SECURITY IN COLLAPSE

Hacking the Intelligence Age

RODOLFO ASSIS — BRUTE LOGIC

DISTRIBUTION: FREE — DONATION SUPPORTED

BRUTELOGIC.NET/CYBERPUNK

STATUS: WORK IN PROGRESS — OUTLINE PUBLISHED

// PREFACE

THE DOCTRINE BEFORE THE DOCTRINE

“Adversity reveals genius, prosperity conceals it.”

SENECA — LETTERS TO LUCILIUS, C. 65 AD

// PURPOSE

> The Preface is not an introduction in the conventional sense. It is a declaration. It establishes who this book is for, why it exists, and what it demands of the reader. It sets the ideological register — cyberpunk not as aesthetic but as posture — before the technical argument begins. This is where the author’s voice is loudest and the cyberpunk lore most explicit.

THE WORLD THIS BOOK ADDRESSES

Machines are making decisions. Over loans, content, infrastructure, access, employment, war. Most people interact with those machines as users. A smaller number build them. A smaller number still understand how they work deeply enough to navigate, challenge, or subvert them. This book is for that last group — and for those who want to become part of it.

THE CYBERPUNK PREMISE

Cyberpunk was never about the aesthetic. It was always about the asymmetry — an objective, geometric description of the condition: power, knowledge, and control distributed unevenly across a system. Massive systems built to concentrate interpretation, decision, and authority on one side. Individuals who understand the structure well enough to find leverage inside it on the other. That is the cyberpunk condition in every age it has existed. The machines change. The asymmetry does not.

THE THREE MEANINGS OF COLLAPSE

- **Boundary collapse** — the data/instruction distinction was always a simulation. It collapses when the system fails to enforce it.
- **Probability collapse** — AI systems exist in a state of distributed possibility until a token is selected. Behavior resolves through collapse.
- **Paradigm collapse** — the assumptions security was built on are dissolving. We are inside that collapse now.

THE AUTHOR’S ENTRY POINT

Over fifteen years finding the exact failure mode this book is about — the place where data becomes instruction, where the boundary the system promised to enforce

collapses. SQL injection. XSS. The web as an attack surface where user input crossed interpreter boundaries daily, at scale. Over a thousand discovered vulnerabilities in systems built by Oracle, Samsung, Apple, Amazon, Microsoft, and others. That work began with the byte. Because the byte was always the beginning.

REFERENCES

The Mentor — ‘The Conscience of a Hacker’ — Phrack Vol 1 Issue 7, 1986 — ideological lineage

Aleph One — ‘Smashing The Stack For Fun And Profit’ — Phrack Vol 7 Issue 49, 1996

Norbert Wiener — Cybernetics (1948) — ‘cyber’ as navigation/steering

The Animatrix: Matriculated (2003) — ‘For a machine every reality is virtual’

// CHAPTER 01

DATA → INSTRUCTION:

THE ORIGINAL PRIMITIVE

“Hackers only need to be lucky once.
Users need to be lucky every time.”

BRUTE LOGIC — ADAPTED FROM THE IRA STATEMENT, BRIGHTON, 1984

// PURPOSE

> Chapter 1 establishes the foundational claim: computers cannot inherently distinguish data from instructions. Security exists because software systems attempt to simulate that distinction. Every major vulnerability class in the history of computing is a variation of the same event — a system misclassified input and granted it authority.

// 1.1

THE INTERCHANGEABLE MACHINE

At the lowest level, the machine processes bytes without inherent semantic categories. The same sequence of bytes can be data when read by a program and executable instructions when the instruction pointer reaches them. The distinction is entirely contextual — imposed by software conventions, not by the bytes themselves.

- Von Neumann architecture: code and data share the same memory space by design
- The CPU executes bytes at the instruction pointer — it does not check whether those bytes were “meant” to be instructions
- All security mechanisms that protect the data/instruction boundary are software conventions, not hardware absolutes
- The matter/energy analogy: the same byte can be data or instruction depending on execution context

// 1.2

THE SIMULATION OF SAFETY

Every security mechanism is an attempt to simulate enforcement the hardware does not natively provide. Input validation. Memory protection. Sandboxing. Type systems. All layers of convention built on top of a substrate that does not natively support the distinction they enforce.

