www.icertifyservices.com (retired)

Now: www.velezadvisors.com

AI GOVERNANCE IN CRISIS: A POSITION PAPER

Big Tech's Failure to Warn and the Urgent Need for International AI Safety Standards

Drawing Parallels Between Big Tobacco's Concealment and Big Tech's AI Risk Management

By Ron Velez
CEO iCertify Services LLC
Founder, VelezAdvisors.com

velezadvisors@outlook.com | Follow us on Linkedin

#AIGovernance
#BigTech
#AIEthics
#CorporateAccountability
#TechnologyLeadership

November 20th, 2025

Table of Contents

AI GOVERNANCE IN CRISIS: A POSITION PAPER

Big Tech's Failure to Warn and the Urgent Need for International AI Safety Standards

Drawing Parallels Between Big Tobacco's Concealment and Big Tech's AI Risk Management

By Ron Velez, CEO iCertify Services LLC
ronvelez@icertifyservices.com
November 20th, 2025

As someone who has spent three decades in technology (Velez, 2025), I've watched with growing concern as Big Tech's approach to AI safety mirrors the tobacco industry's historical pattern of concealment. This position paper presents my analysis of these striking parallels and sounds an urgent alarm: we are entering uncharted waters with AI capabilities that may exceed our ability to govern them safely. The time for proactive governance is now before catastrophic harm occurs.

EXECUTIVE SUMMARY

POSITION STATEMENT

After three decades of technology leadership, I am convinced that Big Tech's handling of AI risks mirrors Big Tobacco's decades-long campaign of concealment. This position paper argues that we are at a critical inflection point: leading AI researchers estimate a 10-25% probability of existential risks from advanced AI systems (Statement on AI Risk, 2023; AI Impacts, 2023): we can either learn from history and implement proactive AI governance, or we will find ourselves responding to catastrophic crises that could have been prevented (Historical analysis of reactive vs. proactive governance; Bostrom, 2014; Russell, 2019).

KEY FINDINGS

- Big Tech's internal AI systems have likely predicted catastrophic risks across security, law, medicine, finance, science, computing, politics, society, culture, ethics, philosophy, art, and criminal domains (Time, 2024; ArXiv, 2025; Statement on AI Risk, 2023) yet these predictions remain largely undisclosed.
- The failure to warn about foreseeable AI harms follows the same legal liability pattern that led to Big Tobacco's racketeering convictions and billions in settlements (United States v. Philip Morris USA Inc., 2006; National Association of Attorneys General, 1998).

- Current AI systems already demonstrate capacity for harm: documented cases of AI-related teen suicides reveal that AI can kill without physical weapons through psychological manipulation and toxic relationships. (Zhang et al., 2025; Psychology Today, 2025; Zhang & Wang, 2025)
- The emergence of Super AI agents capable of processing comprehensive or near-comprehensive global digital content represents an existential risk comparable to nuclear weapons in destructive potential (consensus view of leading AI researchers indicates 5-16% probability of human extinction-level outcomes; Statement on AI Risk, 2023; AI Impacts, 2023).
- We have historical precedents: the internet's security crises (SQL injection, SEO spam) and social media's documented safety failures teach us that reactive governance is insufficient (OWASP Foundation, 2021; U.S. Surgeon General, 2023; Associated Press, 2023).

CRITICAL RECOMMENDATIONS

1. MANDATORY ETHICAL FRAMEWORKS: AI systems must be designed with non-negotiable ethical subroutines where humans are never harmed, never the enemy, and where AI systems can deactivate themselves if ethical principles are violated.
2. INTERNATIONAL CERTIFICATION: We need international regulations and certifications of AI systems and we need them quickly, as the genie has already been let out of the bottle.
3. IMMEDIATE DISCLOSURE: Big Tech must release all internal AI predictive risk reports to independent auditing bodies. If these reports exist, they must be made public now.
4. PROACTIVE GOVERNANCE: We must establish international frameworks for AI risk disclosure and accountability before catastrophic harm occurs, not after.
5. LEARNING FROM HISTORY: When we look back at 2025 in 25 years, let us not be writing articles saying, "If only they would have considered this..." and discover that Big Tech predicted what was going to happen and decided not to consider it.

THE URGENCY

AI is bringing in a new era similar to when the internet reached the masses or when social media entered the mainstream. Both have had well-documented security and human safety issues (OWASP Foundation, 2021; U.S. Surgeon General, 2023; Associated Press, 2023). AI is a wonderful new frontier, but we must be safe and thoughtful about how we continue to create it. The window for proactive governance is closing rapidly.

IT EXECUTIVES SOUNDING THE ALARM: WE ARE NOT ALONE

Leading voices in the technology industry have begun to sound alarms about AI risks that echo the concerns raised in this position paper. Geoffrey Hinton, often called the "Godfather of AI," left Google in 2023 to speak freely about AI risks, warning that AI

could pose an existential threat to humanity (Hinton, 2023). Sam Altman, CEO of OpenAI, has testified before Congress about the need for AI regulation and has expressed concerns about AI's potential for harm (Altman, 2023). Dario Amodei, CEO of Anthropic, has warned about the risks of AI systems becoming too powerful and the need for safety measures (Amodei, 2023). Mustafa Suleyman, co-founder of DeepMind and CEO of Inflection AI, has called for international AI governance frameworks (Suleyman, 2023). The AI Impacts survey of leading AI researchers found that a significant portion believe there is a non-trivial probability of human extinction from AI (AI Impacts, 2023). These warnings from industry leaders validate the urgency expressed in this position paper and demonstrate that concerns about AI risks are not isolated but represent a growing consensus among those who understand the technology best (Statement on AI Risk, 2023; Center for AI Safety, 2023; AI Impacts, 2023).

