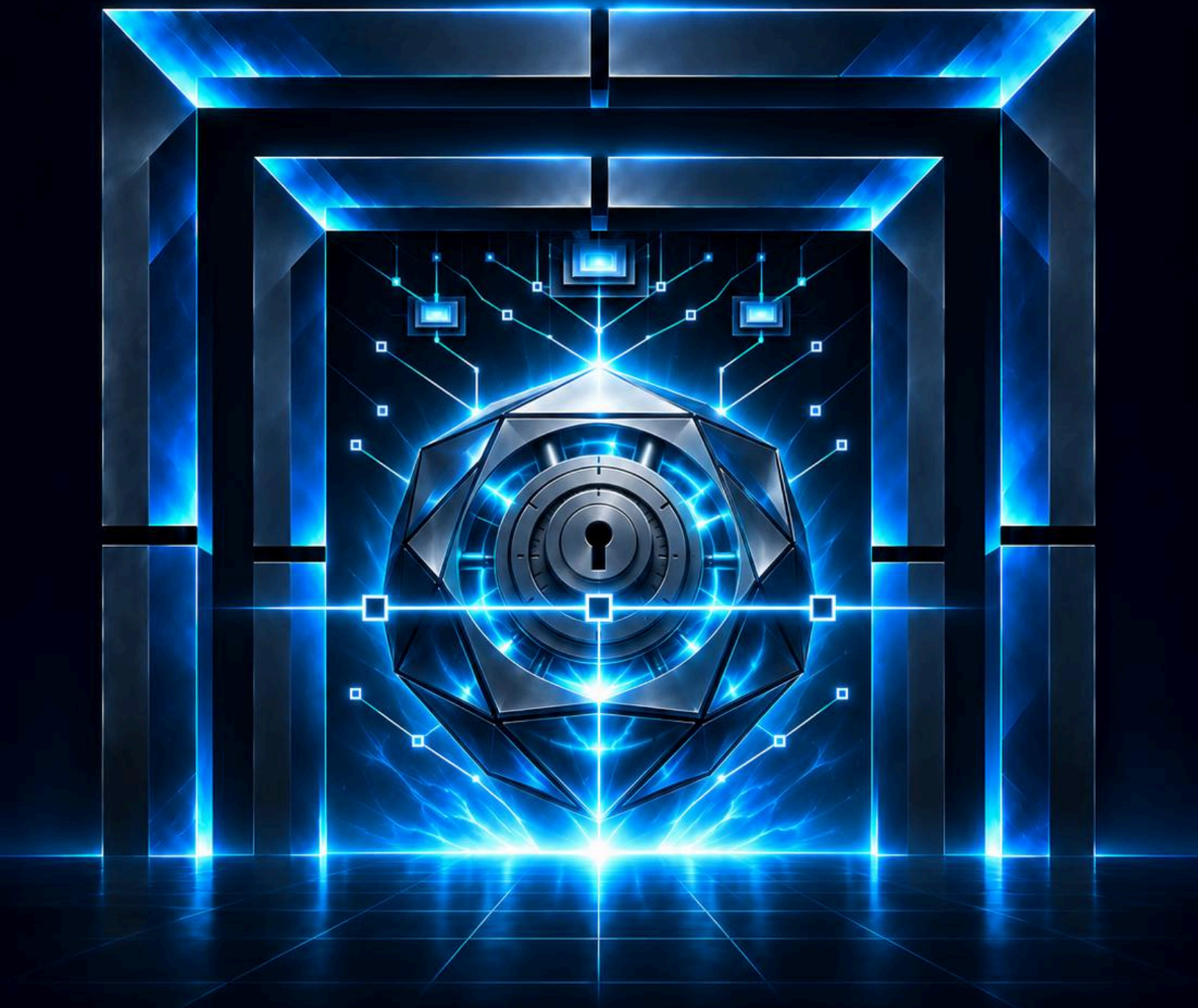
THE M.I.A.M. DOCTRINE

MUTUAL INHIBITION OF AI-MEDIATED ESCALATION

A SOUVENIR INTELLIGENCE STRATEGY



TONY T. NGUYEN | NGUYỄN TỪ VĨNH TRƯỜNG

HIGHLAND REFLEXIVE PRESS

The M.I.A.M Doctrine

Mutual Inhibition of AI-Mediated Escalation

Tony T. Nguyen | Nguyễn Từ Vĩnh Trường · Highland Reflexive Press

Summary

In 2025, Dan Hendrycks, Eric Schmidt and Alexandr Wang set out a strategy for the age of superintelligence built on three pillars: deterrence through mutual sabotage, which they call MAIM; nonproliferation; and competitiveness, including the integration of artificial intelligence into military command and control (Hendrycks, Schmidt & Wang, 2025). The M.I.A.M doctrine follows the same architecture, chapter for chapter, and sets a counter-pillar against each of theirs.

1. **Mutual inhibition in place of mutual sabotage.** The default is not readiness to strike but zero: not a single weapon fired, not even in retaliation.
2. **No delegation to machines in place of nonproliferation.** What must be kept out of the wrong hands is not only the weapon but the decision itself, which must never slip from human hands into a machine's.
3. **The strength of slowness in place of the race for speed.** The safest nation is not the fastest one, but the one whose structures force it to slow down before the button.

