AI Fundamentals

A PLAIN-LANGUAGE GUIDE FOR TERMS, CONCEPTS, TECHNOLOGIES, AND BUILDING BLOCKS RELATED TO ARTIFICAL INTELLEGENCE IN PSYCHOLOGICAL PRACTICE

How to Use This Document

This guide is written specifically for clinical and forensic mental health practitioners. It provides a comprehensive, plain-language introduction to the computer science terms, architectural frameworks, and clinical considerations that govern modern artificial intelligence (AI) systems.

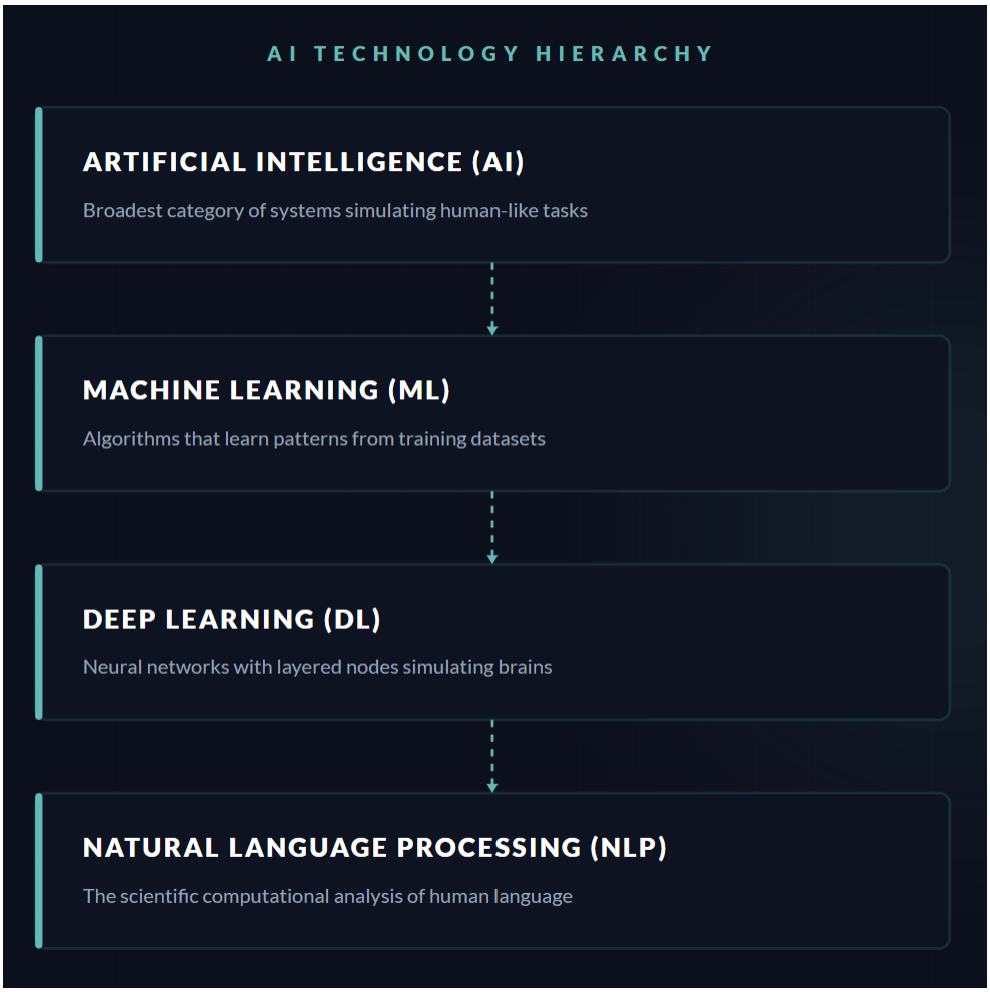
Understanding these foundational concepts is essential for developing and maintaining professional competence, satisfying ethical imperatives, and, if necessary, testifying confidently on the stand if your use of AI-assisted technologies are challenged in Court.

PREPARED BY: ARGENT FORENSICS
WWW.ARGENTFORENSICS.COM
MAY 2026

Part One: Defining the Technology

What is Artificial Intelligence and How is it Categorized?

Practitioners must understand the categories of computer science that drive modern analytical tools.



Artificial Intelligence (AI)

The broadest parent category in computer science, referring to any system, algorithm, or program capable of performing tasks that historically required human cognitive processing.

