AI GLOSSARY

Artificial Intelligence (AI): an umbrella or generic term, like the word 'vehicle,' which refers to a range of technologies. The term is generally meant to reflect computers mimicking human intelligence, for example, emulating cognitive tasks like learning, reasoning, problem-solving, and understanding language.

Agentic AI: a more advanced AI system that can operate autonomously to achieve complex goals with minimal human intervention. For example, ordinary LLMs can plan an itinerary. Agentic AI can independently implement the itinerary by booking flights and hotels. Fully Agentic AI is still in development.

AI Agent: the manifestation of an agentic AI system. AI agents use GenAI to perform tasks autonomously, often employing multi-step processes, without human intervention. Examples of AI agents include chatbots on a website that answer questions and resolve issues without human intervention, AI agents that assist software developers with coding and systems testing, and AI agents on a physician's website that schedule appointments or facilitate renewing a prescription.

Algorithm: a set of instructions to complete a task. A recipe to bake a cake is one type of algorithm. Computer algorithms are sets of instructions given to a computer to complete specific tasks. AI uses algorithms to learn from vast amounts of data, identify patterns, and make predictions. For example, an AI system might use algorithms to review applicants to be selected for a job interview. The neutrality of these algorithms might be challenged in an employment discrimination lawsuit.

Alignment: a field of research that studies ways to design AI systems that behave in a manner consistent with human-centered, ethical principles. AI, though developed by humans, does not have values. Alignment seeks to mitigate the potential for an AI system to produce outputs that are inaccurate, biased, or harmful. It does this by exploring methods to prevent the AI system – which seeks to optimize its goals - from taking steps that are inconsistent with ethical norms. An example of misalignment is an AI system that makes up rules or cheats to win a game against a human. Alignment is an ongoing process that begins with system design, continues through LLM pre-training, includes different types of fine-tuning, such as reinforcement learning, and incorporates continuous monitoring. At the present time, fully realized alignment is aspirational.

Artificial General Intelligence (AGI): a (still) theoretical advanced type of AI. Narrow (weak) AI can perform specialized tasks within a narrow scope. General (strong) AI can perform any task at least as well as a human. Super AI, often referred to as Artificial General Intelligence (AGI), is

theoretical AI that can perform better than a human. There are currently debates over whether and when AGI will be achieved.

Black-box Technology: parts of an AI system in which the mechanisms of operation are opaque and not well understood.

Context Window: the maximum amount of information, measured in parts of words (tokens), that an LLM can process, analyze, and remember at one time. It is the equivalent of the AI system's short-term memory.

Discriminative or Predictive AI: traditional AI that categorizes data or makes predictions based on its training data. In electronic discovery, technology-assisted review (TAR) systems use discriminative AI to distinguish between relevant and non-relevant documents based on labeled training examples.

Frontier Model: the latest and most advanced versions of large GenAI systems. They often have multi-modal capabilities (e.g., text, image, audio, and video). They are capable of complex reasoning, coding, and agentic-like actions. They require massive computing resources, including electricity, to train and deploy. These models are general purpose and adaptable, but they are not always reliable. Because of their uneven abilities across tasks, they are often referred to as having a 'jagged edge.'

Generative AI (GenAI): a subset of AI that involves training on massive data sources—primarily from the internet—to generate content in response to a user prompt. It can converse and replicate specific styles. For example, a user can train an AI model to write in their own voice. GenAI excels at creative tasks and synthesizing and summarizing complex content.

Hallucination: when LLMs generate fluent but factually incorrect, nonsensical, or fabricated information. A hallucination may occur when the model's training was based on limited or contradictory information related to the subject matter of the prompt.

Large Language Model (LLM): advanced AI systems trained on massive textual data sets drawn from the internet and other proprietary sources. Their purpose is to understand, generate, process, and predict language.

Machine Learning (ML): a subset of artificial intelligence that enables a system to learn and improve from experience. There are at least four types of machine learning systems that can accomplish tasks of increasing levels of sophistication:

- **Supervised Learning:** trained by humans using labeled training examples (e.g., this is a relevant document / this is not a relevant document). The system infers rules and patterns from the training data to make educated guesses about new unlabeled data.

- **Unsupervised Learning:** uses algorithms to look for naturally occurring patterns, clusters, groupings, or anomalies in a data set. Programmers don't instruct the system to look for a particular pattern. The system does this automatically. This technology is useful for learning about what is in a dataset.
- **Reinforcement Learning:** machine-learning method where the system is trained by feedback (positive or negative) in response to the system's output. The system learns to optimize for positive behaviors.
- **Deep Learning:** multiple layers of neural networks (algorithms simulating the human brain). Information from each layer is combined with information at the next layer in increasing levels of complexity. At the highest layer, there is either a decision or an action taken. Deep learning enables state-of-the-art AI applications like computer vision, natural speech recognition, and self-driving cars. The precise mechanics of deep learning are not well understood which is why it is often referred to as a black-box technology.

Natural Language Processing (NLP): algorithms that enable computers to understand and generate human language as it is written and spoken. Examples of technologies that use NLP are predictive text like autocorrect, personal assistants (e.g. Alexa and Siri), and Google Translate.

Retrieval Augmented Generation (RAG): a technique that connects LLMs to external, up-to-date, sources (e.g., databases or legal cases) to provide more accurate and factually grounded outputs. RAG can reduce but does not eliminate the LLM hallucination problem.