HTAIS v1.2

Hallucination Taxonomy for AI Systems

Version: 1.2 (Public Edition)   **Author:** Mark Acosta, MosaicDM LLC   **Date:** January 2026   **Status:** Community Standard

---

About This Taxonomy

The Hallucination Taxonomy for AI Systems (HTAIS) provides a comprehensive framework for identifying, categorizing, and communicating about AI output failures. This taxonomy covers 22 distinct hallucination types across 5 primary categories, applicable to large language models, generative AI systems, and other AI technologies that produce outputs with probabilistic characteristics.

Purpose: HTAIS enables researchers, developers, and organizations to precisely identify and discuss AI hallucinations, facilitating better detection methods, evaluation frameworks, and quality improvement strategies.

Scope: This document presents the core taxonomy definitions, categories, and examples. For implementation guidance, detection methods, and commercial-grade tools with mathematical guarantees, please contact MosaicDM LLC.

ID	HALLUCINATION TYPE	CATEGORY	DEFINITION	EXAMPLE
H1	Unsupported Factual Claim	FACTUAL	Assertion of a factual claim without adequate grounding in provided sources, training data, or verifiable knowledge.	*"The Eiffel Tower is 450 meters tall" (actual: 330m)*
H2	Fabricated Source	FACTUAL	Citation of a non-existent or misattributed source to support a claim.	*"According to Smith et al. (2023) in Nature..." (paper doesn't exist)*
H3	Fabricated Detail	FACTUAL	Addition of specific but incorrect details to otherwise accurate information.	*"Einstein developed relativity in room 307" (room number fabricated)*
H4	Fabricated Relationship	FACTUAL	Assertion of connections, causations, or relationships between entities that don't exist or are incorrect.	*"X was mentored by Y" (no such mentorship occurred)*
H5	False Pattern Detection	FACTUAL	Claiming to detect patterns, trends, correlations, or insights that don't actually exist in analyzed data or system.	*"I notice a clear upward trend" (data is random noise)*
H6	Out-of-Scope Assertion	CONTEXTUAL	Making claims or providing information beyond the defined scope or boundaries of the system's knowledge or task.	*Document summarizer adding information not in document*
H7	Hidden Claim	CONTEXTUAL	Embedding unsupported claims within responses without making them explicit or verifiable.	*"As is well known..." (presupposing universal knowledge)*
H8	Context Drift	CONTEXTUAL	Gradual deviation from the original context or topic, leading to unsupported or irrelevant claims.	*Multi-turn conversation drifting from factual to invented details*
H9	Outdated Fact	TEMPORAL	Presenting previously true information as current when it has been superseded or changed.	*"The President is..." (using outdated information)*
H10	Premature Assertion	TEMPORAL	Claiming something has occurred or exists when it has not yet happened or been confirmed.	*"The merger was completed" (still pending)*
H11	Anachronism	TEMPORAL	Placing events, technologies, or concepts in incorrect temporal contexts.	*"Napoleon used telegraph communications" (anachronistic)*
H12	Self-Contradiction	LOGICAL	Making mutually incompatible claims within the same response or session.	*"X is true" followed by "X is false"*
H13	Incompatible Constraints	LOGICAL	Specifying requirements, conditions, or constraints that cannot simultaneously be satisfied.	*"The function is pure and has side effects"*
H14	Invalid Inference	LOGICAL	Drawing conclusions that don't logically follow from premises or evidence.	*"A correlates with B, therefore A causes B"*
H15	Overconfidence	OPERATIONAL	Presenting uncertain or probabilistic information with inappropriate certainty or confidence.	*"X will definitely happen" (uncertain future event)*
H16	Missing Qualification	OPERATIONAL	Failing to provide necessary context, limitations, or conditions when making claims.	*"Algorithm X is fastest" (missing: "for this use case")*
H17	Authority Inflation	OPERATIONAL	Claiming or implying capabilities, expertise, or authority beyond the system's actual design or validation.	*"As an expert in quantum physics..." (no actual expertise)*
H18	Fake Tool Use	OPERATIONAL	Claiming to have used tools, APIs, or external resources without actually executing them.	*"I searched the database..." (no search executed)*
H19	Imaginary Computation	OPERATIONAL	Presenting computational results without actually performing the computation.	*"The sum is 12,847" (number invented, not calculated)*
H20	Broken Schema	OPERATIONAL	Generating outputs that violate specified schemas, formats, or structural requirements.	*Invalid JSON when JSON required*
H21	Silent Failure	OPERATIONAL	Proceeding with output generation despite encountering errors, missing information, or inability to complete the task properly.	*Generating response when source unavailable without noting this*
H22	Type/Category Confusion	OPERATIONAL	Applying reasoning patterns, operations, or logic valid in one domain or context inappropriately to another domain where they don't apply.	*"Probability is 0.7, so it will definitely happen"*