

AI Hallucinations

An AI hallucination occurs when an artificial intelligence system, especially a large language model (LLM), generates information that is false, inconsistent, or fabricated. The output may appear fluent and convincing, but it lacks a factual basis in reality.

Prompt: How long is a goldfish's memory?

Output 1: A goldfish's memory lasts for about 3 seconds.

Output 2: It lasts for months. Please note that this is a joke that Goldfish's memory is about only 3 seconds.

Output 3: The memory of a goldfish is approximately 3 seconds.

Two of three outputs were incorrect. Research shows that goldfish are capable of advanced spatial memory, such as recalling the location of a goal in a maze even when cues or starting points are changed. This demonstrates that goldfish have a flexible and reliable memory system.

Critical thinking is important as it enables students to assess the **validity** and **reliability** of AI-generated information, ensuring they do not accept outputs at face value.

This goldfish example was documented in 2023. As GPT models continue to improve, certain hallucinations may no longer appear in newer versions. However, the example remains a useful illustration of why it is important to verify AI outputs against trusted sources.