A Glossary of AI-Related Terms[1]

Basic Terms and Concepts

AGI (Artificial General Intelligence): An AI system that can approximate human performance on most intellectual tasks and activities. It is generally recognized that AGI has not yet been attained. However, there are different opinions about whether and when this could happen.

AI (Artificial Intelligence): The design and programming of computers to give them the ability to simulate human intelligence. The term was coined in 1956 by pioneering computer scientist John McCarthy.

ANI (Artificial Narrow Intelligence): AI designed to perform a relatively narrow range of functions. It is generally held that all current AI systems should still be classified as ANI. However, more advanced models are demonstrating increasing capabilities in a greater variety of tasks.

ASI (Artificial Superintelligence): An AI system that is more capable than humans in all respects. ASI is still only speculative.

GPTs (Generative Pre-trained Transformers): An AI model that produces human-like outputs based on training data. Different types of GPTs may produce text, audio, images, or videos.

LLMs (Large Language Models): AI models that can interpret and produce human language based on context. They currently function by predicting (generating) the next word, considering the great flexibility of language usage and the larger context of communication.

Machine Learning: A method of training neural networks to recognize patterns and make predictions based on data.

Neural Networks: Computing systems structured like neurons in the human brain.

Prompts: Natural language inputs users give to AI models to elicit responses or perform tasks.

Reinforcement Learning: A type of machine learning in which a model is trained to make decisions by being given mathematical rewards or penalties for its responses.

Training Data: Information that AI learns from to make predictions.

[1] This list was generated with the help of Microsoft Copilot. It was revised and checked for plagiarism with the help of Grammarly.

Additional Terms and Concepts

Agent, Agentic AI: AI that can act autonomously (without human control) to achieve specified goals. AI agents must be able to interact with their environment (whether physical or digital), make decisions, and take resulting action.

Alignment: Refers both to how well AI systems operate according to human values, goals, and ethical principles and to the process of aligning them.

Bias: Comprises systematic model errors that can lead to unfair or inaccurate results. Biased outputs can be due to various factors, including biased training data and how the model is trained. Attempts to correct certain biases in AI systems can introduce other biases.

Chatbots: AI programs designed to simulate conversation with human users.

Compute: When used as a noun, compute refers to the computational power required to train a model or solve a problem.

Context Windows: The amount of text or data that an AI model can consider at one time when making predictions or generating responses.

Deep Learning: A type of machine learning performed with neural networks that have many layers. It produces AI systems that can analyze complex data and generate complex outputs.

Deepfakes: AI-generated images, videos, or audio that appear authentic.

Emergence: The phenomenon in which combining many simple processes results in complex and often surprising capabilities.

Finetuning: Making minor adjustments to a model that has already been trained to optimize its performance in specific tasks.

Frontier Models: Cutting-edge AI models that push the boundaries of what is currently possible in the field. Each generation of Frontier Models has been about an order of magnitude larger than its predecessor in recent years.

Guardrails: Safety measures implemented in AI systems to prevent harmful or unintended outcomes. Guardrails are not foolproof, sometimes allowing themselves to be bypassed or defeated and sometimes blocking a response to a legitimate prompt.

Hallucinations: Instances where AI models generate incorrect or nonsensical information (i.e., they sometimes make things up). As models have advanced, the problem of hallucinations has been reduced but has not been eliminated.

Inference: The phase where an AI model makes predictions or decisions based on new data.

Jailbreaking: Manipulating an AI system to bypass its safety measures or restrictions.

Multimodal Models: AI models that can combine processing of various types of data, such as text, images, video, and audio.

Natural Language Processing (NLP): The ability of computer systems to interact using everyday human language.

Predictive AI: Attempts to determine the probability of future outcomes based on analysis of past data. A typical example is using AI in online stores to suggest future purchases by analyzing the customer's buying history.

Prompt Engineering: Designing prompts to elicit the best possible responses from AI models.

RAG (Retrieval-Augmented Generation): RAG provides documents or other resources to which the AI can refer, enhancing the accuracy and relevance of AI outputs. Many LLMs can respond only based on information before their training cutoff date. However, some models can search the web or other data sources to produce more current responses.

Recommendation Systems: AI systems that suggest products, services, or information based on analysis of past preferences and behavior.

RLHF (Reinforcement Learning from Human Feedback): A technique where AI models are trained using human feedback to improve their performance.

Scraping: The process of extracting data from websites or other sources for AI training. Training data for Frontier Models is often obtained by scraping.

Style Transfer: Using AI to produce one image in the style of another. One debate in AI is whether using an artist’s style to create new images violates the artist’s legal or ethical rights.

Supervised Learning: A type of machine learning in which the model is trained on labeled data to analyze unlabeled data.

Tokens: Units of text (such as words or sub-words) that AI models process when generating or analyzing language.

Transfer Learning: A technique where a pre-trained AI model is adapted to perform a new but related task.

Turing Test: A test proposed in 1950 by mathematician and computer science pioneer Alan Turing. Turing proposed that instead of deciding whether a machine

could reproduce human intelligence, the test should be whether the machine could produce responses that a human could not distinguish from those of another human.

Unsupervised Learning: A type of machine learning where the model is trained on unlabeled data and must find patterns and relationships in the data on its own.