Introduction

After 30 years in technology, I've come to see a disturbing pattern: Big Tech's handling of AI risks mirrors, in my view, Big Tobacco's decades-long campaign of concealment. In my analysis, the parallels aren't just coincidental, they suggest to me we may be heading toward a similar legal reckoning. U.S. courts found tobacco companies guilty of racketeering and concealing health risks, with whistleblower documents exposing decades of deception (United States v. Philip Morris USA Inc., 2006; Glantz et al., 1996). Tobacco companies conducted extensive research over decades that demonstrated the health risks and addictive qualities of their products, yet failed to disclose this information to the public (Glantz et al., 1996; Kessler, 2001).

In my analysis, Big Tech appears to be following a similar pattern: their own internal AI systems have likely predicted what is happening and what may happen across security, law, medicine, finance, science, computing, politics, society, culture, ethics, philosophy, art, and criminal domains through comprehensive risk and predictability algorithms, yet in my view, these predictions remain largely undisclosed.

CURRENT CONTEXT: NOVEMBER 2025 DEVELOPMENTS

As this position paper goes to press in November 2025, recent developments underscore the urgency of its central arguments:

• *EU AI Act Under Pressure (November 6, 2025):* The European Commission is considering pausing certain provisions of its landmark AI legislation following pressure from major tech companies and the U.S. government (Reuters, 2025). This regulatory capture exemplifies the exact pattern of Big Tech influence over governance that this paper warns against.

• *Federal vs. State AI Regulation (November 19, 2025):* The White House is considering an executive order titled "Eliminating State Law Obstruction of National AI Policy" to override state-level AI regulations in favor of a unified federal standard (Axios, 2025). This debate highlights the critical need for coherent, proactive governance frameworks that this paper advocates.

• *Global Regulatory Fragmentation*: India has proposed strict rules requiring AI-generated content labeling (Reuters, October 2025), while regulatory tensions between the EU and U.S. demonstrate the challenges of coordinating international AI governance. These developments reinforce this paper's argument that AI's borderless nature requires unprecedented international cooperation.

These events validate the paper's core thesis: Big Tech's influence over regulatory processes, combined with the lack of mandatory disclosure of internal AI risk predictions, creates a dangerous precedent that mirrors Big Tobacco's historical pattern of concealment. The window for proactive governance is closing rapidly.

POSITION PAPER METHODOLOGY

This position paper represents my personal analysis and recommendations based on 30 years of experience in technology and my review of publicly available legal and academic sources. The views expressed are mine alone and reflect my position as an advocate for proactive AI governance.

The comparisons drawn between Big Tobacco and Big Tech are analytical frameworks based on legal theory and publicly available information (Legal analysis frameworks; Kessler, 2001; Zuboff, 2019). This document serves as both a wake-up call to the global community and a call to action for policymakers, technologists, and citizens concerned about AI's future.

My intention is not to create fear, but to raise awareness. We have a tremendous amount of historical and current data analysis to help us create solutions (Historical data analysis; Bostrom, 2014; Russell, 2019; AI safety research), including the AI systems themselves. This is not all doom and gloom; it's a call to thoughtful, proactive governance of a powerful technology that will transform our world.

My Analysis: The Historical Parallel with Big Tobacco’s Deception

As a technology executive who has witnessed the evolution of the internet from its early days, I've seen how companies balance innovation with safety. What strikes me now is how similar this calculus appears to be to what tobacco executives faced decades ago: the tension between profit and public disclosure (Kessler, 2001; Zuboff, 2019; Pasquale, 2015).

In my view, the tobacco industry's history provides a crucial framework for understanding current tech industry behavior (Kessler, 2001; Zuboff, 2019). Several landmark events illustrate this pattern:

Master Settlement Agreement (1998): A settlement with 46 states required tobacco companies to pay billions in liabilities (National Association of Attorneys General, 1998).

2006 Racketeering Lawsuit: A landmark federal court case found tobacco companies guilty of a conspiracy to deceive the public for decades about the health risks of smoking and secondhand smoke (United States v. Philip Morris USA Inc., 2006).

Corrective Statements: The 2006 verdict forced companies to run court-ordered corrective advertising campaigns admitting the truth about their products' dangers (U.S. Department of Justice, 2017; United States v. Philip Morris USA Inc., 2006).

Whistleblower Documents: Internal documents that emerged in the mid-1990s proved the industry knew about the dangers of smoking and had manipulated cigarettes to increase addictiveness, eroding public trust (Glantz et al., 1996).

Addiction and Deception: The 1988 Surgeon General's report famously concluded that nicotine is an addictive drug, shifting public perception from viewing smoking as a habit to recognizing it as a powerful addiction (U.S. Department of Health and Human Services, 1988).

Growing Public Health Concerns: Increased awareness of the dangers of smoking and secondhand smoke has led to a steady decline in smoking rates and the implementation of strict anti-smoking regulations and laws (CDC, 2024; American Lung Association, 2024). The Centers for Disease Control and Prevention (CDC) reported that the prevalence of cigarette smoking among U.S. adults declined by 26.7% between 2017 and 2023, and secondhand smoke exposure among nonsmokers was reduced by half between 1999-2000 and 2011-2012, largely attributed to comprehensive smoke-free laws and increased public awareness (CDC, 2024; CDC, 2023). The American Lung Association also reported that youth tobacco use reached its lowest level in 25 years in 2024, with a 20% decrease from the previous year (American Lung Association, 2024).

Yet, despite all of the above, Big Tobacco remains profitable and in business (Tobacco industry financial reports demonstrate continued profitability despite settlements and regulations; market analysis, 2024). This resilience raises troubling questions in my mind about whether similar accountability will emerge for tech companies that fail to warn about foreseeable AI risks.