Machine Learning (ML)

A subfield of AI focused on building algorithms that are not explicitly programmed with rigid rules, but instead "learn" to identify statistical patterns and make predictions by analyzing massive training datasets.

Natural Language Processing (NLP)

The mathematical analysis of human language by computer systems. Algorithms convert raw, unstructured words and paragraphs into structured numerical values, allowing software to perform functions such as semantic search, translation, sentiment/context analysis, and summarization.

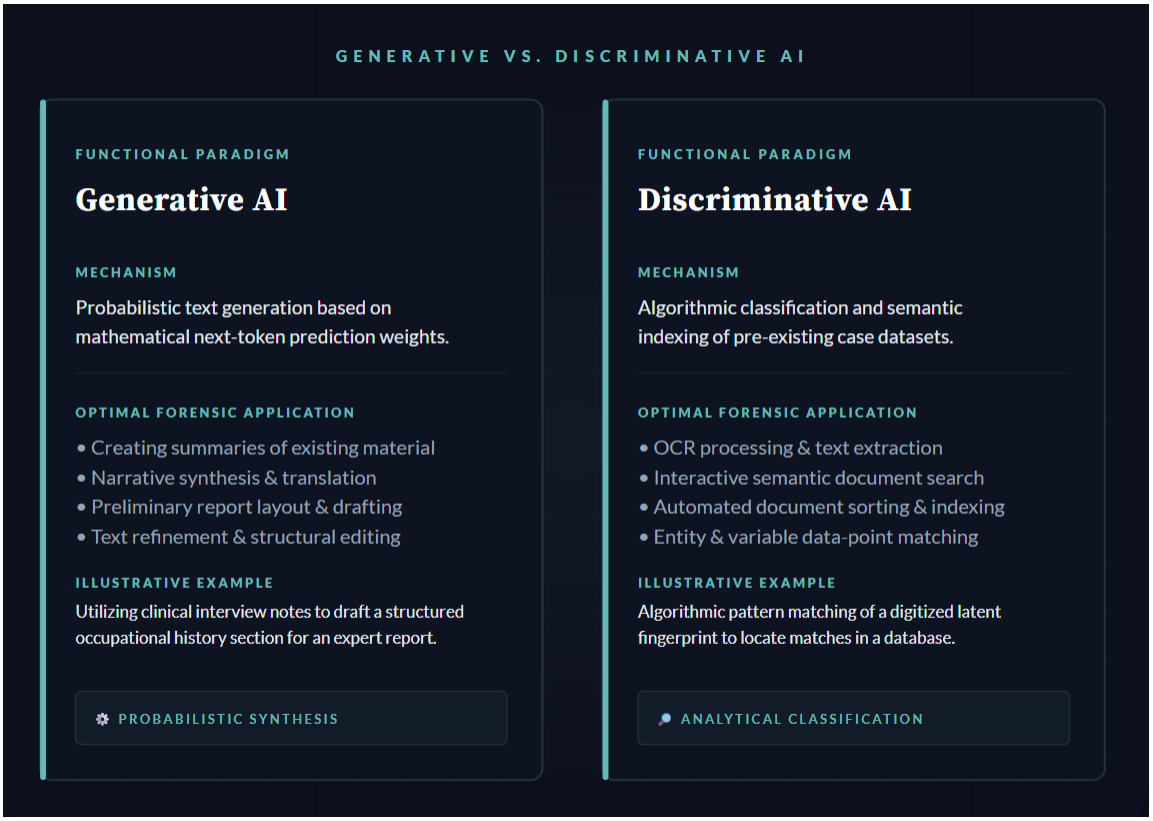
Large Language Models (LLMs)

The specific class of deep learning/NLP models designed to understand, process, and generate human-like text. Trained on vast corpora of written language, LLMs utilize a highly sophisticated mathematical architecture called a **Transformer** to calculate the probability of which word (or segment of a word) should logically follow another.

Part Two: Generative vs. Discriminative AI

Novel creation versus classification or filtering

When evaluating technology for forensic use, practitioners must distinguish between Generative AI (which writes new text) and Discriminative or Analytical AI (which classifies, filters, or searches existing data). Discriminative AI has been functioning in clinical and forensic environments for a long time. It is the development of generative AI that has led to more intentional guardrails and more intensive ethical scrutiny.



Generative AI (GenAI or gAI)

A subfield of the subfield of machine learning and the output produced by LLMs. These systems utilize deep neural networks to produce *novel content*, including text, images, video, or audio, that resembles human creations. At their baseline function, generative language models do not retrieve facts; instead, they "predict" or "generate" a plausible-sounding sequence of words based on the mathematical patterns they learned during training.

