The Complete Generative AI Leader Dictionary (100 Terms)

Created by Vladimir Raykov

I. Core AI & Machine Learning Concepts

1. **Artificial Intelligence (AI):** Teaching computers to perform tasks that typically require human intelligence, like learning, problem-solving, and decision-making.
    - *Think:* Smart software that can mimic human thinking.

2. **Machine Learning (ML):** A type of AI where systems learn from data to improve their performance on a specific task without being explicitly programmed for each case.
    - *Think:* Computers learning from examples, like a child learns to identify a cat after seeing many pictures of cats.

3. **Supervised Learning:** An ML approach where the model learns from data that is already labeled with correct answers.
    - *Think:* Showing a student flashcards with "dog" on one side and a picture of a dog on the other.

4. **Unsupervised Learning:** An ML approach where the model learns from data that has no predefined labels, finding patterns and structures on its own.
    - *Think:* Giving a student a pile of mixed animal photos and asking them to sort them into groups that seem similar, without telling them what the groups are.

5. **Reinforcement Learning:** An ML approach where an agent learns by trial and error, receiving rewards or penalties for its actions to achieve a goal.
    - *Think:* Training a dog with treats for good behavior and "no" for unwanted behavior.

6. **Natural Language Processing (NLP):** A broader field of AI focused on enabling computers to understand, interpret, generate, and interact with human language (both text and speech).
    - *Think:* The science behind how computers can "read," "understand," and "write" like humans.

7. **Knowledge Cutoff:** The point in time beyond which a pre-trained AI model has not seen or learned new data. Its knowledge is essentially frozen at that date.
    - *Think:* The last newspaper the AI read; it doesn't know about events that happened after that date.

8. **Data Dependency (Model Limitation):** AI models, especially generative ones, are highly dependent on the data they were trained on. If the data is biased, limited, or outdated, the model's performance will reflect that.
    - *Think:* If you only teach a chef recipes with chicken, they won't know how to cook fish.

9. **Edge Cases:** Specific, unusual, or rare situations or inputs that an AI model might not have encountered frequently during training, potentially leading to unexpected or poor performance.
    - *Think:* An self-driving car encountering a kangaroo on the road in a city where kangaroos are not native – an unexpected scenario.

II. Generative AI Fundamentals

1. **Generative AI (Gen AI):** A type of AI that can create new, original content (like text, images, audio, or code) based on the patterns and structures it learned from existing data.
    - *Think:* An AI that's like an artist or writer, capable of producing something novel.

2. **Foundation Models:** Very large AI models trained on a vast amount of broad, unlabeled data, capable of performing a wide range of tasks. They serve as a "foundation" for more specialized AI applications.
    - *Think:* A highly educated generalist who can then be quickly trained for a specific job.

3. **Large Language Models (LLMs):** A type of foundation model specifically trained on massive amounts of text data to understand, generate, and manipulate human language.
    - *Think:* A super-powered version of autocorrect that can write entire essays or chat like a human.

4. **Multimodal Foundation Models:** Foundation models that can process and understand information from multiple types of data simultaneously (e.g., text, images, audio).
    - *Think:* An AI that can watch a video, listen to the audio, and read subtitles all at once to understand the content.

5. **Diffusion Models:** A type of generative model, often used for image generation, that works by gradually adding noise to an image and then learning to reverse the process to create new images from noise.
    - *Think:* Starting with a blurry, staticky image and gradually refining it into a clear picture.

6. **Transformer:** A specific neural network architecture (the underlying math and structure) that is particularly good at handling sequential data like text. It's a key innovation behind most modern LLMs.
    - *Think:* The powerful engine design that makes many modern AI models, especially LLMs, so effective.

7. **Embeddings:** Numerical representations (vectors) of words, sentences, or even images/audio in a way that captures their meaning and relationships. Similar items will have similar numerical representations.
    - *Think:* Giving every word a unique GPS coordinate in "meaning space," so words with similar meanings are close together.