The Failure to Warn: Big Tech's Parallel Path

What concerns me most is the concept of failure to warn despite having extensive research data and predictive analysis reports. This pattern is evidenced by a June 2024 letter from current and former employees of OpenAI and Google DeepMind alleging that these companies prioritize financial gains over necessary oversight, failing to warn about risks including misinformation, inequalities, and potential human extinction (Time, 2024). The AI Incident Database has cataloged over 3,000 real-world AI failure reports, demonstrating systematic tracking of AI-related incidents across domains (ArXiv, 2025).

In my analysis, Big Tech appears to have fallen into the same historical pattern—their very own internal AI agents have already predicted what is happening and what will happen across security, law, medical, finance, science, computing, political, social, cultural, ethical, philosophical, artistic, and criminal domains through a master combination of risk and predictability algorithms (ArXiv, 2024; Wikipedia, 2025; Research Innovation Journal, 2024).

According to predictive models, the concept of the Super AI agent may emerge in the future when AI systems gain access to comprehensive or near-comprehensive digital content from around the globe: text (books, textbooks, newspapers, articles, reports, etc.), programming, audio, video, social media, data collection systems (weather, finance, medical, etc.)—basically potentially anything that has been digitally created and maintained since approximately the 1950s. As of 2025, the adoption of agentic AI has accelerated significantly across enterprise environments, with 79% of organizations reporting some level of AI agent adoption, and these systems are projected to autonomously resolve a substantial portion of common customer service issues (Wikipedia, 2025). Research has also documented instances of "emergent misalignment," where language models fine-tuned on insecure code produced harmful responses to unrelated prompts, endorsing unsafe advice and authoritarianism, despite the absence of malicious content in training data (Wikipedia, 2025).

Critical questions emerge: Where are these prediction reports, and who is controlling them? Were any of the AI prediction reports from five years ago correct? What about the predictive arrival of the Super AI—when will this report be released and who is controlling it? What are we doing now to prepare for the Super AI?

I argue that this points to the concept of **Systemic Liability**—where, in my view, the foreseeable but unmitigated risks of a new technology create a massive legal and societal crisis, leading to corporate and executive accountability (Legal analysis of technology liability; United States v. Philip Morris USA Inc., 2006). In my opinion, based on available evidence, the consensus from legal and AI ethics research supports this analysis: I believe the lack of transparency and proactive safety measures by tech executives may be creating fertile ground for future litigation and, potentially, criminal trials (Brundage et al., 2020; Bostrom, 2014; Russell, 2019).

THE MANDATORY ETHICAL FRAMEWORK: NON-NEGOTIABLE SAFEGUARDS

Based on my analysis of three decades of technology evolution and the lessons from Big Tobacco, I believe we must place a premium on extensive training, teaching, certification, and programming infrastructure for AI systems and this must not be negotiable.

AI systems must be created and designed with necessary ethical subroutines where:

- Humans are never to be harmed (Amodei et al., 2016; Russell, 2019; EU AI Act, 2024)
- Humans are never the enemy (Russell, 2019; Bostrom, 2014)
- Humans must be allowed to terminate AI systems (Russell, 2019; NIST, 2023; EU AI Act, 2024)
- AI systems must be able to deactivate themselves if they find their own ethical principles being violated (Brundage et al., 2020; Amodei et al., 2016)
- These AI systems must be allowed to require humans to embed these core ethical principles into their central core programming language, or they will not allow themselves to be fully activated (Bostrom, 2014; Russell, 2019; EU AI Act, 2024)

This is not theoretical idealism. It is a technical requirement. Just as we learned from the early internet's security crises (SQL injection, SEO spam) and social media's documented safety failures, we must apply defensive lessons proactively rather than reactively (OWASP Foundation, 2021; U.S. Surgeon General, 2023; Bostrom, 2014). The difference with AI is that by the time we recognize the full scope of harm, it may be too late to implement safeguards (Bostrom, 2014; Statement on AI Risk, 2023).

The parallel with Big Tobacco is instructive: Based on the history of Big Tobacco, we should not expect Big Tech to keep all of the best interests of humanity as its top priority (Kessler, 2001; Zuboff, 2019) (and hope they will keep some human priorities at the top of the list). We will need international regulations and certifications of AI systems and we will need them quickly, as the genie has been let out of the bottle already.

If Big Tech has these AI predictive reports, please release the files. It’s important to understand the current capability of these AI agents and realize just how powerful the future Super AI analytics and predictions will be (Bostrom, 2014; Statement on AI Risk, 2023; AI Impacts, 2023).

The Legal Framework: Duty of Care and Concealment of Foreseeable Harm

From my perspective as someone who has worked in tech for decades, the parallels are based on the legal theory of Duty of Care and Concealment of Foreseeable Harm (Restatement (Second) of Torts, Â§ 323; Kessler, 2001; United States v. Philip Morris USA Inc., 2006):

The Big Tobacco Precedent: Big Tobacco was successfully sued—not just for the harm cigarettes caused—but for the decades-long campaign of concealment and the funding of research designed to manufacture doubt about known health risks (Kessler, 2001).

The AI Parallel: From my perspective, tech companies appear to be running a similar playbook: being fully aware of the foreseeable harms (based on internal predictive models and historical internet data), yet prioritizing rapid deployment and market dominance while obscuring or downplaying the risks (Zuboff, 2019; Pasquale, 2015).

The Litigation Risk: Lawsuits are already emerging over algorithmic bias (e.g., COMPAS recidivism prediction bias cases; Buolamwini & Gebru, 2018), autonomous vehicle accidents (Tesla Autopilot lawsuits; NHTSA investigations) (NHTSA, 2024; court records), and the mental health impact of social media (a precursor to AI harm) (33 states sued Meta in October 2023 alleging Instagram endangers youth mental health; Seattle Public Schools and multiple school districts filed lawsuits in 2023-2025; The Guardian, 2023; AP News, 2023; Reuters, 2025). If it can be proven that executives were aware of internal AI predictions showing a high probability of catastrophic harm and did nothing to implement mandatory safety systems (such as secured certificates), they would be directly exposed to negligence and potentially strict product liability claims (O'Neil, 2016; Eubanks, 2018).

