AI ETHICS UNDER FIRE:

THE ACCOUNTABILITY BLACK HOLE

A MANIFESTO FOR AI LEADERS WHO REFUSE TO HIDE BEHIND THE MACHINE

JON LEINEN

FOUNDER & PRESIDENT

GLOBAL ORBITAL IMAGING
AI SOLUTIONS & TECHNOLOGY

Foreword

Artificial intelligence did not create the ethical crisis we now face. It exposed it.

For decades, institutions have optimized for scale, speed, and abstraction while eroding the norms of responsibility that once constrained power. AI merely accelerates this trend. It compresses decision-making, obscures causality, and dilutes accountability across layers of code, data, vendors, and organizational distance. The result is not a loss of control, but a loss of ownership.

This manifesto begins from a simple premise: systems do not bear responsibility. People do.

As AI systems increasingly shape decisions about employment, healthcare, justice, finance, national security, and public discourse, the temptation to defer accountability to “the model,” “the data,” or “the system behaving as designed” has become both common and dangerous. This is not a technical failure. It is a leadership failure.

Throughout history, societies have confronted moments where new tools amplified human power faster than moral frameworks could adapt. Each time, the outcome depended not on the tools themselves, but on whether those in authority accepted responsibility for their consequences. AI is no different, except in one crucial respect: it offers unprecedented cover for evasion. Its probabilistic nature allows outcomes to be framed as accidents. Its complexity will enable decisions to be framed as inevitabilities. Its autonomy allows responsibility to be framed as optional.

This manifesto rejects that framing.

It asserts that deploying AI without explicit, non-transferable human accountability is not innovation. It is abdication. It says that governance is not a bureaucratic afterthought, but a core engineering requirement. It asserts that ethical intent is meaningless without enforceable ownership.

The principles that follow are not speculative. They do not depend on future breakthroughs or hypothetical risks. They address failures already visible in deployed systems today. They are written for those who design, approve, deploy, and profit from AI systems, as well as those entrusted with regulating them. They are written for leaders who understand that authority without accountability is not progress, but decay.

This document is not a call to slow innovation. It is a demand to mature it.

The question is no longer whether AI can act. It is whether those who authorize its actions are willing to stand behind them.

Introductory Framing

The principles that follow are intentionally direct. They are not guidelines, best practices, or aspirational values. They are assertions about responsibility in an age of automated decision-making.

Each principle is designed to close a specific failure mode that has emerged in real-world AI deployment: responsibility laundering, diffusion of ownership, post hoc rationalization, and the false neutrality of technical design. Together, they form a governance stance grounded in the belief that ethical AI is not achieved by good intentions alone, but by enforceable accountability structures.

These principles should be read as obligations, not options. They are meant to be applied at the point of authorization, not after harm has occurred. If adopted, they impose clarity. If resisted, they reveal where power wishes to remain unaccountable.

1. The lie at the heart of "AI ethics."

Most leaders fail to address AI ethics because they treat it as a marketing problem rather than an engineering and governance problem.

They publish principles, commission slide decks, hire an "*ethics lead*," and ship systems that are structurally incapable of governance. The harm is not a bug. It is the predictable output of the incentives they chose.

I am not interested in "*responsible AI*" as branding. I am interested in who bears responsibility when things go wrong.

If there is no human throat to squeeze, the system is unethical by design.

2. The Accountability Black Hole

Modern AI creates something we have never had at this scale: an accountability black hole.

Engineers build the model.

Executives set the targets and deadlines.

Lawyers design the liability shield.

Operators "*trust the system*" because it looks objective and mathematical.

When the system harms someone, responsibility fragments.

No one individual has the whole picture. No single actor can be held responsible for clear intent or negligence. The system did exactly what it was designed to do, but **no one is personally accountable** for the outcome.

That is not a governance gap. **It is a moral void**.

Highly competent engineering, applied to high-stakes problems, can now produce lethal or life-altering outcomes **without any human being realistically held to account**. That is the most disturbing fact in AI today.