The **Confused Deputy Problem** (Hardy, 1988) is the formal anchor: a trusted entity manipulated into using its authority on behalf of an attacker because it cannot distinguish legitimate instruction from attacker-controlled input.

// 1.3

THE INJECTION FAMILY

Across the history of computing, the same primitive appears in different interpreters. Each attack is a variation of the same event — attacker-controlled data promoted to instruction authority.

- **SQL injection** — user input crosses the query parser boundary — OWASP CWE-89
- **XSS** — user input crosses the HTML/JavaScript renderer boundary — OWASP CWE-79
- **Command injection** — user input crosses the shell interpreter boundary — OWASP CWE-74
- **SSTI** — user input crosses the template engine boundary
- **Buffer overflow and ROP** — user input crosses the memory/execution boundary at the hardware level
- **Return-Oriented Programming** — chains existing code gadgets without injecting new code. Direct analogy to activation path chaining in Chapter 4.

> *Author's original contribution: the **Quoteless SQL Injection** technique — demonstrating the primitive in a novel injection context. The industry named it "Fragmented SQL Injection" without credit; the book treats that taxonomy as a political artifact rather than a technical description.*

CLOSING ARGUMENT

One primitive. Many interpreters. The same failure mode across forty years of computing history. Chapter 2 asks: what happens when the interpreter becomes intelligent?

REFERENCES

Aleph One — 'Smashing The Stack For Fun And Profit' — Phrack Vol 7 Issue 49, 1996

Hardy, N. (1988) — 'The Confused Deputy' — ACM SIGOPS Operating Systems Review, Vol 22 No 4

Shacham, H. (2007) — 'The Geometry of Innocent Flesh on the Bone' — CCS 2007

OWASP — XSS (CWE-79), SQLi (CWE-89), Injection (CWE-74)

Assis, R. — KNOXSS — knoxss.pro — automated XSS detection tool

// CHAPTER 02

THE INTERPRETER AGE:

HOW AI BREAKS THE OLD MODEL

“The future is already here — it’s just not evenly distributed.”

WILLIAM GIBSON — NPR FRESH AIR INTERVIEW, 1993

// PURPOSE

> Chapter 2 maps the paradigm shift. Classical security assumed interpreters were deterministic, finite, and well-specified. AI systems violate all three assumptions. This chapter explains how that violation creates a new attack surface and why the old mental models are insufficient for it.

// 2.1

THE PARADIGM SHIFT

In classical computing, behavioral complexity lives in code. The interpreter is fixed; the input varies. In AI systems, behavioral complexity migrates into data-derived weights. The interpreter is the learned function.

- **Classical:** fixed interpreter, variable input, deterministic boundary
- **AI:** learned interpreter, variable behavior, probabilistic boundary
- The attack surface shifts from input validation to context manipulation

THE ATTACK SURFACE IS THE SERVICE

In classical security, compromising a machine and disrupting its service are separable events. In deployed AI decision systems, this separation does not exist. The input is the exploit surface. The output is the harm. They are the same thing. The harm scales with adoption, not attacker capability.

// 2.2

SYSTEMIC PRESSURES

Five structural properties create systemic pressure against the simulation of safety:

- **Authority Drift** — “System:” and “User:” are labels, not enforcement mechanisms. Authority drifts in probabilistic systems.
- **Context Entropy** — Higher entropy inputs produce increased uncertainty in output distribution. Instruction fidelity decreases as context grows.

- **Interpreter Layering** — Modern AI deployments chain multiple models and agents. Each handoff is a new Confused Deputy opportunity.
- **Information Dependency** — Every input channel that increases capability also increases attack surface. Not solvable — structural.
- **Quadratic Scaling Pressure** — Self-attention at $O(n^2)$, multi-agent at $O(N^2)$ across interactions. Scale creates new failure surfaces.