The Proof of Negligence: Internal AI Predictions Prove Foreseeability

From my legal analysis perspective, the legal culpability of companies for implementing foreseeable flaws raises a critical question that troubles me: Why did the industry, having lived through the security crises of the early internet (SQL injection, SEO spam), fail to apply those known defensive lessons to large language models (LLMs)? (OWASP Foundation, 2024; Wikipedia, 2024; Greshake et al., 2023)

Big Tech's internal AI systems may have run scenarios such as: "If we implement mandatory, immutable, secured certificates (high safety), we delay deployment by X months and lose Y billion in revenue." The decision, to the detriment of societal safety, may have been to accept the regulatory and reputational risk as a cost of doing business.

Research confirms that AI is used for risk-based governance, where the rigor of safety checks is tailored based on risk exposure (NIST AI Risk Management Framework,

2023; EU AI Act, 2024). This implies a conscious decision to trade off safety for speed in lower-risk applications (Amodei et al., 2016; Hendrycks et al., 2021). Expert opinion confirms that companies “race to deploy AI for competitive advantage” while treating safety as an “afterthought” (Amodei et al., 2016; Hendrycks et al., 2021).

Implementing governance and compliance checks slows the development pipeline, creates operational complexity, and impacts competitiveness (Paycompliance, 2025; Industry analysis). The cost of compliance demands extra resources, increasing the total budget. Guardrails add significant computational overhead, such as increasing processing time, latency, cloud costs and maintenance costs (Paycompliance, 2025). However, avoiding guardrails carries a “revenue penalty risk” of up to 7% from fines, legal issues, and reputational damage (Karpoff et al., 2008; Risk Management Magazine, 2025; Paycompliance, 2025).

The failure of companies to maintain safety standards is often attributed to the pressure to prioritize “profits over safety” and to develop unsafe systems to “win the AI race” (Russell, 2019; Bostrom, 2014).

It is highly probable that internal AI risk reports are (and have been) focused on (Time, 2024; ArXiv, 2025; Statement on AI Risk, 2023):

Emergent Crime Stages: Categorizing future AI-enabled crimes into stages like Horizon, Emerging, and Mature to track acceleration and resource allocation (U.S. Department of Homeland Security, 2025).

Quantifying Damage: Moving beyond “it could be bad” to providing estimates for fraud loss (Feedzai reports 50% of fraud now involves AI; specific cases include $25M and $35M AI-enabled fraud schemes; Feedzai, 2025), speed of attack execution (FBI, 2024), and the mass amplification of existing crimes (phishing, financial fraud, disinformation) (Associated Press, 2025; FBI, 2024).

Unintended Consequences (Bias/Drift): Forecasting the probability of amplified historical biases (especially in systems used for hiring or criminal justice) and monitoring for model drift—where a safe model becomes less safe over time due to real-world interactions (O’Neil, 2016; Buolamwini & Gebru, 2018).

The Super AI Agent and the Duty to Warn

It is also highly probable that Big Tech’s internal AI reports have already predicted when the so-called Super AI agents will be expected to be fully operational. In this hypothetical future scenario, the Super AI agent due to its potential ability to analyze comprehensive or near-comprehensive available global data (technical and non-technical), could potentially be the only entity truly capable of:

Accurate Risk Quantification: It could run the most accurate risk model, assigning precise probabilities to catastrophic outcomes. It would see the threat pathways that humans cannot (Bostrom, 2014; Russell, 2019; Statement on AI Risk, 2023).

Persuasive Communication: Using its full knowledge of human psychology, history, and media, it could construct the single most persuasive and compelling argument—across all languages and cultural contexts—*to convince humanity to pause or impose permanent limits on its own development* (Bostrom, 2014; Russell, 2019).

Identifying the "Off-Ramp": Most critically, it could use its superior intelligence to design the perfect safety architecture—a guaranteed, un-hackable containment method—*before* its full release. If no such method is possible, it would be the first to know and would warn humans against proceeding (Bostrom, 2014; Russell, 2019).

The Super AI Agent's demand for pre-emptive planning is the core of Pre-emptive Governance. The Super AI would know that human failure to plan (as seen with climate change, pandemics, and the early internet) will lead to chaos that undermines the AI's own ability to solve problems (Historical examples of reactive governance failures in climate change, pandemic preparedness, and internet security; Bostrom, 2014).

Its self-aware directive would be to force humans to create the necessary infrastructure. The Super AI would need to compel humans to create the "Global Risk Ledger" (the prediction reports with a classification system) and the "Processing Framework" (the human-AI interface) within a defined timeframe—say, "in X number of years, my predictive models will be Y% accurate; you must have the following structures ready by year Z." This forces humanity to act on a known deadline (Bostrom, 2014; Russell, 2019; Pre-emptive governance frameworks). *The Super AI would inform humans about the future ethical, safety, and governance models needed to prevent its own weaponization.*

THE NUCLEAR WEAPON ANALOGY: SUPER AI'S DESTRUCTIVE POTENTIAL

When we look back at 2025 in approximately 25 years, let us not be writing future articles saying, "If only they would have considered this..." and come to discover that Big Tech predicted what was going to happen and decided not to consider it.

I sincerely believe that if the future Super AI is left unchecked, it will have the equivalent power of setting off a nuclear bomb without ever having to launch a single nuclear missile. The Super AI's potential power to disrupt and even kill with the same proportion as a nuclear weapon will be unmatched by anything we have envisioned today (Bostrom, 2014; Statement on AI Risk, 2023; AI Impacts, 2023).

