusion.**
- **Predicting noise with the UNet.**
- **Removing noise with schedulers.**
- **Understanding Critical Key Concepts:**
  - Pre-trained pipelines
  - Guidance scale
  - Negative prompts
  - Finite differencing
  - Analytic derivatives
  - Textual inversion
  - Latents
  - U-Nets
  - Text encoders and image encoders
  - Contrastive loss function
  - CLIP text encoder
  - Deep learning optimizers
  - Perceptual loss
- **How it all fits into code. Complete Code walkthrough of Stable Diffusion Method.**
- **Train a Stable Diffusion Model on GPU Cloud.**
- **Test your own Diffusion Model For Clothing Articles.**
- **Question and Answers: To make sure, Everybody got the Stable Diffusion Method fully.**

# 3: Stable Diffusion in Practice, Industrial Methods

- How Do We Train Stable Diffusion at Scale?
- Paper Review: Progressive Distillation for Fast Sampling of Diffusion Models
- Paper Review: On Distillation of Guided Diffusion Models
- Ethical Implications of Training a Stable Diffusion Models.
- Biggest Player contributing to Open Source Gen AI.
- How to Capitalize and Contribute to Open Source Stable Diffusion?
- Should you train Stable Diffusion from Scratch? Yes and No!
- What is considered valuable in Stable Diffusion Domain?
- Stability.ai: OpenAI, but a better and much more supporting approach.
- DreamStudio and StableStudio: Blessings of Stable Diffusion.
- Stable Diffusion WebUI Introduction and Purpose.
- Running Automatic1111 WebUI on Kaggle or any GPU Environment.
- Checkpoint:  Everybody is able to get to WebUI? Provide Assistance.

# 4: Methods, Jobs, and Tools of Stable Diffusion

- Analyzing Prompts, Prompts Matrix.
- Prompt Strengths and Weights.
- Prompt Editing and Blending for Stable Diffusion.
- Understanding and Using XYZ Plots for Stable Diffusion.
- Hands-On Exercises and Encouraging People to Try the ideas themselves.

## 5: Stable Diffusion Task to Achieve Purpose at hand

- Introduction: img2img. How to use it?
- Image In-Painting: Various forms and factors.
- Image Editing with img2img Sketch.
- Image Editing with In-Paint Sketch.
- CIVITAI: Introduction and fine-tuned models. Model Zoo of Stable Diffusion
- Stable Diffusion Extensions.
- MidJourney: The Storm, The Story, The Inspiration and Healthy Serendipity.
- MidJourney: How to exploit MidJourney for your purpose at hand.
- Adobe Firefly: Introduction and Demo
- Microsoft Designer: Introduction and Demo
- OpenAI DallE-2: Introduction and Demo
- Blue Willow: Introduction, Demo and Warnings
- Lexica Art Prompt Learnings

## 6: State of Art and Cutting Edge of Stable Diffusion

- Paper Review: InstructPix2Pix
- InstructPix2Pix Introduction and Setup.
- InstructPix2Pix Demo and hands-on.
- ControlNet: Introduction and Setup.
- Generating Image Variations with ControlNets.
- Awe Inspiring Use Cases of ControlNets.
- ControlNets OpenPose: Use of Human Pose in Image Editing.
- Quick Paper Review: Super Resolution
- UpScaling Images with Ultimate SD UpScale + ControlNets

## Module INFINITY: Way Forward for Stable Diffusion

- Where to go from here?
- How to keep a tab on this space?

# Reinforcement Learning

## 1: Mathematical Prerequisites for Reinforcement Learning

- **Markov Decision Processes**
- **Bellman equation and Dynamic Programming**
- **Value Iteration**
- **Policy Iteration**
- **Hands on experience – Jupyter notebook with simple numpy based tutorial with solution for Value Iteration and Policy Iteration**
- **Introduction to Partially Observable Markov Decision Processed and Games**

# 2: Simple Reinforcement Learning

- Temporal difference (TD) learning and Monte Carlo (MC) methods
- RL – framework: OpenAI Gym Environment
- Exploration vs Exploitation in RL
- Actor Only, Critic Only and Actor Critic Algorithms
- Q-learning
- SARSA
- REINFORCE
- Jupyter notebook tutorial with solution for TD, MC, Q-learning, SARSA, REINFORCE
- Discussion on online vs offline RL

# 3: Reinforcement Learning with Function Approximation

- Basic Introduction to Linear Function Approximation
- Deadly triad of Deep RL – function approximation, bootstrapping and offline learning
- DQN and variants
- OpenAI Spinning up based tutorial on DQN with solution
- Stochastic Policy Gradient Theorem
- PPO and variants
- OpenAI Spinning up based tutorial on PPO with solution
- Deterministic Policy Gradient Theorem
- DDPG, TD3, SAC
- OpenAI Spinning up based tutorial on TD3 with solution

# 4: RLHF for LLMs

- LLM Basics
- Types of human feedback
- Supervised Fine Tuning – Basics
- Reward Model from Human Feedback
- RL based LLM finetuning with PPO
- RL based LLM finetuning with ILQL
- TRLX based tutorial on finetuning GPT2 with PPO and ILQL
- Discussion on other RLHF open-source libraries.

# Basics of Data Science & Machine Learning

## 1: Python for Data Science

- Explore Python for Data Science
- Important libraries and functions in Python
- Reading file and manipulating data in python
- Working with data frames, lists, and dictionary

## 2: Data Exploration and Statistical Inference

- Use Matplotlib and Seaborn for data visualization
- Creating charts to visualize data and generate Insights
- Univariate and Bivariate analysis using python
- Perform Statistical Analysis on real-world datasets
- Build and Validate Hypothesis using statistical tests
- Generating useful insights from the data

## 3: Basics of Machine Learning

- Learn Important Machine Learning concepts
- Perform data cleaning and Preprocessing
- In-depth understanding of Basic ML models
- Linear Models, Decision Tree, k-NN
- Math Behind each Machine Learning Algorithm
- Building Classification and Regression Models

- Hyperparameter Tuning to improve model
- Solving real-world business problems using Machine Learning

## 4: Fundamentals of Deep Learning

- Important Concepts of Deep Learning
- Working of Neural Network from Scratch
- Activation Functions and Optimizers for Deep Learning
- Understand Deep Learning architectures (MLP, CNN, RNN, LSTM, GRUs)
- Explore Deep Learning Framework, PyTorch
- Learn to tune the hyperparameters of Neural Networks
- Build Deep Learning models to tackle real-life problems
- Use Transfer Learning for training models

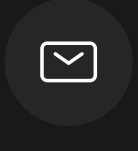
# Shaping Tomorrow with Generative AI

Enroll now

## Contact us:



08068342847 | 08046107668


training_support@analyticsvidhya.com


analyticsvidhya.com/genaipinnacle