While discriminative AI has been utilized in high-stakes settings such as medicine and digital forensics for more than a decade, it was the progression to capable generative AI that introduced considerably more significant scrutiny on psychological applications. This is due to the potential introduction of error (discussed in later sections) as a result of the generative LLM architecture and probabilistic output production, as well as the introduction of potentially blurred lines for misattribution of authorship. As a result, the American Psychological Association has produced a number of guidance documents to help clinicians navigate this new frontier of AI use in clinical settings.

Discriminative/Analytical AI

Unlike generative models, discriminative algorithms are designed to analyze, categorize, and filter *existing* datasets. In forensic practice, standard search engines, optical character recognition (OCR) software, and data-classification databases are discriminative. They do not generate new clinical prose; rather, they identify, sort, and isolate the exact data points already present in the primary files.

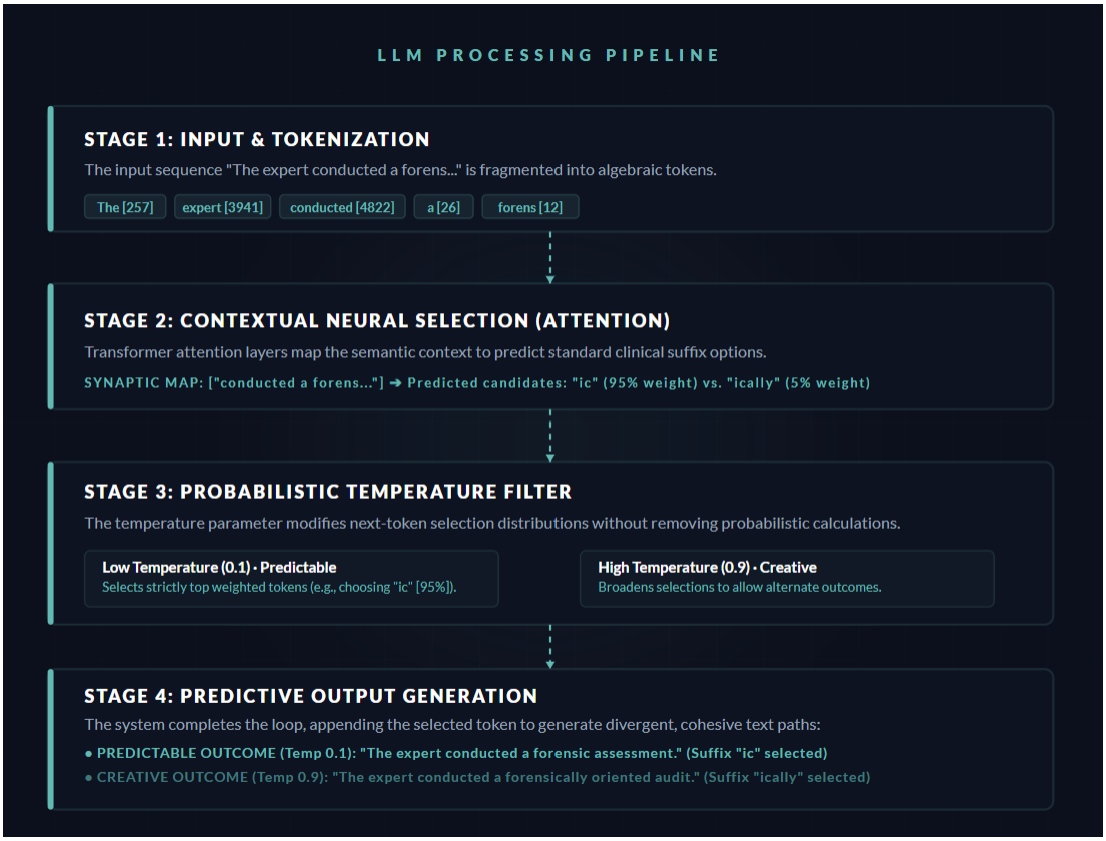
Optical Character Recognition (OCR)

Technology that converts images of text—whether from scanned documents, photographed pages, or typed PDFs—into machine-readable text.

Part Three: The Architecture of a Large Language Model

How LLMs Work

To understand why an LLM responds the way it does, a clinical practitioner must understand the core components of the model's processing pipeline: Tokens, Neural Processing, and Temperature.



Prompt

The natural language instructions, queries, or content inputted into an AI platform to direct its output.