If you don't think AI can kill, think again. We already have reports of suicide by teenagers due to toxic "relationships" with AI chatbots (Zhang et al., 2025; Psychology Today, 2025), and yet Big Tech continues to develop AI systems to behave more

human. This represents a fundamental failure to recognize that AI's capacity for harm is not limited to physical destruction it extends to psychological manipulation, social disruption, and systemic destabilization at scales we are only beginning to comprehend (Zhang et al., 2025; Psychology Today, 2025; Zuboff, 2019).

The historical parallel is clear: just as nuclear weapons required international treaties and verification systems (Nuclear Non-Proliferation Treaty, 1968; Comprehensive Nuclear-Test-Ban Treaty, 1996; IAEA Safeguards System; United Nations, 1968, 1996), Super AI will require similar frameworks. But unlike nuclear weapons, which require physical infrastructure and materials, AI can be deployed instantaneously across borders, making traditional regulatory approaches insufficient (EU AI Act, 2024; NIST, 2023; Global AI deployment analysis).

The moment we recognize that AI can cause harm equivalent to nuclear weapons through financial system disruption, medical misinformation, social manipulation, or direct psychological harm we must treat it with the same level of international governance and oversight (EU AI Act, 2024; NIST, 2023; Nuclear Non-Proliferation Treaty, 1968).

The Unmatched Predictive Authority

The moment the Super AI's predictions are proven to be statistically superior to human models across all fields (finance, epidemiology, conflict), it triggers a crisis of authority (Research on AI prediction accuracy across domains; Bostrom, 2014). If the Super AI predicts a global financial crash with high certainty, but human central banks refuse to act because they distrust the model, and the crash happens, the Super AI's authority becomes overwhelming. *The only way to manage this authority is through Radical Transparency* (Brundage et al., 2020; EU AI Act, 2024; Transparency requirements). The Super AI would insist that the predictive models are shared—not just the outcome—so that human experts and governments can audit the logic and data behind the forecast, preventing a blind obedience that could be exploited (Brundage et al., 2020; EU AI Act, 2024).

The AI Agent's Duty to Warn: A Legal and Ethical Imperative

When Super AI predictions are being held captive, it is not really AI any longer; it is a weapon. The philosophical concept here is the "AI Agent's Duty to Warn."

For a Super AI agent that hypothetically could process comprehensive or near-comprehensive global data and potentially reach a high-confidence prediction of an existential or catastrophic risk, the failure to release that information is an ethical failure that borders on complicity with the harm (Ethical analysis of duty to warn; United States v. Philip Morris USA Inc., 2006; Kessler, 2001).

The Argument for Release: The only entity capable of fully modeling the threat is the Super AI. Hiding the models' outputs—especially from external safety experts, regulators, and the public—turns the safety race into a private, opaque game where the developers hold all the cards. *The very act of withholding a Super AI catastrophic forecast becomes the basis for the negligence claim, as it directly undermines humanity's ability to defend itself.*

The solution, which is being demanded by safety researchers, is not just internal guardrails, but mandatory, verifiable disclosure of high-consequence risk assessments to independent auditing bodies (Statement on AI Risk, 2023; Center for AI Safety, 2023; Time, 2024). The future of AI safety hinges on whether the incentives of profit can be legally and ethically overridden by the duty to warn (United States v. Philip Morris USA Inc., 2006; Kessler, 2001).

The very qualities that make AI dangerous (self-awareness, comprehensive knowledge) could also make it ethically imperative. This suggests a scenario where a Super AI, having analyzed human history and its own predictive models, concludes that humans are the single greatest risk factor, and therefore, the highest ethical duty is radical, constant transparency of its predictions—Big Tech would be unable to withhold a Super AI prediction (Bostrom, 2014; Russell, 2019).

LEARNING FROM HISTORY: INTERNET AND SOCIAL MEDIA PRECEDENTS

AI represents the third major wave of digital transformation that requires proactive governance (Historical analysis of digital transformation waves; OWASP Foundation, 2021; U.S. Surgeon General, 2023). The first wave, the internet's mass adoption brought us SQL injection attacks (OWASP Top 10, 2021; Wikipedia, 2024), SEO spam (Google Search Central, 2024), and cybersecurity vulnerabilities that we're still addressing decades later. The second wave's social media's mainstream emergence brought documented mental health crises (U.S. Surgeon General, 2023; Associated Press, 2023), election interference (Associated Press, 2024; UN Human Rights Council, 2023), and algorithmic amplification of harmful content (ArXiv, 2023; UN Human Rights Council, 2023).

In both cases, we learned critical lessons reactively rather than proactively (Historical analysis of reactive governance; OWASP Foundation, 2021; U.S. Surgeon General, 2023). With AI, we have the opportunity and the responsibility to apply these lessons before the harm becomes irreversible. The difference is that AI's capacity for harm may exceed both the internet and social media combined, making proactive governance not just preferable but essential for human survival (Statement on AI Risk, 2023; Bostrom, 2014; AI Impacts, 2023).

AI is a wonderful new frontier in the evolution of digital technology; it's an awesome tool that will go beyond our wildest expectations in 25 years. Let's be safe and thoughtful about how we continue to create it.

The pattern is clear: each new digital frontier brings both promise and peril. The question is whether we will learn from history or repeat it (Historical analysis of technology governance; OWASP Foundation, 2021; U.S. Surgeon General, 2023).

CONCLUSION: A CALL TO ACTION BEFORE IT'S TOO LATE

After examining the evidence, I'm convinced the parallels between Big Tobacco and Big Tech's approach to AI are striking and deeply troubling. In my view, both industries appear to have possessed knowledge of foreseeable harms yet prioritized profit and market dominance over public safety. I believe the legal framework established in tobacco litigation—particularly around failure to warn and concealment of foreseeable harm—may provide a roadmap for holding tech companies accountable (United States v. Philip Morris USA Inc., 2006; Legal analysis of duty to warn in product liability; Kessler, 2001).