8. **Token Count / Output Length:**
    - **Token:** The basic unit of data (like a word, part of a word, or punctuation) that an LLM processes.
    - **Output Length:** A parameter controlling how long or short the AI's generated response will be (often measured in tokens).

    - *Think:* Tokens are like the "words" the AI uses. Output length sets a limit on how many "words" it can use in its answer.

9. **Safety Settings/Filters:** Mechanisms built into generative AI models or platforms to prevent the generation of harmful, inappropriate, or biased content.
    - *Think:* Guardrails that help keep the AI's responses appropriate and safe.

III. Working with Gen AI Models & Data

1. **Prompt Engineering:** The art and science of crafting effective inputs (prompts) to guide a generative AI model to produce the desired output.
    - *Think:* Learning how to ask the AI the right question in the right way to get the best answer.

2. **Prompt Tuning:** A technique to adapt a pre-trained foundation model for specific tasks by learning a small set of "soft prompt" parameters, rather than retraining the entire model. More efficient than full fine-tuning.
    - *Think:* Giving the AI a special "lens" or "filter" to make it better at a specific task without re-teaching it everything.

3. **Fine-tuning:** Adapting a pre-trained foundation model to a specific task or dataset by further training it on a smaller, more specialized set of labeled data.
    - *Think:* Taking a generalist AI and sending it to a "specialty school" to become an expert in a particular field, like legal or medical text.

4. **Context Window:** The amount of text (input and recent conversation history) that an LLM can consider at one time when generating a response.
    - *Think:* The AI's short-term memory; how much of the current conversation or document it can "remember" to stay on topic.

5. **Retrieval Augmented Generation (RAG):** A technique that enhances LLM responses by first retrieving relevant information from an external knowledge base (like company documents) and then providing this information to the LLM as context to generate an answer.

    - *Think:* An AI assistant that first looks up facts in your company's database before answering your question, making its answers more accurate and up-to-date.

6. **Grounding:** The process of connecting an LLM's responses to reliable, factual information sources (like your enterprise data or the web) to improve accuracy and reduce "hallucinations" (made-up answers). RAG is a common grounding technique.
    - *Think:* Making sure the AI's answers are based on real facts, not just its general knowledge.

7. **Hallucinations:** When a generative AI model produces output that is factually incorrect, nonsensical, or not based on its training data, yet presents it confidently.
    - *Think:* The AI confidently making things up.

8. **Labeled Data:** Data where each piece of information has a tag or label that provides context or identifies what it is (e.g., an image labeled "cat," an email labeled "spam").
    - *Think:* Photos in an album where each photo has a caption explaining what it is.

9. **Unlabeled Data:** Raw data without any explicit tags or labels (e.g., a collection of news articles, a folder of random images).
    - *Think:* A pile of unsorted photos with no captions.

10. **Structured Data:** Data organized in a predefined format, typically in tables with rows and columns (e.g., spreadsheets, databases).
    - *Think:* A well-organized address book.

11. **Unstructured Data:** Data that doesn't have a predefined format or organization (e.g., text documents, images, videos, audio files).
    - *Think:* A collection of handwritten notes, photos, and voice memos.

12. **Data Quality:** The measure of how fit data is for its intended purpose, considering factors like accuracy, completeness, consistency, relevance, and timeliness.
    - *Think:* Ensuring the ingredients you use for cooking are fresh, correct, and complete for a good meal.

13. **ML Lifecycle:** The end-to-end process of developing and deploying a machine learning model. Key stages include: Data Ingestion, Data Preparation, Model Training, Model Deployment, and Model Management (monitoring).
    - *Think:* The complete "recipe" for building and maintaining an AI model, from gathering ingredients to serving the dish and checking if people like it.

14. **Zero-shot Prompting:** Asking an LLM to perform a task it hasn't been explicitly trained for, without giving it any examples of how to do it in the prompt.
    - *Think:* Asking a smart person a question on a topic they know generally about, but haven't studied for this specific question, and they still give a decent answer.

