AI TERMINOLOGY

EXECUTIVE BRIEF

A Comprehensive Guide for Business Leaders

Understanding the Language of Artificial Intelligence

5	50+	15+	4
SECTIONS	KEY TERMS	TABLES	INTERFACES

December 2024

EXECUTIVE SUMMARY

The AI revolution has introduced a new vocabulary that executives must master to make informed strategic decisions. This brief provides a structured overview of essential AI terminology, organized by practical application and business context.

Understanding these terms is not merely academic—it directly impacts **vendor negotiations**, **product evaluation**, **team communication**, and **strategic planning**. An executive who understands the difference between a hallucination and a context window limitation can make better decisions about AI deployment and risk management.

What This Brief Covers

- **AI Interface Categories:** IDE/Terminal, Web, Code, and App—with capability comparisons
- **Core Technical Concepts:** LLMs, tokens, context windows, and model architecture
- **Prompts & Prompt Engineering:** The art and science of AI communication
- **Hallucinations & Limitations:** Understanding when and why AI fails
- **Industry Terminology:** The slang, acronyms, and jargon of the AI ecosystem

KEY TAKEAWAY

AI interfaces are not created equal. Terminal/IDE access provides significantly more capability than consumer apps. Understanding this hierarchy enables better tool selection and more realistic expectations.

1 AI INTERFACE CATEGORIES

Not all AI access points are created equal. The same underlying model can behave very differently depending on how you interact with it. This section breaks down the four primary interface categories.

Interface Comparison Matrix

CATEGORY	IDE/TERMINAL	WEB	CODE/API	MOBILE
Access Level	Full system	Browser	Programmatic	Mobile-optimized
File Operations	Full read/write	Upload only	Full automation	Limited
Automation	Scripts & pipelines	Manual only	Full integration	None
Context Size	Maximum	Standard	Configurable	Reduced
Use Case	Dev & automation	General queries	Production	Quick questions
Technical Skill	Advanced	Beginner	Developer	None required

1.1 IDE/Terminal (Command Line Interface)

Definition: The Integrated Development Environment (IDE) or terminal provides the most powerful interface to AI systems. This is the command-line access available on all PCs (Terminal on Mac, Command Prompt/PowerShell on Windows, Bash on Linux).

Why It Matters: Terminal access enables capabilities that consumer interfaces cannot match. Think of it as the difference between using a calculator app versus having full access to Excel with macros—same basic math, vastly different capabilities.

Key Capabilities:

- Execute code directly on your system
- Read and write files without manual upload/download
- Chain multiple operations together (pipelines)
- Automate repetitive tasks with scripts
- Access larger context windows for complex tasks

Executive Insight

Organizations that only use consumer interfaces are leaving 60-80% of AI capability on the table. Always ask about terminal/API access.

Terminal AI Tools

TOOL	PROVIDER	PRIMARY USE
Claude Code	Anthropic	Agentic coding, file operations
GitHub Copilot CLI	GitHub/Microsoft	Code generation in terminal
Cursor	Cursor Inc.	AI-native IDE
Aider	Open Source	Git-integrated coding

1.2 Web Interface

Definition: Browser-based access to AI systems, such as ChatGPT, Claude.ai, or Google's Gemini. This is the most common way executives interact with AI today.

Strengths: Immediate access, no installation required, intuitive interface, good for quick queries.

Limitations: Manual file handling, no automation capability, session-based context, cannot integrate with other systems natively.

1.3 Code/API Access

Definition: Application Programming Interface (API) access allows software to communicate with AI systems programmatically. This is how AI gets embedded into products and workflows—where AI transitions from "tool" to "infrastructure."

1.4 Mobile App

Reality Check: Mobile apps typically have the most limited capabilities—smaller context windows, no file system access, and optimized for brief interactions rather than complex work. Best for voice queries, quick fact-checking, and idea capture while traveling.

2 CORE TECHNICAL CONCEPTS

These foundational concepts are essential for understanding AI capabilities and limitations. Mastering this vocabulary enables more productive conversations with technical teams and vendors.

2.1 Large Language Models (LLMs)

What It Is: An LLM is a type of AI trained on massive amounts of text data to understand and generate human language. Think of it as a sophisticated pattern-matching system that predicts what words should come next based on context.

Key Players

MODEL	PROVIDER	STRENGTHS	BEST FOR
GPT-4 / GPT-4o	OpenAI	Versatility, vision, reasoning	General purpose
Claude 3.5/4	Anthropic	Long context, safety, coding	Documents, coding
Gemini 1.5/2.0	Google	Multimodal, massive context	Research
LLaMA 3	Meta (Open)	Self-hosting, customization	Private deployment
Mistral	Mistral AI	Efficiency, EU hosting	Cost-conscious

2.2 Tokens

Definition: Tokens are the basic units that AI models process. A token is roughly 4 characters or about 0.75 words in English. The sentence "Hello, how are you?" is approximately 5-6 tokens.

Why It Matters: AI pricing is typically based on tokens processed. Understanding token counts helps estimate costs and manage context window limits.

