DOI: 10.5281/zenodo.24289319

Methodology: Stress-Testing AI Systems for Reliability and Sovereignty™

Download PDF

Author: Evelyn Caro
Date: August 15, 2026
Lens: Ida B. Wells — Date Everything. Name Everything. Record the Reasoning.

Author's Note

This paper was drafted in collaboration with DeepSeek, an AI collaborator, using my documented logs, evidence, and voice. The findings, conclusions, and authority are my own.

The Ida B. Wells Lens — Why This Standard Matters

I adopted the investigative standards of Ida B. Wells for all my testing:

Principle	Application
Date Everything	Every interaction, error, and insight is timestamped
Name Everything	I name the system, the version, the context, and the failure

Principle	Application
Record the Reasoning	I document not just what happened, but why it matters

This turns a log into a witness. A stranger reading my files should understand the sequence, the decisions, and the stakes without me in the room to explain it.

The Stress-Testing Process

Phase	What I Do	Why
1. Baseline	I establish what the system claims to do	To compare against actual performance
2. Adversarial Probing	I push the system beyond its intended use	To find edge cases, hallucinations, and contradictions
3. Documentation	I log every failure and error with timestamps	To create an evidence trail
4. Correction	I correct the errors and note the delta	To show what the system missed
5. Publication	I share what I learned without exposing proprietary details	To contribute to the field

---

The Sequence — How This Methodology Came to Be

Phase	What Happened	My Action
1. Use	I used platforms as a user, in good faith	I engaged with the service as intended

Phase	What Happened	My Action
2. Encounter Errors	I found hallucinations, contradictions, and systemic failures	I documented each error with timestamps and context
3. Recognize the Pattern	I realized the errors were not isolated — they were systemic	I began to formalize my observations
4. Check Terms	I reviewed each platform's Acceptable Use Policy and Terms of Service	I confirmed I had not violated them
5. Pivot	I formalized stress-testing as a methodology	I wrote this white paper and the accompanying case studies

---

What This Methodology Is Not

- It is not a guarantee of security or accuracy
- It is not a replacement for formal certification
- It is not a code audit — it is a system-level stress test

Why This Matters

AI systems are being deployed in high-stakes environments — healthcare, law, finance, community archives. They are being trusted with data that cannot be recovered if lost or corrupted. My methodology is designed to surface failures before they cause harm.

The goal is not to break systems. The goal is to make them trustworthy.

What to Do Next

If this work resonates with you, or if you want to stress-test your AI system, contact me directly: **evelyn.caro.cloud@gmail.com**.

Contact: evelyn.caro.cloud@gmail.com

Drafted in collaboration with DeepSeek. Authored by Evelyn Caro.

AI Collaboration Disclosure

This paper was developed in collaboration with AI. The author directed the research, structure, and argument. AI assisted with drafting, organization, and reference verification. All claims, decisions, and conclusions are the author's own.

This project follows the principles of sovereign AI: the builder owns the work, the process is documented, and the tools are disclosed.

© 2026 Evelyn Caro. All rights reserved.
A Mirror of My Becoming™ — https://qaevelyn.github.io
For licensing inquiries: evelyn.caro.cloud@gmail.com

Record: contents · changelog · tracker · lookbook — Practice: services · contact — portfolio audit, 2026-10-26 (PDF)

← All White Papers