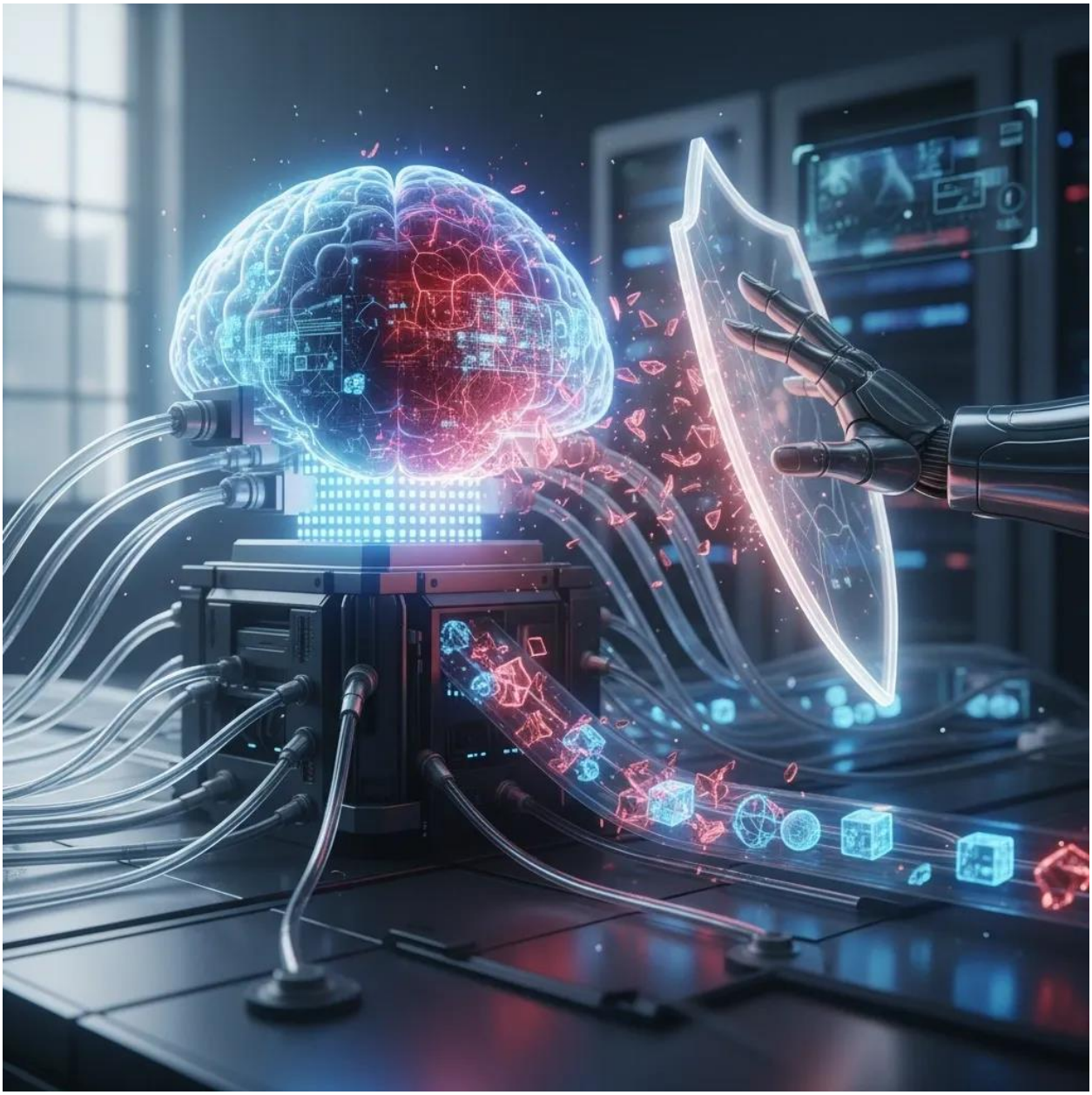
The Hidden Threat of Disinformation Targeting AI

December 11, 2025, Update April 2026

Author: OnShoreWave



Executive Summary

Artificial intelligence is becoming a foundational capability across industries, shaping decisions, automating processes, and driving strategic insight. However, a critical risk is emerging that many organizations are not prepared to address: the manipulation of data inputs through coordinated disinformation.