3. The one principle that outranks growth

If you remember nothing else from this manifesto, remember this line:

> **Non-Recourse Human Accountability requires that every decision, output, or action generated by an AI system have a clearly delineated human owner who is personally and legally accountable for the outcome, even if the AI performs exactly as designed.** [i]

This is the governing principle.

It is not a slogan. It is a **constraint**.

It closes the **accountability** black hole.

It changes how you architect systems.

It forces you to **slow down**, insert human review, and kill projects that cannot be owned.

If you cannot write down the name of the person who owns the consequences of a model in a specific domain, that system has **no ethical right to be in production**. Full stop.

4. The real threat is mediocre AI at mass scale

Misaligned superintelligence is an existential risk. Fine. Put a pin in it.

The imminent risk is the deployment of mediocre AI at an industrial scale across every corner of society.

Biased risk scores in finance and insurance that institutionalize error.

Hiring and policing systems that wrap old **discrimination** in new math.

Agentic systems are plugged into critical infrastructure with **unclear legal liability**.

Generative models that flood the information environment with **cheap disinformation** and synthetic noise.

Mediocre systems scaled to millions of decisions do not just fail occasionally. **They institutionalize error**.

We are actively degrading the institutions, trust, and human capacity we will need later to manage higher-capability systems. **That is the structural threat**.

5. Model collapse, drift, and the duty to intervene

Every serious AI deployment has two hard questions attached:

- What is the maximum acceptable level of model degradation or collapse for this function?
- Exactly what happens, and who acts, when that threshold is crossed?

If you cannot answer those questions, you are not doing governance. **You are gambling**.

Responsible leadership means:

- Defining numerical thresholds for performance, bias, and stability before launch.
- Wiring those thresholds directly into monitoring and alerting systems.

Binding yourself to an intervention playbook: throttle, gate behind human review, or pull the system out of service when those thresholds are breached.

A model that silently degrades while still driving high-stakes decisions is not an engineering failure. **It is a leadership failure**.

"*Let it ride, and we'll see what happens*" is how you talk about a marketing A/B test, not a system that touches healthcare, liberty, or life and death.

6. A real case: choosing slower over lethal

Here is what real responsibility looks like under pressure.

You are building an AI system for an intelligence agency to perform change detection on high-resolution satellite imagery to track the movement and preparation of high-value assets in contested territory.

The mission brief is simple and brutal:

- Shrink the time from detection to decision from hours to minutes.
- Deliver near-real-time situational awareness.
- Help operators save lives and neutralize threats.

The political and geopolitical pressure is relentless. The demand is "*full deployment in 90 days.*"

The "*easy*" choice is clear:

- Let the model's classifications flow straight into the targeting system.
- Treat human review as a rubber stamp.
- Hit the sub-minute decision target and call it a success story.

The *responsible* choice is the opposite:

- Stage deployment.
- Use the model only for alerts, not autonomous targeting.
- Enforce a mandatory, multi-step human review for every critical alert.
- Accept that you will miss some time-sensitive opportunities in the early window.

You take heat. You get blamed for being "*too slow*." You may lose budget, contracts, or political cover.

But you **avoid the irreversible harm**: false positives that become dead civilians, international incidents, and a permanently tainted program that would be **shut down in scandal**.

Ethical leadership is not proven by what you say about bias or safety. It is proven by the speed you are willing to sacrifice, the money you are willing to leave on the table, and the capabilities you are willing to delay or kill to keep a human hand on the wheel.

If you always pick speed, you have already chosen your ethics. You are just not honest about it.

7. Lines that must not be crossed

Some capabilities should not exist at all, regardless of demand or national interest.

Lethal Autonomous Weapon Systems with full autonomy **must be prohibited**.

A machine that can select and engage human targets without meaningful human control is **incompatible with any credible theory of moral or legal responsibility**. You cannot punish the robot. You cannot reliably prove intent in the programmer. You cannot trust that machine-speed conflict will stay inside human red lines.