CONTENT TYPE	APPROXIMATE TOKENS
1 word	1.3 tokens
1 sentence	15-25 tokens
1 paragraph	75-150 tokens

1 page (500 words)	650-700 tokens
10-page report	6,500-7,000 tokens

Cost Implications

MODEL	INPUT (per 1M tokens)	OUTPUT (per 1M tokens)
GPT-4o	$2.50	$10.00
GPT-4o-mini	$0.15	$0.60
Claude 3.5 Sonnet	$3.00	$15.00
Claude 3 Haiku	$0.25	$1.25
Gemini 1.5 Pro	$1.25	$5.00

2.3 Context Window

Definition: The context window is the maximum amount of text (measured in tokens) that an AI can "see" and process at once. Think of it as the AI's working memory. Everything the AI needs—your question, documents, conversation history, and space for response—must fit.

MODEL	CONTEXT WINDOW	REAL-WORLD EQUIVALENT
GPT-3.5	4,096 tokens	~6 pages
GPT-4 Turbo	128,000 tokens	~200 pages
Claude 3.5 Sonnet	200,000 tokens	~300 pages
Gemini 1.5 Pro	1,000,000 tokens	~1,500 pages
Gemini 1.5 Pro (ext)	2,000,000 tokens	~3,000 pages

■ Business Implication

Larger context windows enable analysis of entire documents, codebases, or conversation histories. Claude and Gemini are preferred for document-heavy work.

2.4 Temperature

Definition: Temperature is a setting (0.0 to 2.0) that controls how "creative" or "random" the AI's responses are. Lower temperature = more predictable; higher temperature = more varied.

TEMPERATURE	BEHAVIOR	BEST FOR
0.0 - 0.3	Deterministic, consistent	Code, factual queries, data extraction
0.4 - 0.7	Balanced	General conversation, analysis
0.8 - 1.2	Creative, varied	Brainstorming, creative writing
1.3 - 2.0	Highly random	Experimental, unusual ideas

2.5 Fine-Tuning vs. RAG

Two methods exist to customize AI behavior for specific use cases:

FACTOR	FINE-TUNING	RAG
Method	Permanently adjusts model weights	Dynamically retrieves info at query time
Cost	High upfront ($10K-$100K+)	Lower, ongoing
Flexibility	Fixed after training	Dynamic, updateable
Data Required	Thousands of examples	Any document corpus
Best For	Style/behavior changes	Knowledge retrieval
Time to Deploy	Weeks	Days

3 PROMPTS & PROMPT ENGINEERING

The quality of AI output depends enormously on how you communicate with it. Prompt engineering is the discipline of crafting effective instructions for AI systems.

3.1 What Is a Prompt?

Definition: A prompt is any input you provide to an AI system—questions, instructions, examples, or context. Everything you type (or that's programmatically sent) is part of the prompt.

COMPONENT	DESCRIPTION	EXAMPLE
System Prompt	Background instructions (often hidden)	"You are a helpful financial analyst..."
User Prompt	Your actual question/request	"Analyze Q3 revenue trends"
Context	Documents, data, history	[Uploaded spreadsheet]
Examples	Sample inputs/outputs (few-shot)	"Here's how I want it formatted..."

3.2 The CRISP Framework for Executives

LETTER	ELEMENT	EXAMPLE
C	Context	"I'm preparing for a board meeting on AI strategy..."
R	Role	"Act as a management consultant with M&A expertise..."
I	Instructions	"Create a 5-point executive summary..."
S	Specifics	"Focus on risks, costs, and timeline..."
P	Parameters	"Keep it under 500 words, use bullet points..."

3.3 Core Principles

PRINCIPLE	BAD EXAMPLE	GOOD EXAMPLE
Be Specific	"Analyze this data"	"Calculate YoY growth rate for each product category"

Provide Context	"Write a memo"	"Write a memo to the CFO about Q3 budget overruns"
Define Format	"Summarize this"	"Summarize in 5 bullet points, max 20 words each"
Use Examples	"Format it nicely"	"Format like this example: [provide sample]"
Chain of Thought	"What's the answer?"	"Think step by step, then provide your answer"

3.4 Common Prompt Types

TYPE	USE CASE	EXAMPLE START
Analysis	Data interpretation	"Analyze [X] and identify..."
Synthesis	Combining information	"Compare [A] and [B], then..."
Generation	Creating content	"Write a [format] about [topic]..."
Transformation	Reformatting	"Convert this [format] to [new format]..."
Extraction	Pulling specific info	"Extract all [items] from..."
Evaluation	Assessment	"Evaluate [X] against criteria [Y]..."

4 HALLUCINATIONS & LIMITATIONS

Understanding AI failures is as important as understanding capabilities. This section covers what can go wrong and how to mitigate risks.

4.1 What Are Hallucinations?

Definition: Hallucinations occur when an AI generates information that sounds plausible but is factually incorrect, fabricated, or nonsensical. The AI presents false information with the same confidence as accurate information.