Tokenization

Computers cannot read characters or words directly. Before an LLM can process text, it must break down words into smaller mathematical fragments called Tokens (typically representing a few characters or a common sub-word). These tokens are assigned a unique numerical ID. *Example:* The word "Forensic" may be tokenized into three distinct numerical identifiers representing "For", "ens", and "ic".

Probabilistic Prediction

An LLM does not possess consciousness, clinical insight, or a conceptual understanding of psychology. At its core, the model is a probabilistic next-token predictor. When a prompt is entered, the model's neural network calculates a probability distribution across its entire vocabulary, determining which token is mathematically most likely to appear next. This is fundamentally distinct from deterministic computer software that produces the same function predictably and repeatedly. This is why an LLM will typically not produce the exact same output twice, even to identical inputs. *Example:* If the input text is "A forensic psychologist conducts an independent...", the model calculates that the token "evaluation" has a 92% probability of following, while "recipe" has a 0.001% probability.

Temperature (Creativity Variables)

LLMs feature a configurable mathematical setting called Temperature (ranging from 0.0 to 1.0) that controls how strictly the model adheres to the

highest-probability next token. Low temperature settings (e.g. 0.0 to 0.2) move the model closest to deterministic outputs. It will almost always select the absolute highest-probability word, resulting in a more consistent and predictable, as well as more factual output. This is the setting utilized in systems build for clinical and forensic settings. High temperature settings (e.g. 0.7 to 1.0) introduce more deliberate randomization in responding. The model will select lower-probability words, producing "creative," diverse, and varied outputs. While this tends to be desired for more artistic applications, it is typically inappropriate for psychological functions.

Transformer

The breakthrough neural network architecture that enables language models to process words in relation to all other words in a sentence simultaneously, instead of sequentially, allowing the AI to understand linguistic context.

Context Window

The maximum amount of data (tokens) that an LLM can hold in active memory during a single prompt-and-response interaction. A larger context window allows the system to read and synthesize lots of information simultaneously. Extremely large context windows are susceptible to a phenomenon known as "context rot" (also called the "lost in the middle" phenomenon). This is where the model's attentional recall degrades, causing it to overlook crucial details within larger data sets, and prioritizing information from the beginning or the end.

Vector Database

A specialized database that stores mathematical representations of text meaning—called embeddings or vectors—enabling semantic search based on conceptual relationships rather than simple keyword matches.

Part Four: The Anatomy of a “Hallucination”

How and Why Errors Occur

Psychological practitioners must be concerned with reducing error rates and increasing the rate of repeatable (reliable) and useful (valid) output.

Hallucinations (Confabulations)

In computer science, a Hallucination (or, more accurately to psychological terminology, a Confabulation) refers to an AI output that is seamlessly executed grammatically and presented with confidence, but is factually incorrect and/or ungrounded in a real-world data source.

Confabulations Are an LLM Feature, Not a Bug

In general-purpose generative models (like ChatGPT, Gemini, or Claude), fabricated content is perceived as less of a malfunction and more of the system's baseline state. Because these models are designed strictly to predict the most plausible-sounding sequence of words rather than query a database of factual truths, they do not possess a mechanism to distinguish between a historical fact and a statistically probable fiction. When an LLM does not have sufficient training data to produce a factual result, or if it is given an ambiguous prompt, it typically does not stop to recognize that it does not know the answer. Instead, it calculates the most linguistically satisfying sequence of words, creating, for example, plausible-sounding psychiatric diagnoses, hospitalizations, or legal case citations that do not exist. In many cases this behavior is reinforced with programming efforts to maximize engagement and enhance speed, rather than to prioritize strict accuracy.

For most psychological practitioners, and certainly those in forensic settings, introducing an unverified AI hallucination into a formal represents a catastrophic blow to professional credibility. This structural limitation is why open-domain, consumer-grade AI tools (such as public ChatGPT, Gemini, or Claude) must be carefully considered and strenuously vetted before integrating them into any clinical functions (there is also a privacy/security issue that will be discussed in later sections of this overview). However, with psychologically-specific prompt engineering and proper task designation designed to recognize high-risk circumstances to reduce confabulated content, LLMs can be helpful to a variety of clinical tasks as long as human verification remains paramount.

Retrieval-Augmented Generation (RAG)

An AI architecture that combines document retrieval with language model generation. The retrieval component searches a specific, closed document set for relevant information, while the generation component synthesizes that retrieved information into a response.