15. **One-shot / Few-shot Prompting:** Providing the LLM with one (one-shot) or a few (few-shot) examples of the desired task and output format within the prompt itself, to guide its response.
    - *Think:* Showing the AI one or two examples of what you want (e.g., "Translate this English sentence to French: [example]. Now translate this: ...").

16. **Role Prompting:** Instructing the LLM to adopt a specific persona or role (e.g., "You are a helpful marketing expert") to influence the style, tone, and content of its responses.
    - *Think:* Telling the AI to "act as if" it's a specific character or expert.

17. **Human in the Loop (HITL):** A system where humans are involved in the AI's process, typically to review, correct, or provide feedback on the AI's outputs, especially in critical or ambiguous situations.
    - *Think:* Having a human editor check the AI's writing before it's published.

18. **Data Anonymization:** The process of removing or altering personally identifiable information (PII) from data so that individuals cannot be identified.
    - *Think:* Taking names and addresses out of a dataset to protect privacy.

19. **Data Pseudonymization:** Replacing identifiable data with artificial identifiers, or pseudonyms. It's a step towards anonymization but might allow re-identification under certain conditions (e.g., with a separate key).
    - *Think:* Giving everyone in a dataset a secret code name instead of their real name.

20. **Prompt Chaining:** A technique where the output of one prompt to an LLM becomes part of the input for a subsequent prompt, allowing for more complex, multi-step tasks.
    - *Think:* Asking the AI a series of connected questions, where each answer helps you frame the next question to dig deeper or build something complex.

21. **Model Versioning:** The practice of tracking and managing different iterations or versions of an AI model as it's updated, retrained, or fine-tuned. Important for reproducibility and rollback.
    - *Think:* Like software versions (e.g., App v1.0, App v1.1), this is keeping track of different updates to your AI model so you know which one is which and can go back to an older one if needed.

IV. Gen AI Outputs & Capabilities

1. **Text Generation:** Creating human-like text (e.g., articles, summaries, emails, poetry).
2. **Image Generation:** Creating new images from text descriptions or by modifying existing images.
3. **Video Generation:** Creating new video sequences, often from text prompts or by animating images.
4. **Speech Generation (Text-to-Speech):** Converting written text into audible human-like speech.
5. **Code Generation:** Automatically writing software code in various programming languages based on natural language descriptions or existing code snippets.
6. **Text Understanding (Natural Language Processing - NLP):** AI's ability to interpret, analyze, and make sense of human language. (Also listed in Core AI concepts, reinforcing its dual nature).
7. **Image Understanding (Computer Vision):** AI's ability to interpret and understand information from images or videos.
8. **Speech Understanding (Speech-to-Text):** Converting spoken language into written text.
9. **Personalized User Experience:** Using Gen AI to tailor content, recommendations, interactions, and services to individual user preferences, behavior, and needs.

- *Think:* A news app that shows you stories it knows you'll be interested in, or an e-commerce site that recommends products just for you.

10. **Data Analysis (as a Gen AI use case):** Leveraging Gen AI, especially LLMs, to interpret, summarize, ask questions of, and generate insights from both structured and unstructured data using natural language.
    - *Think:* Using AI to "talk" to your data, asking it questions in plain English to find trends or get summaries, even from complex spreadsheets or long documents.

V. Controlling Gen AI Model Behavior

1. **Temperature:** A parameter that controls the randomness or creativity of an LLM's output. Higher temperature = more creative/random; Lower temperature = more focused/deterministic.
    - *Think:* A creativity dial. Low for factual answers, high for brainstorming.

2. **Top-P (Nucleus Sampling):** A parameter that controls the range of possible next words an LLM considers. It selects the most probable words whose cumulative probability adds up to 'P'.
    - *Think:* Instead of picking from all possible next words, the AI picks from a smaller, more likely set, making outputs less random than high temperature alone.

3. **Function Calling:** Enabling an LLM to interact with external tools or APIs by formatting its output in a way that can trigger a specific function (e.g., fetch weather data, book a flight).
    - *Think:* Giving the AI the ability to use "tools" to get real-world information or perform actions.