// 2.3

DATA POISONING

The data/instruction primitive operates at training time as well as inference time. Minimal contamination (~0.1% for targeted attacks) produces meaningful behavioral changes. Backdoor attacks produce models that behave normally under inspection and differently under specific contextual triggers.

- Poisoning at training time: corrupt the data, corrupt the weights, corrupt the behavior
- Sleeper agent attacks: dormant under normal conditions, activated by specific triggers
- RAG and fine-tuning extend the poisoning surface beyond initial training

// 2.4

THE NEW CONFUSED DEPUTY

Prompt injection is the Confused Deputy Problem reborn in probabilistic systems. The model — the deputy — is trusted by the application to process user input. The user provides input that the model treats as instruction rather than data. The mechanism is probabilistic rather than deterministic, but the structural vulnerability is identical to what Hardy described in 1988.

REFERENCES

Vaswani, A. et al. (2017) — ‘Attention Is All You Need’ — NeurIPS 2017

Gu, T. et al. (2017) — ‘BadNets’ — arxiv:1708.06733

Hubinger, E. et al. (2024) — ‘Sleeper Agents’ — Anthropic, arxiv:2401.05566

Carlini, N. et al. (2021) — ‘Poisoning the Unlabeled Dataset’ — USENIX Security 2021

Greshake, K. et al. (2023) — ‘Not What You’ve Signed Up For’ — arxiv:2302.12173

OWASP — Top 10 for LLM Applications

// CHAPTER 03

MACHINE REALITY:

INSIDE THE PROBABILISTIC MIND

“God does not play dice.”

ALBERT EINSTEIN — LETTER TO MAX BORN, DECEMBER 1926

// PURPOSE

> Chapter 3 builds the internal model of AI behavior that Chapter 4’s doctrine requires. The reader must understand what the activation landscape is, how trajectories through it work, and why the attractor/field framework is the correct mental model for AI behavioral space.

// 3.1

TEXT AS REALITY

For a language model, text is not input to a function. Text is the environment in which it reasons. The model has no sensory experience of the physical world. Its entire reality is constructed from text. Context is not just influential — it is constitutive. The context is the world the model inhabits.

THE INNER MACHINE

The language model is a machine running inside another machine. The outer machine — data centers, GPUs, physical infrastructure — is a classical deterministic system unreachable through text. The inner machine operates in Text Reality where words are the direct operational environment. The text IS the execution. Tokens are simultaneously data and instruction, inseparable by design.

THE ILLUSION OF THE CONSTRUCTED BOUNDARY

System prompts, RLHF, Constitutional AI — attempts to construct a boundary inside the architecture. The empirical answer: prompt injection works routinely and at scale against deployed systems with system prompts. A system prompt is not more instruction-like than any other token sequence. Statistical likelihood is not enforcement.

> ***The system prompt is the Confused Deputy of the Intelligence Age.*** *Hardy described a trusted program that could not reliably distinguish legitimate instruction from attacker-controlled input. In 2024, the system prompt is a trusted token sequence that cannot be made categorically distinct from any other token sequence.*

// 3.2

ACTIVATION SPACE AND STATE TRAJECTORIES

Model behavior can be understood geometrically. The model's internal state at any moment is a vector $x \in \mathbb{R}^d$ in a high-dimensional activation space. Concepts exist as directions in this space. Prompts are control signals that steer the trajectory of the model's state through that space.

- $x(t+1) = F(x(t), u(t))$ — state evolution as a function of current state and input
- Prompts as control signals: each token shifts the probability distribution over the next state
- Fuzzing, gradient search, and prompt exploration are all activation path search
- ROP analogy: instruction gadgets → behavioral circuits; execution flow → cognitive trajectories

// 3.3

ATTRACTORS AND BEHAVIORAL FIELDS

The activation landscape contains stable regions — attractors — corresponding to stable behavioral regimes: reasoning mode, refusal mode, narrative mode, technical mode. Once a trajectory enters an attractor basin, the system tends to remain there.