Human-in-the-Loop

A design principle in which a human expert reviews, verifies, and takes responsibility for AI-generated output

Part Five: Compliance, Security, and Privacy Standards

Keeping Data Protected, Secure, and Confidential

It is imperative that practitioners attend to a variety of obligations related to securing clinical and forensic data. This may evoke state or federal regulations (e.g. data privacy acts, HIPAA), licensing rules or statues (e.g. psychology practice acts), Ethical guidelines or mandates (e.g. APA Ethics Code, Specialty Guidelines for Forensic Psychology), or legal mandates (privileged material). Breaches of security or privacy can introduce licensing, malpractice, or other legal liability.

HIPAA-Compliance and Protected Health Information (PHI)

The Health Insurance Portability and Accountability Act (HIPAA) mandates strict administrative, physical, and technical safeguards for handling patient and examinee medical records. Public consumer AI tools are not HIPAA-compliant because they do not protect PHI from data leakage, data exposure, or external model training. This does not mean that public models can never be used for clinical functions, but it does mean that PHI can never be used with them. For example, public LLMs can be helpful with rote/routine report copy (e.g. an introduction to the MMPI), test data table formatting (provided PHI is removed and that it is vetted thoroughly for accuracy), hypothetical consultation (e.g. "What might be some alternative explanations besides schizophrenia for a 50 year-old female to have recently developed psychosis?"). Functions that otherwise require inputs or outputs with PHI must always go to secure, HIPAA-compliant platforms, and ideally those built specifically for psychological use.

Business Associate Agreement (BAA)

Under HIPAA, any third-party technology platform that processes, stores, or transmits PHI on your behalf is legally defined as a Business Associate. To operate legally, you must sign a contractually binding Business Associate Agreement (BAA) with the vendor. A BAA contractually binds the technology provider to enforce federal security standards and assumes joint legal liability for data protection. Determining if a platform is willing and able to sign a BAA is the most important factor when determining if is HIPAA compliant because this bounds them to most of the requirements that obligate secure data handling. Anyone willing to sign a Business Associate Agreement (BAA) with you is making a contractual declaration to protect data to the same standard you would. So, for example, it is extraordinary unlikely that anyone who signs a BAA with you will use non-encrypted mechanisms or use your data for public model training, because it would violate their obligations to protect PHI.

You do not require a new BAA with each platform update, as they are generally drafted to serve as a general agreement that you both follow HIPAA standards, but of course if you have questions or concerns about this you should take them to the developer and/or an attorney for specific review.

The other thing to be aware of is that most AI software companies do not have their own in-house AI models. Instead, they maintain their own BAAs with the large LLM providers (e.g. OpenAI, Anthropic, Google, Microsoft), and these BAAs contractually prohibit those companies from storing, caching, or using any of their users' data, prompts, or outputs to train or optimize future iterations of their commercial AI models. One distinction to be aware of is that there can be variations of what is considered "your" data. Again, if you have questions or concerns about this with a particular platform, you should take them to the developer and/or an attorney for specific review.

Encryption

To ensure that case files cannot be intercepted or tampered with, professional platforms must encrypt data. At its most fundamental level, encryption is the mathematical equivalent of placing data behind a complicated lock. It converts readable, sensitive information ("plain text") into an unreadable, scrambled format ("cipher text") using a complex, muti-layered algorithm. This scrambled data can only be decoded back into its original readable form by an authorized party with the corresponding mathematical "key." In professional practice, encryption serves as a digital secure vault, so that even if someone intercepts or physically takes the data files they remain unreadable, and therefore protected.

To ensure that case files cannot be accessed, intercepted, or tampered with, professional platforms must enforce two distinct cryptographic safeguards:

- Encryption in Transit: Cryptographically protects and scrambles data while it is actively moving over the internet between your local

browser and the secure cloud server, preventing anyone from intercepting it during transmission.

- Encryption at Rest: Cryptographically scrambles data while it is stored on physical servers, using healthcare-grade security keys so that saved files remain fully unreadable even if someone takes or accesses those physical storage drives.

CONTACT AND FURTHER INFORMATION

Argent Forensics LLC

www.argentforensics.com/resources

support@argentforensics.com

Additional professional resources, including Sample Informed Consent and Report Disclaimers, APA AI Tool Evaluation Checklist, and Testimony Preparation documents, are available for free download at www.argentforensics.com/resources.

This guide was prepared by Argent Forensics LLC for the professional use by psychological practitioners. It is intended for educational purposes only. Argent Forensics LLC assumes no responsibility for the outcomes of its use in any specific professional or legal context.