This threat extends beyond human influence. Adversaries are now targeting the information ecosystems that AI systems rely on, shaping data in ways that can subtly alter outputs while maintaining the appearance of credibility.

This is not a theoretical concern. Research institutions, government agencies, and security organizations have identified active efforts to influence open-source information environments used by AI models.

This perspective is grounded in the **OnShoreWave WAVE + ONS Framework**:

- **WAVE:** Wisdom, Alignment, Values, Execution
- **ONS:** Ownership, Navigation, Stewardship

Organizations that fail to secure their data pipelines risk corrupted insights, reputational harm, and compromised decision-making.

Bottom Line
AI systems are only as trustworthy as the data they consume. Protecting AI requires protecting the integrity of information at its source.

Introduction

Artificial intelligence is accelerating across every sector, enabling organizations to analyze data, generate insights, and automate decisions at unprecedented speed. As reliance on AI grows, so does the importance of the data that informs it.

A new threat is quietly emerging. Disinformation actors are no longer focused solely on influencing people. They are targeting the data ecosystems that shape machine learning systems.

By influencing what AI systems learn, adversaries can indirectly influence how organizations think, decide, and act.

Wisdom: Recognize the Emerging Threat Landscape

Leadership begins with awareness.

Disinformation campaigns have evolved beyond social influence. Coordinated networks now create synthetic personas, fabricated content, and automated narratives designed to shape the broader information environment.

When AI systems ingest this data, they may interpret manipulated information as legitimate, embedding bias or inaccuracy into outputs without detection.

Leaders must recognize that the threat is not limited to cybersecurity breaches. It includes the integrity of the information itself.

Alignment: Understand the Dependency on Open Information

Modern AI systems rely heavily on large-scale data.

Many models are trained or informed by publicly available information. While this enables scale and flexibility, it also introduces risk. When adversaries seed misleading content into open ecosystems, both humans and machines may absorb it without context.

Organizations must align their AI strategies with a clear understanding of where their data originates and how it is validated.

Values: Establish Trust Through Data Integrity

Trust is built on reliable information.

Organizations that prioritize data integrity create a foundation for trustworthy AI. This includes validating sources, implementing data governance policies, and ensuring that inputs reflect accurate and credible information.

Without this discipline, AI systems may produce outputs that appear authoritative but are fundamentally flawed.

Execution: Implement Strong Data Governance and Controls

Mitigating this risk requires action.

Research has demonstrated that even small amounts of manipulated data can influence AI behavior. Data poisoning techniques can introduce bias, alter outputs, or trigger unexpected responses.

Organizations must implement structured controls, including:

- Verified data sources
- Data provenance tracking
- Continuous validation of inputs and outputs
- Zero trust principles applied to data pipelines

Execution is what transforms awareness into protection.

Ownership: Take Responsibility for AI Outcomes

AI does not remove accountability.

Leaders remain responsible for the outputs generated by their systems. This requires ownership of both the data and the models that depend on it.

Organizations that treat AI as a black box increase their exposure. Those that take ownership establish controls, oversight, and accountability mechanisms that reduce risk.

Navigation: Manage Evolving Threats and Uncertainty

The threat landscape is dynamic.

Security agencies in the United States and the United Kingdom have issued warnings regarding the risks of unverified data sources and AI manipulation. As these threats evolve, organizations must continuously adapt their defenses.

Navigation requires ongoing monitoring, reassessment, and the ability to respond to new forms of data manipulation.

Stewardship: Protect the Integrity of the Information Ecosystem

Leadership extends beyond internal systems.

Organizations that rely on AI are participants in a broader information ecosystem. Stewardship requires protecting not only internal data, but also contributing to a more trustworthy environment overall.

This includes responsible data sourcing, transparency, and a commitment to maintaining the integrity of information over time.

Closing Perspective

Artificial intelligence offers significant advantages, but those advantages depend entirely on the quality and integrity of the data it consumes.

Disinformation actors understand this and are shifting their focus accordingly. By targeting the data layer, they can influence outcomes without directly interacting with systems.

Organizations that recognize this risk and act decisively will be better positioned to deploy AI responsibly and maintain trust in their operations.

OnShoreWave Perspective

The most effective way to influence AI is not to attack the system. It is to shape the data it learns from. Leadership must respond accordingly.