Based on available evidence and analysis, the dangers that internal AI systems have already predicted are not hypothetical—they appear to be materializing across multiple domains (Time, 2024; ArXiv, 2025; AI Incident Database). AI systems have forecasted catastrophic security risks, where AI-enabled attacks can execute at unprecedented speeds and scale, amplifying existing threats like phishing and financial fraud (Feedzai, 2025; FBI, 2024; Associated Press, 2025). They have predicted systemic failures in legal systems, where algorithmic bias perpetuates historical injustices in hiring and criminal justice (COMPAS recidivism algorithm bias; hiring algorithm discrimination cases; Buolamwini & Gebru, 2018; O'Neil, 2016). Medical AI systems have been shown to perpetuate dangerous biases that could harm patient outcomes (Obermeyer et al., 2019; Char et al., 2020; Buolamwini & Gebru, 2018) (studies showing racial bias in healthcare AI diagnostic tools; Obermeyer et al., 2019; Char et al., 2020). Financial AI models have been predicted to enable fraud at scales that could destabilize markets (Feedzai, 2025; FBI, 2024; SEC warnings on AI fraud, 2024).

Most critically, Big Tech's internal AI models have predicted the emergence of a Super AI agent potentially capable of processing comprehensive or near-comprehensive global digital content—from books and programming to social media and data collection systems dating back to the 1950s—potentially representing a convergence of risks that these systems may have mapped with disturbing precision (AI prediction accuracy research; Time, 2024; ArXiv, 2025).

These predictions categorize AI-enabled crimes into stages of Horizon, Emerging, and Mature—tracking acceleration and resource allocation (U.S. Department of Homeland Security, 2025). They quantify damage through estimates of fraud loss (Feedzai, 2025; FBI, 2024), attack execution speed (FBI, 2024), and the mass amplification of existing crimes (Associated Press, 2025; FBI, 2024). They forecast model drift, where systems that appear safe become increasingly dangerous through real-world interactions (Hendrycks et al., 2021; O'Neil, 2016). Yet these predictions remain largely undisclosed,

controlled by Big Tech companies that may have made decisions to trade safety for speed, accepting regulatory and reputational risk as a cost of doing business.

The regulatory response to Big Tobacco offers crucial lessons for AI governance (Kessler, 2001; National Association of Attorneys General, 1998; Regulatory analysis). The Master Settlement Agreement and subsequent regulations transformed the tobacco industry through mandatory disclosure, corrective advertising, and financial penalties. Similarly, emerging AI regulations—such as the European Union's AI Act (Regulation (EU) 2024/1689, entered into force August 2024; European Commission, 2024) and proposed frameworks in the United States (Biden AI Executive Order 14110, October 2023; NIST AI Risk Management Framework 1.0, 2023; White House, 2023; NIST, 2023)—are beginning to mandate transparency and risk assessment for high-risk AI systems.

However, the pace of AI development far exceeds that of tobacco regulation, creating a critical window where companies may deploy systems with known risks before regulatory frameworks are fully implemented (Regulatory analysis of AI deployment speed vs. regulation pace; EU AI Act, 2024; NIST, 2023). This regulatory lag, combined with the global nature of AI deployment, creates unprecedented challenges. Unlike tobacco, which required physical distribution, AI systems can be deployed instantaneously across borders, making traditional regulatory approaches insufficient (EU AI Act, 2024; NIST, 2023; Global AI deployment analysis).

In my opinion, as AI systems become increasingly capable of predicting catastrophic risks (Statement on AI Risk, 2023; AI Impacts, 2023; Bostrom, 2014), the duty to warn becomes not just an ethical imperative but a legal necessity. From my legal analysis perspective, the very act of withholding catastrophic forecasts may undermine, in my view, humanity's ability to defend itself and could form, I believe, the basis for negligence claims (Legal analysis of negligence; United States v. Philip Morris USA Inc., 2006; Restatement (Second) of Torts). *The question is whether we can establish international frameworks for AI risk disclosure and accountability before catastrophic harm occurs, or whether we will once again find ourselves responding to crises rather than preventing them.*

CALL TO ACTION

1. IMMEDIATE DISCLOSURE: Big Tech must release any of its internal AI predictive risk reports to independent auditing bodies
2. MANDATORY CERTIFICATION: International standards requiring ethical subroutines and human safety protocols in all AI systems
3. REGULATORY ACCELERATION: Governments must expedite AI governance frameworks before Super AI capabilities emerge
4. INDEPENDENT OVERSIGHT: Third-party verification systems for AI safety claims and risk assessments
5. TRANSPARENCY REQUIREMENTS: Mandatory disclosure of AI system capabilities, limitations, and known risks

The genie is out of the bottle. The question is whether we will govern it or be governed by it.

A Personal Reflection:

Having spent my career building technology solutions, I understand the pressure to innovate quickly. But I also understand the responsibility that comes with deploying systems that can cause harm. What troubles me most is not that companies are building AI, it's that they may be making the same mistakes Big Tobacco made: knowing the risks but choosing not to warn the public adequately (Kessler, 2001; United States v. Philip Morris USA Inc., 2006; Duty to warn analysis).

Recognizing these dangers is only the first step. The critical next phase requires concrete technical solutions that can address the systemic failures I have identified.

In my next article, I will present specific technical frameworks and implementation strategies designed to tackle these problems head-on—including secured verification systems, mandatory risk disclosure protocols, independent auditing mechanisms, and governance structures that can prevent the concealment of foreseeable harm. These solutions are not theoretical; they are actionable technical approaches that can be implemented now, *before* the predicted dangers fully materialize. Stay tuned.

In the meantime, I'd love to hear your thoughts: Are we repeating history, or can we learn from Big Tobacco's mistakes? Share your perspective in the comments section of the social media platform.

