LAB-1   **Date:**23/01/2026

Name: Ankana Mandal	**SRN:** PES2UG23CS076	**Section: B**

Documenting the Gen-AI Handson

Hugging Face can be thought of as the GitHub for Artificial Intelligence, where a wide range of open-source models, datasets, and projects are available. These resources demonstrate how modern AI models work and allow developers to reuse pre-trained models instead of building everything from scratch.

The transformers library acts as a bridge between Hugging Face and our code. It provides simple APIs to download and use pre-trained models in personal projects and supports multiple deep learning frameworks.

The pipeline() function is a powerful abstraction provided by the library. It combines preprocessing, model inference, and postprocessing into a single callable function, making it easy to perform complex NLP tasks with minimal code.

In this hands-on session, we compared two generative models: distilgpt2 and gpt2. The distilgpt2 model is a smaller, distilled, and faster version of gpt2, requiring less memory and providing lower latency. The standard gpt2 model, on the other hand, is larger and produces higher-quality outputs.

A random seed was set initially to make the experiment reproducible. This ensures that the same inputs produce consistent outputs across multiple runs, which is important for experimental evaluation.

When comparing the generated text from both models for the same prompt, it was observed that the gpt2 model produced more coherent and grammatically correct text. The distilgpt2 model, while reasonably good, showed repetition and gradually lost context over longer outputs.

Tokenization is a critical preprocessing step before passing text into a model. Since models cannot directly understand words, the text is broken down into smaller meaningful units called tokens, which are then converted into unique numerical IDs for efficient processing.

The temperature parameter controls randomness in text generation. Higher temperature values allow the model to choose less probable words, making the output more creative, while lower values make the output more deterministic and focused.

Part-of-Speech (POS) tagging is another important preprocessing step, where grammatical labels are assigned to words. For example, identifying whether the word "will" is used as a noun or a modal verb helps the model understand sentence structure more accurately.

Named Entity Recognition (NER) is used to identify and classify entities such as names, organizations, and locations. For instance, the word "Apple" may refer to a fruit or the company, and NER helps distinguish between these meanings based on context.