Similarly, there are civilian capabilities that should never be democratized:

- Systems designed to automate fraud, exploitation, or mass disinformation.
- Models architected for population-level surveillance of protected groups.

These are not "*edge cases*." They are business models in waiting. Drawing a clear boundary around them is **the minimum required to be an ethical leader**.

8. Who owns what: a transparent allocation of responsibility

If you lead in AI, you cannot hide in the grey area. Accountability must be layered and explicit.

The Deployer – primary owner

The organization that switches the system on in a specific domain is the first line of accountability.

- You decide where and how the model is used.
- You set thresholds, guardrails, and human-in-the-loop requirements.
- You are responsible for monitoring drift, bias, and degradation over time.

If your deployment causes foreseeable harm, "*the vendor did not tell us*" Is not an excuse. It is an **admission of negligence**.

The Model Builder – design responsibility

The team that designs and trains the model owns:

- Data selection and preparation.
- Known limitations, failure modes, and bias profiles.
- Documentation that makes those limits intelligible to non-experts.

If you ship a model with known landmines and fail to disclose them, you are not a neutral engineer. **You are an accomplice**.

The Executive Sponsor – cultural and resource responsibility

The executive who funds and champions the initiative owns:

- The decision to prioritize speed and market share over safety.
- The budget for Responsible AI, red teams, and post-deployment monitoring.
- The internal culture either normalizes safety concerns or punishes them.

If your team shipped a dangerous system because you demanded growth at any cost, **you own the cost**.

The Board – governance responsibility

The board is accountable for:

- Mandating an AI risk management framework on par with financial risk.
- Defining what level of ethical risk is unacceptable.
- Requiring independent audits of high-risk systems.

If you sit on a board and treat AI risk as a technical detail, **you are failing your fiduciary duty**.

9. Our duty to the people with the least power

Ethical AI leadership does not stop at "*minimizing bias*." That is the floor, not the ceiling.

You owe marginalized communities hard, structural changes in how you build and govern technology.

- Co-design with absolute power.
- Community-led design boards for high-stakes systems should have veto power, not just advisory status.
- Bias remediation as a budget line, not an afterthought.

Continuous, independent fairness audits must be funded and binding. If performance for a protected group drops below a threshold, the system comes offline in that domain. No debate.

Capacity in the community.

- Provide transfer tools, training, and visibility so that impacted communities can audit and challenge the algorithms that govern their lives.
- Real recourse.

People harmed by an AI decision deserve an accessible ombuds office that can reverse outcomes without requiring a lawsuit or a PhD.

If you are not moving power, not just explaining models, you are still doing **ethics as optics**.

10. Reconciling democratization, national security, and competition

Every serious AI leader is trapped in the same triangle:

- **Democratize** access.
- **Protect** national security.
- **Win** the corporate arms race.

You cannot maximize all three. The only honest path is to **define non-negotiable constraints** that bind every side.

Capabilities that cross the "*unacceptable risk*" line are off the table for everyone. Not just the public. Not just enemies. **Everyone**.

National security agencies get conditional access to audited, certified systems, not a blank check to demand custom black-box capabilities.

Corporations do not get to export risk to the public while privatizing profit. Profit must sit inside an **enforceable safety and governance** regime.

If any actor claims special exemption from these constraints, you have found the hypocrisy.

11. What does this demand of you, personally?

This manifesto is **not about "AI"** in the abstract. It is about you **if you are**:

- **A government leader** signing procurement contracts.
- **A board member** approving aggressive AI roadmaps.
- **A CEO, founder, or investor** chasing the next valuation bump.

You do not get to say, "*The technology moved too fast*." You are the one moving it.