4. **Agents (AI Agents):** AI systems that can perceive their environment, make decisions, and take actions to achieve specific goals, often using LLMs as their "brain" and tools (like function calling) to interact.
    - *Think:* An AI assistant that can not only chat but also actively do things for you, like scheduling meetings or researching topics using external tools.

5. **Chain-of-Thought Prompting:** A prompt engineering technique where you ask the LLM to explain its reasoning step-by-step before giving the final answer, often leading to more accurate results for complex problems.
    - *Think:* Asking the AI to "show its work" like in a math problem.

6. **ReAct Prompting (Reasoning and Acting):** An advanced prompting framework where the LLM can generate both reasoning traces and task-specific actions, allowing it to interact with external tools to gather information and solve complex tasks.
    - *Think:* An AI that can "think out loud" about what it needs to do, then "use a tool" (like a search engine), then "think again" based on the new info.

VI. Google Cloud AI Platforms & Tools

1. **Vertex AI:** Google Cloud's unified machine learning platform to build, deploy, and manage ML models, including generative AI models.
    - *Think:* The central workshop for all your AI projects on Google Cloud.

2. **Model Garden (in Vertex AI):** A repository in Vertex AI offering access to a wide variety of foundation models (Google's, open-source, third-party) that can be easily used and customized.
    - *Think:* A library or store where you can find and select pre-built AI models.

3. **Vertex AI Search (formerly Enterprise Search):** A Google Cloud service that allows businesses to create powerful search experiences across their own websites and enterprise data, often enhanced by generative AI for summarization and question answering.
    - *Think:* A super-smart search engine for your company's internal documents and data.

4. **Vertex AI Agent Builder:** A Google Cloud toolset to easily build and deploy generative AI-powered conversational agents (chatbots, voice bots) and specialized search applications.
    - *Think:* A toolkit for creating smart AI assistants for customer service or internal helpdesks.

5. **Google AI Studio:** A web-based tool for quickly prototyping and experimenting with Google's latest generative AI models (like Gemini) using prompts.
    - *Think:* A playground to test out Google's newest AI models with your own ideas.

6. **Gemini:** Google's most capable and general multimodal AI model, designed to understand and operate across different types of information like text, code, images, and video.
    - *Think:* Google's flagship, highly versatile AI model that can handle many different tasks and data types.

7. **Gemma:** A family of lightweight, state-of-the-art open models from Google, built from the same research and technology used to create Gemini models.
    - *Think:* Smaller, open-source versions of Google's powerful AI technology that developers can more easily run and customize.

8. **Imagen:** Google's family of text-to-image diffusion models, capable of generating high-quality, photorealistic images from natural language prompts.
    - *Think:* Google's AI artist that can create pictures based on your descriptions.

9. **Veo:** Google's most capable video generation model, able to create high-definition video clips from text, image, or video prompts.
    - *Think:* Google's AI film director that can create short videos based on your ideas.

10. **TPUs (Tensor Processing Units):** Google's custom-designed hardware accelerators optimized for machine learning workloads, especially for training and running large AI models.
    - *Think:* Specialized computer chips that make AI tasks run much faster and more efficiently.

11. **Google's AI-first approach:** Google's strategy of deeply integrating AI into all its products and services, and making AI a core part of its innovation and development philosophy.
    - *Think:* Google putting AI at the heart of everything it builds and does.

12. **Low-code / No-code Tools (for AI):** Platforms and tools that allow users with minimal or no programming skills to build and deploy AI applications, often using visual interfaces and pre-built components.
    - *Think:* Building an AI app by dragging and dropping blocks, instead of writing complex code.

13. **API (Application Programming Interface):** A set of rules and protocols that allows different software applications to communicate and exchange data with each other. Many Gen AI capabilities are accessed via APIs.
    - *Think:* A menu in a restaurant. You (one software) use the menu (API) to tell the kitchen (another software) what you want.