Why They Happen:

- LLMs predict probable next words, not truth
- Training data contains errors and contradictions
- Models have no real-world verification mechanism
- Pressure to provide an answer vs. admitting uncertainty

Hallucination Types

TYPE	DESCRIPTION	EXAMPLE
Factual	Wrong facts	"The Eiffel Tower was built in 1920"
Citation	Fake sources	"According to Smith et al. (2019)..." [doesn't exist]
Numerical	Wrong numbers	Incorrect statistics, dates, prices
Logical	Flawed reasoning	Valid-sounding but incorrect conclusions
Entity	Wrong attributions	Assigning quotes/actions to wrong people

4.2 High-Risk Scenarios

DOMAIN	RISK LEVEL	MITIGATION REQUIRED
Legal citations	■ VERY HIGH	Always verify case law
Medical advice	■ VERY HIGH	Professional verification mandatory

Financial data	■ HIGH	Cross-reference original sources
Historical facts	■ MEDIUM-HIGH	Verify key claims
Technical specs	■ MEDIUM	Check documentation
Creative writing	■ LOW	Accuracy often not required

4.3 Mitigation Strategies

For Individuals:

- Verify critical claims independently
- Ask for sources, then check if they exist
- Cross-reference against known reliable sources
- Use retrieval tools (web search, RAG)

For Organizations:

- Implement human review for critical workflows
- Use RAG systems with verified documents
- Set temperature low for factual tasks
- Create verification checklists

■■ **CRITICAL WARNING:** Never use AI output for high-stakes decisions without human verification. This includes: legal documents, medical diagnoses, financial advice, safety-critical systems.

5 INDUSTRY TERMINOLOGY & SLANG

The AI industry has developed its own vocabulary. This glossary covers terms you'll encounter in vendor pitches, technical discussions, and industry publications.

5.1 Model & Architecture Terms

TERM	DEFINITION	WHY IT MATTERS
Transformer	Neural network architecture behind LLMs	Foundation of all major AI
Attention	How models weigh input importance	Enables context understanding
Inference	Running a trained model for outputs	What you pay for in API calls
Weights	Learned parameters in a model	What makes each model unique
Embeddings	Numerical representations of text	Enables semantic search
Vector DB	Storage for embeddings	Powers RAG and search

5.2 Training & Alignment Terms

TERM	DEFINITION	BUSINESS RELEVANCE
Pre-training	Initial training on massive text	Creates base capabilities
RLHF	Reinforcement Learning from Human Feedback	Makes models helpful/safe
Alignment	Ensuring AI behaves as intended	Critical for enterprise safety
Jailbreaking	Bypassing AI safety restrictions	Security concern
Red Teaming	Adversarial vulnerability testing	Required for responsible deployment
Synthetic Data	AI-generated training data	Reduces need for real data

5.3 Common Slang & Abbreviations

TERM	MEANING	CONTEXT
AGI	Artificial General Intelligence (human-level)	Long-term AI discussions
Agentic AI	AI that takes actions autonomously	Automation, workflows
CoT	Chain of Thought (reasoning technique)	Prompt engineering
Few-shot	Providing examples in the prompt	Improving output quality
Zero-shot	No examples provided	Testing base capabilities
Grounding	Connecting AI to real data sources	Reducing hallucinations
Multimodal	Handles text, images, audio, video	Model capabilities
SLM	Small Language Model (efficient)	Cost-conscious deployments
Wrapper	Product built on top of another AI	Vendor due diligence

5.4 Emerging & Trending Terms

TERM	DEFINITION	WHY IT'S HOT
MCP	Model Context Protocol - AI-to-tool connections	Enables AI to use external tools
Function Calling	AI's ability to trigger external functions	Automation, integrations
Mixture of Experts	Architecture using specialized sub-models	Efficiency at scale
Constitutional AI	Training AI with explicit principles	Safety and alignment
Inference-Time Compute	More thinking during response	Better reasoning (o1, Claude)
Distillation	Training smaller models from larger	Cost reduction, edge deployment

QUICK REFERENCE CARD

Interface Hierarchy (Most → Least Capable)

RANK	INTERFACE	CAPABILITY	BEST FOR
1	IDE/Terminal	■■■■■	Automation, development, complex tasks
2	Code/API	■■■■■	Production systems, integrations
3	Web Interface	■■■	General use, document review
4	Mobile App	■■	Quick questions, on-the-go

Decision Framework: When to Trust AI Output

RISK LEVEL	EXAMPLE TASKS	VERIFICATION REQUIRED
■ **Low Risk**	Brainstorming, first drafts, summaries	Light review
■ **Medium Risk**	Analysis, recommendations, research	Fact-check key claims
■ **High Risk**	Financial data, technical specs, citations	Full verification required
■ **Critical Risk**	Legal, medical, safety-critical	Professional review mandatory

Cost Estimation Quick Guide

TASK	EST. TOKENS	GPT-4o COST	CLAUDE COST
Simple question	500	~$0.005	~$0.008
Document summary (10 pg)	8,000	~$0.08	~$0.12
Full report analysis	50,000	~$0.50	~$0.75
Large codebase review	200,000	~$2.00	~$3.00

FINAL TAKEAWAYS FOR EXECUTIVES

1) Interface matters more than model choice. 2) Never trust AI for high-stakes without verification. 3) Prompt engineering is learnable. 4) Context window determines what's possible. 5) Master the vocabulary.