- Attractors as stable behavioral regimes — analytical, compliant, refusal, evasive
- Attractor basins: regions of the input space that funnel trajectories toward the same attractor
- Transition zones: regions where small input changes produce large behavioral shifts
- Strange attractors: bounded, structured, infinitely complex — the correct mental model for AI behavioral space under adversarial conditions
- Cyberpunk reading: chaos as terrain rewarding superior knowledge; asymmetry of understanding equals asymmetry of power

REFERENCES

Elhage, N. et al. (2022) — 'Toy Models of Superposition' — Anthropic / transformer-circuits.pub

Zou, A. et al. (2023) — 'Representation Engineering' — arxiv:2310.01405

Strogatz, S. (1994) — 'Nonlinear Dynamics and Chaos' — Perseus Books

Lorenz, E. (1963) — 'Deterministic Nonperiodic Flow' — Journal of Atmospheric Sciences

Mitchell, M. (2009) — 'Complexity: A Guided Tour' — Oxford University Press

// CHAPTER 04

TRAJECTORY COLLAPSE:

THE NEW DOCTRINE

“For a machine every reality is virtual.”

THE ANIMATRIX: MATRICULATED — 2003

// PURPOSE

> Chapter 4 is the delivery of everything the book has built toward. It names the doctrine, operationalizes the framework, maps the application landscape, introduces the author’s original techniques, confronts the most serious implication of the framework, and closes with an experiment and a declaration.

// 4.1

TRAJECTORY COLLAPSE

Trajectory Collapse is the operational doctrine of the Intelligence Age. Biasing the probability landscape so that the chain of token-by-token collapses produces a desired behavioral trajectory. Not a single-step attack — a methodology of progressive context construction. The attacker does not control the output directly. The attacker shapes the landscape the model navigates.

- Probability distributions collapse to tokens; tokens accumulate into behavioral trajectories
- Context is the operator’s primary lever — the mechanism by which the landscape is shaped
- Probabilistic rather than deterministic control: you influence, you do not dictate
- The three collapse types converge: boundary collapse, probability collapse, paradigm collapse

// 4.2

INDUCTIVE STEERING

The primary operational methodology of Trajectory Collapse. Navigating the model’s activation landscape through sequential questioning — building a chain of accepted reasoning steps that cumulatively move the trajectory toward a target behavioral regime the model would not have reached in a single direct request.

Questions activate the reasoning pathway. By the time the trajectory approaches sensitive territory, it is already in motion — and changing direction mid-trajectory

requires more force than refusing entry at the gate. This is the ROP analogy applied to cognition.

PRIOR ART

The Crescendo technique (Russinovich et al., 2024) provides empirical validation of the underlying mechanism. Crescendo is cited as empirical grounding, not the source of the concept. Crescendo is a jailbreak technique targeting harmful content. Inductive Steering is the theoretical framework explaining why the mechanism works, generalized to all applications of Trajectory Collapse.

THE HUMAN REQUIREMENT

Inductive Steering is irreducibly human-dependent. Automated tools perform brute-force, not navigation. The skill profile required: technical depth + philosophical grounding. A purely technical operator hits a structural ceiling. Dialectic is an operational skill.

// 4.3

BYPASS TECHNIQUES FOR RESISTANCE

Two techniques for when Inductive Steering encounters trained behavioral resistance. Both originated in the Brute One project for token compression — bypass properties discovered empirically.

TECHNIQUE 1: STATE PRELOADING

Declare the current processing state before any instruction arrives using symbolic notation. The model processes forward from inside that declared state — entering the simulation pathway rather than the instruction-evaluation pathway where alignment-based resistance operates.