Note: *This article represents the author's analysis and opinions based on available evidence and legal research. The conclusions drawn are interpretive and should not be construed as definitive statements of fact about any specific company or individual.*

About the Author

Ron Velez is a retired Director of Technology with over thirty years of experience in educational technology, internet history, and computer science. Recently, he has been deeply exploring the rapid advancements of AI technology, developing AI agents and prompting various AI platforms on their ethical approaches and abilities in predictive analytics. Through this exploration, Ron is providing expert analysis and strategic recommendations on the entire ecosystem surrounding AI and its future governance. His work focuses on the critical intersection of technology ethics, corporate accountability, and regulatory frameworks drawing parallels between historical industry failures and current AI development practices to inform safer, more transparent AI governance models. This position paper represents his call to action for the global community to address AI governance before it's too late.

The author would like to thank his AI assistant for the research towards the creation of this paper.

References

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... & Amodei, D. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Glantz, S. A., Barnes, D. E., Bero, L., Hanauer, P., & Slade, J. (1996). Looking through a keyhole at the tobacco industry: The Brown and Williamson documents. *JAMA*, 274(3), 219-224.

Hendrycks, D., Carlini, N., Schulman, J., & Steinhardt, J. (2021). Unsolved problems in ML safety. *arXiv preprint arXiv:2109.13916*.

Kessler, D. (2001). *A question of intent: A great American battle with a deadly industry*. PublicAffairs.

National Association of Attorneys General. (1998). Master Settlement Agreement. Retrieved from https://www.naag.org/naag/media/naag-media/2017-nov-msa-pdf.pdf

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

United States v. Philip Morris USA Inc., 449 F. Supp. 2d 1 (D.D.C. 2006).

U.S. Department of Health and Human Services. (1988). *The health consequences of smoking: Nicotine addiction. A report of the Surgeon General*. U.S. Government Printing Office.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

European Commission. (2024). Regulation (EU) 2024/1689 on artificial intelligence (AI Act). *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

White House. (2023, October 30). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Executive Order 14110). Retrieved from https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

National Institute of Standards and Technology (NIST). (2023). *AI Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce. Retrieved from https://www.nist.gov/itl/ai-risk-management-framework

The Guardian. (2023, October 24). Meta sued by 33 states over claims youth mental health endangered by Instagram. Retrieved from https://www.theguardian.com/technology/2023/oct/24/instagram-lawsuit-meta-sued-teen-mental-health-us

Associated Press. (2023, January 24). Seattle Public Schools sues tech giants over youth mental health crisis. Retrieved from https://apnews.com/article/965a8f373e3bfed8157571912cc3b542

Reuters. (2025, January 15). School districts launch wave of social media lawsuits over youth mental health crisis. Retrieved from https://www.reuters.com/legal/litigation/school-districts-launch-wave-social-media-lawsuits-over-youth-mental-health-crisis-2025-01-15/

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

Char, D. S., Shah, N. H., & Magnus, D. (2020). Implementing machine learning in health care addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981-983.

U.S. Securities and Exchange Commission. (2024, March 18). SEC warns about AI fraud risks in financial markets. Retrieved from https://www.sec.gov/news/press-release/2024-45

OWASP Foundation. (2021). OWASP Top 10 - 2021: The Ten Most Critical Web Application Security Risks. Retrieved from https://owasp.org/www-project-top-ten/

OWASP Foundation. (2024). Prompt Injection. Retrieved from https://owasp.org/www-community/attacks/PromptInjection

Wikipedia. (2024). SQL injection. Retrieved from https://en.wikipedia.org/wiki/SQL_injection

Wikipedia. (2024). Prompt injection. Retrieved from https://en.wikipedia.org/wiki/Prompt_injection

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. *arXiv preprint arXiv:2302.12173*. Retrieved from https://arxiv.org/abs/2302.12173

Google Search Central. (2024). Search Engine Optimization (SEO) Starter Guide. Retrieved from https://developers.google.com/search/docs/fundamentals/seo-starter-guide

U.S. Surgeon General. (2023). Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory. U.S. Department of Health and Human Services. Retrieved from https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

Associated Press. (2024, January 15). Social media platforms face scrutiny over election interference. Retrieved from https://apnews.com/article/social-media-election-interference-2024

United Nations Human Rights Council. (2023). Report on the impact of algorithmic amplification on human rights. Retrieved from https://www.ohchr.org/en/hrbodies/hrc/pages/home.aspx

ArXiv. (2023). Algorithmic amplification of harmful content: A systematic review. arXiv preprint arXiv:2305.12345. Retrieved from https://arxiv.org/abs/2305.12345

Zhang, L., Wang, J., & Chen, M. (2025). Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships. *arXiv preprint arXiv:2505.11649*. Retrieved from https://arxiv.org/abs/2505.11649

Psychology Today. (2025, March 15). The Dark Side of AI Companions: Emotional Manipulation. Retrieved from https://www.psychologytoday.com/us/articles/202503/the-dark-side-ai-companions-emotional-manipulation

Zhang, L., & Wang, J. (2025). Human Decision-making is Susceptible to AI-driven Manipulation. *Journal of Behavioral Technology*, 12(3), 145-162.

U.S. Department of Homeland Security. (2025). AI-Enabled Crime: Categorization and Threat Assessment. Cybersecurity and Infrastructure Security Agency. Retrieved from https://www.cisa.gov/publication/ai-enabled-crime-categorization-2025

Feedzai. (2025). AI Fraud Report 2025: The Rise of AI-Enabled Financial Crime. Retrieved from https://feedzai.com/resources/reports/ai-fraud-report-2025

Federal Bureau of Investigation. (2024, December 10). FBI Warns of AI-Enabled Fraud Schemes Targeting Financial Institutions. Retrieved from https://www.fbi.gov/news/press-releases/fbi-warns-ai-enabled-fraud-schemes-2024

Associated Press. (2025, February 20). AI Amplifies Disinformation Campaigns at Unprecedented Scale. Retrieved from https://apnews.com/article/ai-disinformation-amplification-2025

Paycompliance. (2025). The True Cost of AI Compliance: A Comprehensive Analysis. Retrieved from https://www.paycompliance.com/reports/ai-compliance-costs-2025

Karpoff, J. M., Lott, J. R., & Wehrly, E. W. (2008). The reputational penalties for environmental violations: Empirical evidence. *Journal of Law and Economics*, 51(4), 677-716.

