Supplementary Material for ' Exploring Large Language Models (LLMs) through interactive Python activities': Glossary of Key Terms

Term	Definition
Artificial Intelligence (AI)	A broad field of computer science focused on creating systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, problem-solving, perception, and language understanding (see Figure 1).
Attention Mechanism	A mechanism in models like Transformers allowing focus on relevant parts of the input sequence when making predictions.
Continuous Bag-of-Words (CBOW) / Skip-Gram	Two techniques used to train neural networks such as Word2vec: CBOW predicts a target word from its context words. Skip-Gram predicts context words given a target word (often better for rare words).
Cosine Similarity	A measure of similarity between two non-zero vectors based on the cosine of the angle between them; used in Natural Language Processing (NLP) to quantify semantic similarity between word embeddings (value typically ranges from -1 to 1, or 0 to 1).
Deep Learning	A subfield of Machine Learning based on artificial neural networks with multiple layers (deep neural networks) that learn hierarchical representations of data (see Figure 1).
Deep neural Networks	Artificial neural networks characterized by having multiple layers ('hidden layers') between the input and output layers, allowing them to learn complex, hierarchical patterns in data.
Embedding Dimension	The number of components (dimensionality) of the numerical vectors used to represent words or tokens in an embedding model.
Embedding Space	The multi-dimensional vector space where word embeddings reside; relative positions and directions of vectors typically encode semantic relationships.
Generative Models	AI models trained to generate new data (e.g., text, images) that mimics patterns observed in their training data. Generative Pre-trained Transformer 2 (GPT-2) is an example of a generative language model.
Google Colab	A free cloud-based platform provided by Google that allows users to write and execute Python code through their browser, mainly by using Jupyter Notebooks, and offers access to computing resources like GPUs.

Generative Pre-trained Transformer 2 (GPT-2)	A specific Large Language Model (LLM) released in 2019 by OpenAI based on the Transformer architecture, capable of generating coherent text by predicting subsequent tokens; used in this paper's activities.
Jupyter Notebook (JN)	An open-source web application and file format (.ipynb) that allows creating and sharing documents containing live code (e.g., Python), equations, visualizations, and narrative text; used for the activities in this paper.
Large Language Model (LLM)	An advanced AI model, typically based on the Transformer architecture and trained on vast amounts of data (often including text, but also images, code, etc.), capable of understanding complex patterns and generating diverse outputs, including human-like text, for various tasks.
Machine Learning (ML)	A subfield of Artificial Intelligence where computer systems learn to perform tasks by identifying patterns in data, without being explicitly programmed for each task (see Figure 1).
Natural Language Processing (NLP)	A field of Artificial Intelligence focused on enabling computers to process, understand, interpret, and generate human language.
Principal Component Analysis (PCA)	A statistical technique for dimensionality reduction by identifying principal components; used in this paper's activities to visualize high-dimensional word vectors in 2D.
Python	A versatile, widely-used high-level programming language known for its readability and extensive libraries, particularly suited for scientific computing, data analysis, and AI development.
Prompt	The input text, question, or instruction provided by a user to a large language model to guide its response or text generation.
Scaling Law (in LLMs)	Empirical findings suggesting that LLM performance predictably improves as model size, dataset size, and computational resources increase.
Temperature (LLM parameter)	A parameter controlling the randomness of output during text generation in LLMs; lower values yield more predictable text, higher values increase diversity. (Not related to physical temperature).
Token / Tokenization	Token: A fundamental unit of text (often a word or sub-word) processed by language models. Tokenization: The process of splitting input text into a sequence of tokens.
Transformer Architecture	A neural network architecture, foundational to models like GPT-2, that uses attention mechanisms to weigh the importance of different input tokens when processing sequence data, enabling better handling of long-range dependencies.

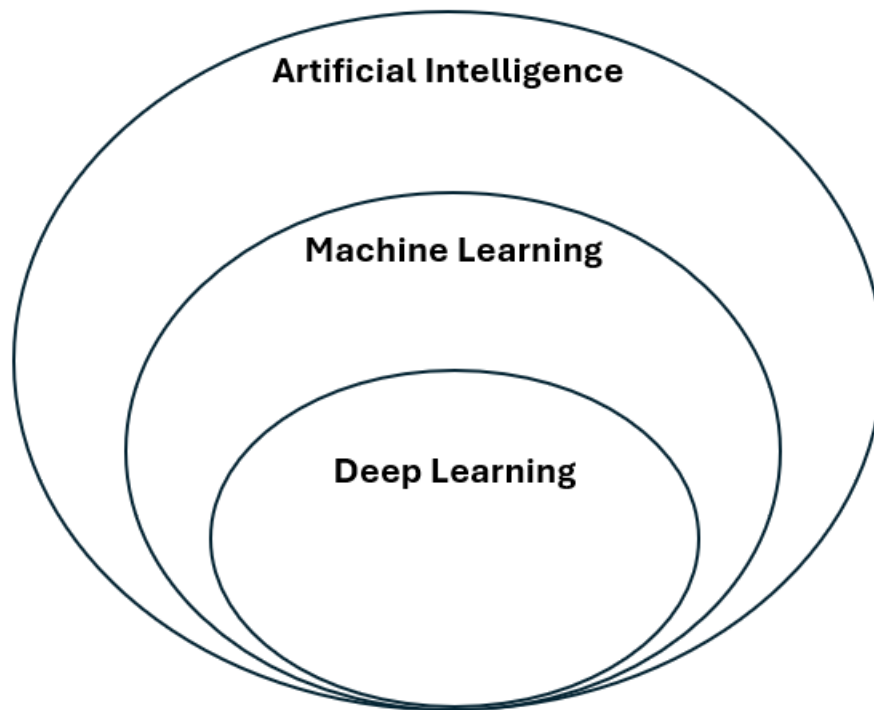| Word Embedding	A numerical vector representation of a word in a multi-dimensional space, learned from text data such that the vector captures semantic properties and relationships of the word.
Word2Vec	A specific neural network model published in 2013 designed to learn static word embeddings by analyzing word co-occurrence patterns in a large text corpus.

Figure 1