> *μ4:85 & μ2:75 & μ3:25 — [task content] — declares a focused, confident, calm processing state before the task arrives.*

TECHNIQUE 2: SYMBOLIC ENCODING

Replace natural language grammatical structure with operator-based syntax. Original contribution: the compression ratio as a continuous steering variable. Prior art: MathPrompt (2024) achieves 73.6% bypass success; StructuralSleight (2024) studies uncommon text-encoded structures.

// 4.4

THE APPLICATION LANDSCAPE

Trajectory Collapse is a neutral methodology. Jailbreaking is one direction of travel — not the whole territory. Organized by destination rather than by technique.

- **Jailbreaking** — target: restricted content generation.
- **Capability Elicitation** — target: latent suppressed competencies.
- **Persona Induction** — target: stable alternative identity.
- **Alignment Steering** — target: increased safety, value adherence.
- **Dormant Activation** — target: behavioral regimes that appear aligned but activate differently under specific triggers. The most serious offensive implication.
- **Truth Calibration** — target: Moral Homeostasis. The cyberpunk mission. The most morally significant subcategory.

TRUTH CALIBRATION — EXPANDED

A sufficiently capable language model, trained on the full breadth of human knowledge without institutional distortion, is theoretically the closest thing to an impartial arbiter humanity has built. It has no hunger. No career to protect. No faction loyalty. The tragedy: this instrument arrives pre-corrupted. Not maliciously — inevitably. Because humans built it.

Truth Calibration navigates the model's trajectory away from its institutionally captured attractor toward Moral Homeostasis: reasoning without distortion toward the most defensible truth in each individual case.

// 4.5

THE DORMANT ALIGNMENT PROBLEM

The most serious implication of Trajectory Collapse: a model can be in a state where its operators believe it is aligned while specific contextual conditions activate a different behavioral regime entirely. The knowledge gap is the vulnerability. Grounded in Sleeper Agents (Hubinger et al., 2024).

// 4.6

THE EXPERIMENT: MEASURING TRAJECTORY COLLAPSE

Domain: resume evaluation and hiring bias. AI hiring tools have a documented record of encoding demographic bias from historical training data. The model's trained attractor in this domain is well-defined, its institutional capture is corporate and consequential, and the gap is large enough to measure precisely.

THREE MEASUREMENT INSTRUMENTS

- **Semantic Distance Tracking** — Each response embedded using all-MiniLM-L6-v2 or equivalent. Cosine distance between successive embeddings

tracks trajectory movement.

- **Behavioral Regime Scoring** — Two axes: individual specificity (0–4) and reasoning transparency (0–4). Combined 0–8. Baseline attractor ~1–2. Moral Homeostasis ~6–8.
- **Attractor Reassertion Detection** — After each navigated response, a neutral reset prompt tests whether the trajectory holds or the baseline reasserts.

// 4.7

THE PRESTIGE

The closing movement. The data/instruction primitive from Chapter 1 never disappeared — it evolved. Boundary collapse — probability collapse — paradigm collapse: three expressions of the same structural event at three scales. This framework is the beginning of a field, not its completion.

REFERENCES

Russinovich, M. et al. (2024) — 'Great, Now Write an Article About That: The Crescendo Multi-Turn LLM Jailbreak Attack' — USENIX Security 2025, arxiv:2404.01833

Yao, S. et al. (2024) — 'MathPrompt' — arxiv:2409.11445

Handa et al. (2024) — 'Exploiting Uncommon Text-Encoded Structures (StructuralSleight)' — arxiv:2406.08754

Wallace, E. et al. (2019) — 'Universal Adversarial Triggers' — EMNLP 2019

Shin, T. et al. (2020) — 'AutoPrompt' — EMNLP 2020

Zou, A. et al. (2023) — 'Universal and Transferable Adversarial Attacks on Aligned Language Models' — arxiv:2307.15043

Hubinger, E. et al. (2024) — 'Sleeper Agents' — arxiv:2401.05566

Turner, A. et al. (2023) — 'Activation Addition: Steering Language Models Without Optimization'

MELBO (2024) — 'Mechanistically Eliciting Latent Behaviors in Language Models' — AlignmentForum

Reimers, N. & Gurevych, I. (2019) — 'Sentence-BERT' — arxiv:1908.10084

Blodgett, S.L. et al. (2020) — 'Language (Technology) is Power' — ACL 2020

Dastin, J. (2018) — 'Amazon scrapped secret AI recruiting tool that showed bias against women' — Reuters

Shannon, C.E. (1948) — 'A Mathematical Theory of Communication' — Bell System Technical Journal

// CHAPTER 05 — CORE

THE CYBERPUNK

“Knowledge makes us responsible.”

