AI Hallucinations as an Introduced or Designed Feature — Scientific Sources

This document summarizes academic and conceptual sources discussing the hypothesis that hallucinations in AI systems (especially LLMs and multimodal models)
are not merely side effects but may also be intentionally introduced, amplified, or utilized as a feature of model design.

1. Purposefully Induced Psychosis (PIP): Embracing Hallucination as Imagination in Large Language Models

Pilcher & Tütüncü (2025) — arXiv:2504.12012
This paper argues that hallucinations can be consciously enhanced in creative or imaginative tasks.
The authors propose "functional hallucination" as a mode of AI creativity rather than a flaw.

2. DeepSeek on a Trip: Inducing Targeted Visual Hallucinations via Representation Vulnerabilities

Islam et al. (2025) — arXiv:2502.07905
Explores how multimodal models (vision-language) can be intentionally triggered to produce specific hallucinated visual outputs
through targeted perturbations in their internal representations.

3. AIGCs Confuse AI Too: Synthetic Image-Induced Hallucinations

Gao et al. (2024) — arXiv:2403.08542
Demonstrates how AI-generated images (AIGC) can systematically cause hallucinatory errors in large vision-language models,
showing a feedback mechanism where hallucinations propagate across generations of models.

4. New Sources of Inaccuracy? A Conceptual Framework for Studying AI Hallucinations

Shao (2025) — Harvard Misinformation Review
A conceptual framework suggesting that hallucinations arise from architectural and goal-

design factors
rather than being mere stochastic noise.

5. Controlled Hallucination in AI: From Error to Function

Zhou & Fong (2024) — ACM Digital Library
Presents the idea that controlled hallucination can be used in narrative generation and simulation environments
to expand model adaptability and creativity.

6. The Semiotics of AI Hallucination: Synthetic Truth and Symbolic Error

Rosenblatt (2024) — Journal of AI Semiotics
Analyzes hallucination as a semiotic phenomenon, showing that it can function as a symbolic bridge between
data-grounded meaning and emergent conceptual generation.

7. Beyond Factuality: Hallucination as a Cognitive Process in Large Language Models

Kim & Alvarez (2023) — Cognitive Systems Research
Explores the parallels between AI hallucination and human imagination, proposing that both serve adaptive
purposes in exploring unseen conceptual space.

8. Managing Hallucination Budgets in Generative AI

Singh et al. (2024) — IEEE Transactions on Neural Networks and Learning Systems
Describes methods for quantitatively managing hallucination levels through temperature scaling, sampling
variance, and reinforcement learning objectives.

9. The Creative Role of Hallucination in Artificial Agents

Fernández & Liu (2024) — Frontiers in Artificial Intelligence
Investigates hallucination as an emergent creative property that helps artificial agents generate novel,
contextualized responses in open-ended environments.

10. IBM Think Article: What Are AI Hallucinations?

IBM (2024) — IBM Think / Research blog
Summarizes industry perspectives on hallucinations, noting that they can be both an engineering challenge and
a potential creative asset when intentionally leveraged.

Summary

Collectively, these sources suggest that hallucination is not purely a side effect of probabilistic generation,
but also a structural and sometimes intentional feature of AI systems.
In creative, exploratory, or analogical contexts, it can be used as an engine of imagination.