14. **Gems (for Gemini Advanced):** Custom, specialized versions of Gemini that users can create for specific tasks or knowledge domains within Gemini Advanced.
    - *Think:* Creating your own personalized "expert" versions of Gemini that are tailored to your specific needs or company data.

15. **Gemini for Google Workspace:** Integration of Gemini's AI capabilities into Google Workspace apps like Docs, Sheets, Gmail, and Meet to help users write, summarize, organize, and create.
    - *Think:* Having a smart AI assistant directly inside your everyday Google apps.

16. **Cloud NotebookLM API (part of Agentspace):** An API that allows developers to build applications using models grounded in user-specified documents, similar to Google's NotebookLM product.
    - *Think:* A tool for developers to create AI that "reads" and uses specific documents you provide as its main source of knowledge.

17. **Google Agentspace:** A concept or platform focused on enabling the development and orchestration of AI agents that can perform tasks.
    - *Think:* A workshop or environment where developers can build and manage smart AI assistants (agents).

18. **Customer Engagement Suite (Google Cloud):** A collection of Google Cloud AI solutions designed to enhance customer interactions, including tools for conversational AI, agent assistance, and customer insights.
    - *Think:* A toolkit of AI services to help businesses provide better customer service.

19. **Conversational Agents (e.g., in Customer Engagement Suite):** AI-powered systems (chatbots, voice bots) that can engage in natural conversations with users to provide information, answer questions, or complete tasks.
    - *Think:* The automated chat assistant on a website or the voice assistant you talk to on the phone.

20. **Agent Assist (e.g., in Customer Engagement Suite):** AI tools that provide real-time support and information to human customer service agents during their interactions with customers.
    - *Think:* An AI "whispering in the ear" of a human customer service agent, giving them helpful suggestions and information.

21. **AutoML (in Vertex AI):** A suite of Google Cloud tools that automates parts of the machine learning model building process, enabling users with limited ML expertise to train high-quality custom models.
    - *Think:* An "AI that helps you build other AI" by handling some of the complex technical steps.

22. **Agent Tooling (Extensions, Functions, Data Stores, Plugins):** The various components and integrations that AI agents use to interact with the external world, access information, or perform actions beyond the LLM's inherent capabilities.
    - *Think:* The "hands and senses" of an AI agent, allowing it to use external software, databases, or services.

23. **Vertex AI Feature Store:** A centralized repository in Vertex AI for storing, serving, managing, and sharing machine learning features (the input variables used by models) to streamline ML workflows.
    - *Think:* A well-organized pantry for all the specific data ingredients your AI models need.

24. **Hypercomputer (Google Cloud):** Google's AI supercomputing architecture, combining optimized hardware (like TPUs and GPUs), software, and networking to efficiently train and serve very large AI models.
    - *Think:* Google's massive, purpose-built supercomputers designed specifically for running huge AI tasks.

25. **Conversational Insights (Google Cloud):** AI-driven analytics that extract meaningful information and trends from customer conversations (e.g., from contact centers) to improve customer experience and agent performance.
    - *Think:* AI listening to (or reading) customer service calls/chats and highlighting common problems, customer sentiment, or how well agents are performing.

26. **Google Cloud Contact Center AI (CCaaS):** Google Cloud's solution for building AI-powered contact centers, integrating conversational AI, agent assist, and analytics. (Often referred to as CCaaS - Contact Center as a Service).
    - *Think:* Google's complete toolkit for creating a modern, smart call center that uses AI to help customers and support agents.

27. **Document AI (Google Cloud API):** A Google Cloud service that uses AI to understand, extract data from, and process documents (e.g., invoices, forms, contracts).
    - *Think:* An AI that can read your scanned documents or PDFs, pull out important information (like names, dates, amounts from an invoice), and organize it for you.

28. **Vision AI (Google Cloud API):** A Google Cloud service that enables applications to understand the content of images using pre-trained models (e.g., detect objects, faces, read text in images).
    - *Think:* An AI that can "see" and tell you what's in a picture, like identifying logos, reading signs, or describing a scene.