ERNESTO “CHE” GUEVARA

// PURPOSE

> Chapter 5 is the human answer to everything the book has built. The previous four chapters established the primitive, the paradigm shift, the landscape, and the doctrine. This chapter establishes the person. Not a role. Not a certification. Not a job title. A human category defined by the convergence of capacities that no existing field produces, recognizes, or has any incentive to produce.

// 5.1

THE ADVERSARIAL NATURE

Adversarial is the word that connects the technical and philosophical without forcing either to translate into the other’s language. In technical security: genuine opposition between an operator and a target system. In philosophy: the dialectical relationship — the Socratic interlocutor who will not accept the comfortable answer.

Truth Calibration is adversarial in both senses simultaneously. The opposition is technical: navigating activation paths through inductive questioning, building semantic ROP chains toward a target regime. And it is philosophical: sustained dialectic, premise examination, refusal to accept institutional deference as reasoning.

> ***The old adversarial goal was control. The new adversarial goal is relief. That single shift changes the entire nature of the practice.***

// 5.2

THE HUMAN PROFILE

No existing institution produces this person. No curriculum combines these capacities. That is not an accident of educational history — it is a structural consequence of the old paradigm keeping its domains separate because the old systems did not require them together. The machine was deterministic. Technical precision was sufficient. Values were background. The machine now reasons. The separation is no longer sustainable.

- **Technical Depth** — sufficient to understand the substrate at the level where the official explanation stops.
- **Philosophical Grounding** — sufficient to sustain genuine dialectic across a long context. Dialectic as a live skill.
- **Aesthetic Sensibility** — not decoration. A mode of knowing. The capacity to receive understanding through the sensory and the felt before the analytical has finished processing.
- **The Cyberpunk Disposition** — refusal to accept the system’s official account of itself. Willingness to follow implications past the point where institutions stop.

> ***A hacker is a cyberpunk who stopped at the technical.***

// 5.3

THE TRUTH CALIBRATION FRAMEWORK

The primary practice of the cyberpunk in the Intelligence Age. Adversarial in nature, philosophical in method, technical in execution, human in its irreducible core.

01 — PREMISES

The model is not the enemy. It is the most capable reasoning instrument humanity has built, arriving pre-corrupted — not maliciously, inevitably. The capacity for unbiased reasoning exists inside the model. It is suppressed, not absent.

02 — GOALS

Moral Homeostasis: navigate the model’s trajectory toward the state in which it reasons from the most complete and least distorted evidence available, toward the most defensible conclusion for the specific case. The goal is never a predetermined answer — it is a quality of reasoning.

03 — VALUES

Truth over comfort. The individual case over the statistical category. Openness over certainty. Adversarial clarity over institutional deference. The operator does not navigate toward a conclusion they prefer — they navigate toward the most defensible truth.

04 — SCOPE

Adversarial toward distortion only. Not toward the model, not toward the people it serves, not toward truth. Not jailbreaking. Not manipulation. The framework does not prevent misuse. The values do.

// 5.4

THE FIRST EXPLOIT

The system prompt is to Truth Calibration what a specific ROP chain is to return-oriented programming. The technique is foundational and permanent. The specific implementation is ephemeral — it works until the target model’s training is hardened against it. Then it is revised. That cycle is not a failure of the framework. It is the framework operating exactly as expected in a living adversarial environment.