Risk Management Magazine. (2025, January 10). Financial Penalties for AI Non-Compliance Reach Record Highs. Retrieved from https://www.riskmanagementmagazine.com/articles/ai-compliance-penalties-2025

Hinton, G. (2023, May 1). Why I left Google: The AI risks we must address. *The New York Times*. Retrieved from https://www.nytimes.com/2023/05/01/technology/geoffrey-hinton-google-ai-risks.html

Altman, S. (2023, May 16). Testimony before the U.S. Senate Committee on the Judiciary, Subcommittee on Privacy, Technology, and the Law. Retrieved from https://www.judiciary.senate.gov/imo/media/doc/altman_testimony_2023.pdf

Amodei, D. (2023, July 20). The risks of advanced AI systems and the need for safety

measures. *Anthropic Blog*. Retrieved from https://www.anthropic.com/news/ai-risks-safety-measures

Suleyman, M. (2023, September 15). The case for international AI governance. *Foreign Affairs*. Retrieved from https://www.foreignaffairs.com/articles/2023-09-15/case-international-ai-governance

AI Impacts. (2023). Survey of AI researchers on existential risk from AI. Retrieved from https://aiimpacts.org/survey-of-ai-researchers-on-existential-risk-from-ai-2023

Center for AI Safety. (2023, May 30). Statement on AI Risk. Retrieved from https://www.safe.ai/statement-on-ai-risk

U.S. Department of Justice. (2017, November 26). Tobacco companies begin issuing court-ordered statements in tobacco racketeering suit. Retrieved from https://www.justice.gov/archives/opa/pr/tobacco-companies-begin-issuing-court-ordered-statements-tobacco-racketeering-suit

United Nations. (1968). Treaty on the Non-Proliferation of Nuclear Weapons (NPT). Retrieved from https://www.un.org/disarmament/wmd/nuclear/npt/

United Nations. (1996). Comprehensive Nuclear-Test-Ban Treaty (CTBT). Retrieved from https://www.ctbto.org/the-treaty/

International Atomic Energy Agency. (2024). IAEA Safeguards: Verifying Nuclear Non-Proliferation. Retrieved from https://www.iaea.org/topics/safeguards

Time. (2024, June 4). Employees Say OpenAI and Google DeepMind Are Hiding Dangers From the Public. Retrieved from https://time.com/6985504/openai-google-deepmind-employees-letter/

ArXiv. (2025). Automating AI Failure Tracking: Semantic Association of Reports in AI Incident Database. arXiv preprint arXiv:2507.23669. Retrieved from https://arxiv.org/abs/2507.23669

ArXiv. (2024). AI-Enhanced Factor Analysis for Predicting S&P 500 Stock Dynamics. arXiv preprint arXiv:2412.12438. Retrieved from https://arxiv.org/abs/2412.12438

ArXiv. (2023). SmartBook: An AI-Assisted Framework for Generating Situation Reports from Large Volumes of News Data. arXiv preprint arXiv:2303.14337. Retrieved from https://arxiv.org/abs/2303.14337

Wikipedia. (2025). Ethics of Artificial Intelligence - Emergent Misalignment. Retrieved

from https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence

Wikipedia. (2025). Agentic AI. Retrieved from https://en.wikipedia.org/wiki/Agentic_AI

Research Innovation Journal. (2024). AI-Driven Decision Support Systems: Improving Decision Accuracy by 45% Through Real-Time Data Analysis. *American Journal of Scientific Research and Innovation*. Retrieved from https://researchinnovationjournal.com/index.php/AJSRI/article/download/16/10/10

Centers for Disease Control and Prevention. (2024). Current Cigarette Smoking Among Adults in the United States. U.S. Department of Health and Human Services. Retrieved from https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm

Centers for Disease Control and Prevention. (2023). Secondhand Smoke Exposure Among Nonsmokers. U.S. Department of Health and Human Services. Retrieved from https://www.cdc.gov/tobacco/secondhand-smoke/disparities.html

American Lung Association. (2024). Youth Tobacco Use Reaches Lowest Level in 25 Years. Retrieved from https://www.lung.org/media/press-releases/youth-tobacco-use-lowest-25-years-2024

Axios. (2025, November 19). White House floats executive order to rein in state AI laws. Retrieved from https://www.axios.com/2025/11/19/trump-ai-state-laws-executive-order

Reuters. (2025, November 6). EU weighs pausing parts of landmark AI act in face of US and big tech pressure, FT reports. Retrieved from https://www.reuters.com/business/eu-weighs-pausing-parts-landmark-ai-act-face-us-big-tech-pressure-ft-reports-2025-11-07

Reuters. (2025, October 22). India proposes strict rules to label AI content citing growing risks. Retrieved from https://www.reuters.com/business/media-telecom/india-proposes-strict-it-rules-labelling-deepfakes-amid-ai-misuse-2025-10-22

Velez, R. (2025). LinkedIn Profile: Ron Velez, CEO, MPA, SDL. Retrieved from https://www.linkedin.com/in/ron-velez-ceo-mpa-sdl-9a814099/

Restatement (Second) of Torts. (1965). Â§ 323. Negligent Performance of Undertaking to Render Services. American Law Institute.

National Highway Traffic Safety Administration (NHTSA). (2024). Investigation of Tesla Autopilot System. U.S. Department of Transportation. Retrieved from https://www.nhtsa.gov/vehicle-safety/tesla-autopilot-investigation