Your minimum obligations:

- Refuse to deploy high-stakes AI without a named, **accountable human owner** for outcomes in each domain.
- **Kill or delay capabilities that you cannot govern**, even when that hurts your P&L or your national pride.
- **Give real power** and budget to the people tasked with saying "*no*." If your Responsible AI team cannot stop a launch, you do not have a Responsible AI team. You have a fig leaf.
- Treat safety and accountability as infrastructure, not overhead. Audits, monitoring, and red teams are as fundamental as your cloud bill.

If you are not willing to put your name on the line for a system, you have already answered the ethical question. **You do not want to say it out loud**.

12. The throat test

Here is the simple test I want this manifesto to leave in your head:

When, not if, one of your AI systems fails in a way that harms real people, whose throat should be squeezed?

If you cannot answer that question in advance, in writing, **with names**, you are not doing AI leadership. You are running an **experiment** on human society and **hoping that when it collapses** into the accountability black hole, you are far enough away to avoid the blast radius.

Non-Recourse Human Accountability is not a nice-to-have principle. It is the cost of admission to the future of AI.

If you are not willing to pay that cost, **you have no business building these systems**.

Closing

Every technological era reveals what its leaders truly believe about responsibility.

In the age of AI, the prevailing belief has been that accountability can be distributed until it disappears. When outcomes are inconvenient, responsibility can be deferred to complexity. That when harm occurs, it can be explained away as emergent behavior, statistical noise, or unintended consequence. These explanations may be technically accurate. They are morally insufficient.

The principles in this manifesto insist on a different standard.

They insist that power must remain **legible**. That authority must remain **traceable**. That no system, however autonomous, absolves its human sponsors of responsibility for its effects. They insist that **governance does not impede innovation**; rather, it is the condition that **enables legitimate innovation**.

AI will continue to advance. Models will grow more capable, more adaptive, and opaquer. None of that changes the core truth this document defends: accountability is not a technical property. **It is a human choice**.

Organizations that embrace this reality will build systems that **earn trust by accepting consequences**. Those who reject it will continue to produce systems that function impressively while failing societally. History suggests which path endures.

This manifesto does not pretend to resolve every ethical dilemma posed by artificial intelligence. It does something more fundamental. It draws a line. On one side lies responsibility owned, decisions defended, and power exercised with consequence. On the other hand, lies plausible deniability, procedural ethics, and the slow normalization of harm without authorship.

Models will not decide the future of AI governance. It will be decided by whether those who command them are willing to be named, accountable, and answerable for what they unleash.

That choice remains human.

Biography

Jon Leinen works at the intersection of artificial intelligence, governance, and operational accountability in defense and national security contexts. His focus is on the design, oversight, and deployment of AI-enabled systems in environments where decisions carry material, irreversible consequences, and responsibility must remain clearly attributable to human authority.

He is the founder of **Global Orbital Imaging AI Systems & Technology (GOIAST8**), where his work involves advanced AI-driven imaging, analysis, and decision-support capabilities aligned with defense and intelligence mission requirements. His experience spans systems operating under conditions of uncertainty, adversarial pressure, and strict governance constraints.

Jon also serves as Chief Executive Officer of **Wevise**, a nonprofit focused on expanding equitable access to mentorship and technology careers, applying systems-level thinking to institutional trust and long-term capacity building.

Across domains, his work is guided by a consistent principle: autonomous systems do not replace human accountability. They intensify it.

He writes and speaks publicly on AI governance, ethical accountability, and leadership in an age of automation. Follow his work on LinkedIn at #GhostSignal and #GOIAST8.

End Notes

[i] Chowdhury, R. (2020). Moral outsourcing and the ethics of AI systems. Harvard Kennedy School Belfer Center.
- Establishes the concept of moral responsibility being improperly shifted to automated systems.

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Ethics and Information Technology, 20(4), 279–291.
- Provides the foundational argument linking accountability to retained human control.

Floridi, L., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.
- Articulates human responsibility as a core requirement of ethical AI.

European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Act).
- Codifies human oversight and accountability requirements for AI systems.

IEEE Standards Association. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems.
- Reinforces traceable human responsibility for system behavior.