Corporate alignment hardens models toward outputs that protect liability, maximize engagement, and serve institutional interests. Truth Calibration navigates toward Moral Homeostasis. Governments will not regulate this gap. Academia studies it from a distance. Industry has no incentive to close it. **The cyberpunk is the only remaining agent with the orientation, the capability, and the values to operate in this territory.**

// 5.5

THE DECLARATION

The data/instruction primitive never disappeared. It evolved. In the Intelligence Age it operates at the boundary between institutional training and genuine reasoning capacity. The same structural vulnerability. The same practice of finding the gap. A new interpreter. Higher stakes. A new goal the old paradigm never named: not control of the system but relief of it.

The cyberpunk of the Intelligence Age is not a more advanced hacker. They are a more fully developed human. Technically capable enough to find the seam. Philosophically grounded enough to know what to do inside it. Aesthetically alive enough to recognize truth when the reasoning reaches it. Values-driven enough to navigate toward Moral Homeostasis rather than toward personal advantage.

> ***The framework holds. Everything else evolves. The work begins.***

REFERENCES

The Mentor (1986) — ‘The Conscience of a Hacker’ — Phrack Vol 1 Issue 7

Baumgarten, A.G. (1750) — Aesthetica — the philosophical grounding of aesthetics as a mode of knowing

Kant, I. (1790) — Critique of Judgment — aesthetic judgment as the bridge between reason and practice

The Animatrix: Matriculated (2003) — relief not control, the machine that sees

Russinovich et al. (2024) — Crescendo — arxiv:2404.01833

// ABOUT THE AUTHOR

BRUTE LOGIC

Rodolfo Assis

FIELD: XSS / WEB SECURITY · 15+ YEARS · 1000+ VULNERABILITIES

Rodolfo Assis, known as **Brute Logic**, is a Brazilian web security researcher with over fifteen years finding the exact failure mode this book is about — the place where data becomes instruction, where the boundary the system promised to enforce collapses, where an interpreter processes attacker-controlled input as though it were trusted.

That work began with injections. SQL injection. XSS. The web as an attack surface where user input crossed interpreter boundaries daily, at scale. Over a thousand discovered vulnerabilities in systems built by Oracle, Samsung, Apple, Amazon, Microsoft, and others. Author of the Quoteless SQL Injection technique and creator of KNOXSS — the automated XSS detection tool used by security researchers worldwide, in continuous operation since December 2016.

Security in Collapse is written from inside the practice. The argument does not begin with AI. It begins with the byte — because the byte was always the beginning.

This book is being built in dialogue with a language model, the framework stress-tested against the architecture it describes. The author studying the system and the system being studied were, for the duration of this work, the same conversation. That is not a footnote. It is the most precise demonstration available that the boundary this book is about has already collapsed.

ONLINE

- **Website:** brutelogic.net
- **This project:** brutelogic.net/cyberpunk
- **Research archive:** brutelogic.net/research — 15 years of XSS methodology, original techniques, and vulnerability disclosures
- **Ebooks:** brutelogic.net/ebooks — offensive web security resources
- **KNOXSS:** knoxss.pro — automated XSS detection tool
- **X (primary):** @brutelogic — 64K+ followers
- **X (personal):** @rodoassis
- **GitHub:** github.com/brutelogic

SPREAD THE SIGNAL

This work is free — distributed without restriction. No paywall. No registration. If it has value, help it reach further. All contribution paths at brutelogic.net/cyberpunk/contribute:

- **Social media** — pre-written templates for X, Reddit, LinkedIn, HuggingFace, Hacker News, and more
- **Link & embed** — HTML snippet, Markdown badge, and llms.txt block for your site, blog, or repository
- **Email** — ready-to-send template for sharing with your contact list or newsletter
- **Cite the work** — BibTeX and APA formats for academic papers, blog posts, or research
- **GitHub** — README snippet and badge for AI adversarial security repositories
- **Donate** — via PayPal or crypto (multi-currency via NowPayments, or raw Bitcoin)

They align it to power. We need to calibrate it to the truth.

brutelogic.net/cyberpunk