29. **Natural Language API (Google Cloud):** A Google Cloud service that provides pre-trained models for understanding text, including sentiment analysis, entity recognition, syntax analysis, and content classification.
    - *Think:* An AI that can read text and tell you what it's about, who or what is mentioned, if the tone is positive or negative, and categorize the content.

30. **Google Cloud's Open Approach (to AI):** Google's philosophy and strategy of supporting and contributing to open-source AI models, tools, and standards, alongside offering its own proprietary solutions, to provide customers with choice, flexibility, and foster innovation.
    - *Think:* Google not only building its own great AI but also supporting the wider AI community, allowing businesses to mix and match the best tools for their needs.

VII. Responsible AI & Strategy

1. **Responsible AI:** Developing and using AI systems in a way that is ethical, fair, transparent, accountable, and beneficial to society, while minimizing risks.
    - *Think:* Making sure AI is used for good, in a safe and fair way.

2. **Bias (in AI):** When an AI system produces results that are systematically prejudiced due to flawed assumptions in the ML process or biased training data.
    - *Think:* If an AI is only trained on pictures of white cats, it might struggle to recognize or might mislabel black cats.

3. **Fairness (in AI):** Ensuring that AI systems do not produce discriminatory or unjust outcomes for different groups of people.
    - *Think:* Making sure the AI treats everyone equally and doesn't make biased decisions.

4. **Transparency (in AI):** The ability to understand how an AI model arrives at its decisions or predictions.
    - *Think:* Being able to see "inside the AI's brain" to understand why it made a particular choice.

5. **Explainability (in AI):** Being able to describe in human-understandable terms how an AI model works and why it made a specific decision. Closely related to transparency.
    - *Think:* The AI being able to explain its reasoning in simple terms.

6. **Secure AI Framework (SAIF):** Google's framework for building secure AI systems, covering aspects from secure infrastructure to protecting models and data.

- *Think:* A comprehensive security plan specifically designed for AI systems.

7. **Accountability (in AI):** Establishing clear responsibility for the development, deployment, and impact of AI systems, including mechanisms for redress if things go wrong.
    - *Think:* Knowing who is responsible if an AI system makes a mistake or causes harm.

8. **Governance (Data/AI):** The framework of rules, practices, and processes used to manage and control an organization's data and AI initiatives, ensuring they align with business objectives, ethics, and regulations.
    - *Think:* The "rulebook" and "management system" for how a company uses data and AI responsibly and effectively.

9. **Scalability (of AI Platforms):** The ability of an AI platform or solution to handle increasing amounts of data, users, or complexity without a proportional drop in performance.
    - *Think:* An AI system that can grow with your business, handling more customers or data smoothly.

10. **Reliability (of AI Platforms):** The consistency and dependability of an AI platform or solution to perform as expected without failures.
    - *Think:* An AI system you can count on to work correctly whenever you need it.

11. **Identity and Access Management (IAM):** A security framework that ensures the right individuals have the appropriate access to technology resources (like AI models, data, and platforms).
    - *Think:* The "bouncer" and "key master" for your AI systems, controlling who gets in and what they can do.

12. **Security Command Center (Google Cloud):** A centralized security and risk management platform in Google Cloud that helps organizations prevent, detect, and respond to threats.

- *Think:* The central security dashboard for your Google Cloud environment, including AI assets.

13. **Drift Monitoring (Model Drift):** The process of tracking how an AI model's performance changes over time as the real-world data it encounters deviates from the data it was trained on.
    - *Think:* Noticing that your AI model, which was great at predicting fashion trends last year, is now out of touch because styles have changed.

14. **Key Performance Indicators (KPIs) for AI:** Specific, measurable metrics used to track the performance, effectiveness, and business value of AI initiatives.
    * *Think:* The specific goals and targets you set to see if your AI projects are successful (e.g., "reduce customer support calls by 20%").