Both doctrines speak of zero, but they mean different zeros. MAIM's is the zero of paralysis: no one can pull ahead, because any project that rises will be crippled by its rivals. M.I.A.M's is the zero of nothing happening: the button is simply never pressed. The root

of that second zero lies in one of the oldest ideas in Buddhism, the teaching of non-self. When there is no longer a "self" to defend or avenge, no one needs the button.

MAIM seeks to stop any one nation from monopolizing superintelligence. M.I.A.M ensures that the effort to stop it, or a false alarm about it, never becomes the last day of the world.

Opening

Some nuclear warheads are born to liven up the offices of heads of state; others are born too, but to shoulder an entirely different role. What that role might be, dear reader, must wait for the fifth scenario, where all shall be revealed in due course.

And this season, in the very midst of those lively offices, whispers began to circulate that somewhere in the remote West Texas desert, a person who files fingernails by day had dared to set pen to paper on a doctrine of the nuclear button. More than a few learned eyebrows rose with an air of condescension, the familiar disdain that greets any outsider who wanders by mistake into a ballroom reserved for those in academic evening dress.

This essay begs leave to return the courtesy with the bare minimum of politeness: since when did the smell of acetone and the nail file strip a person of the standing to worry about the fate of humankind? Is it only the bewigged minds of the reception hall who may concern themselves with the moment the earth closes in one dry blast, a moment after which no one will survive to take notes, to peer-review, or to argue over who was right?

The elite may exhaust themselves debating whose powerful finger will press that fateful button; the writer of these lines does not. For an enraged head of state and a mindless machine in a fit of delusion, when all is said and done, bestow exactly the same favour.

The only thing the wise should engrave upon their hearts is this: once the button sinks, all of humanity will walk hand in hand into nothingness, in an extinction of perfect equality. There, no living soul will remain to assign blame, and, fortunately, no editorial board will stay up past midnight to send out one more courteous letter of rejection: *Regrettably*!!!

The M.I.A.M doctrine pursues a single outcome: however aggressive anyone becomes, the number of nuclear weapons fired must remain zero. This is not one goal among many to be traded against others, because in the nuclear world a single detonation is enough to set off the whole chain, and afterwards no institution will be left intact enough to learn the lesson.

Throughout the nuclear age, humanity was saved not by precise machines but by people who doubted the machines and still had time to hesitate. Artificial intelligence is taking away both at once: it makes warnings look more credible, and it makes the time to verify them shorter, or erases it altogether. The great powers have begun to say that humans must hold the final decision, but no one has yet said under what conditions those representatives must deliberate, or whether their decision may seal the fate of the whole planet, of eight billion people and every living species.

This doctrine answers that question. It consists of one central proposition, six principles and an architecture of three locks, built from lessons that history paid for with near-losses of everything.

Part One. The Button Is Humanity's Security Problem

Artificial intelligence has become a matter of national security. M.I.A.M goes one step further: the nuclear button is a matter of human security, because no nation survives alone once it is pressed.

Hendrycks and his colleagues divide the danger into four domains: strategic competition, terrorism, loss of control, and the strategies already on the table. This part walks through the same four domains, but looks at each from the button's side: the times the world nearly lost control, the four routes by which AI pushes people toward escalation, the limits of classical deterrence, and the doctrines that came before.

Three Times the World Was Saved by a Human Who Hesitated

Even before artificial intelligence, the moments when humanity stood closest to the edge shared one structure: a warning that looked real, a compressed window of time, and one particular person who refused to do what the system was pushing them toward.

27 October 1962, submarine B-59. Picture a Soviet submarine lying beneath the Caribbean at the height of the Cuban Missile Crisis, its batteries nearly drained, its cooling broken, the heat in the compartments so stifling that sailors fainted one after another, and no word from Moscow for days. Above them, American destroyers kept dropping depth charges, the blasts ringing against the steel hull like hammers on a cooking pot, and no one aboard knew whether these were signalling charges meant to force them up or whether the war had already begun. In the belly of the boat lay a nuclear torpedo the Americans knew nothing about, and launching it required all three senior officers to agree. The captain, by the account of those present, wanted to fire and was ready to die along with the enemy; the political officer sided with him; only one signature was missing. Vasily Arkhipov did not sign. He persuaded the boat to surface amid the encirclement and await orders, and when the hatch opened, what greeted them was the air of a world still intact, not the shadow of a mushroom cloud.

The submarine was not the only fuse burning that day. Washington would later call 27 October 1962 Black Saturday: American strategic forces were already at DEFCON 2, a U-2 reconnaissance plane was shot down over Cuba, and another U-2 strayed into Soviet airspace, sending fighters aloft on both sides. Down at the bottom of the sea, the men of B-59 had no way of knowing that the Americans were dropping practice depth charges as a signal to surface, and no way of asking Moscow. The weapon aboard was a T-5 torpedo with a nuclear warhead, often described as comparable in yield to the Hiroshima bomb. Much of what we know about the air inside that compartment comes from the account of signals-intelligence officer Vadim Orlov, published in 2002; other veterans have told it more mildly, so the most dramatic details should be read as one man's testimony rather than as a record.

What matters most in the story is not Arkhipov's character but a rule: firing the torpedo required the agreement of three men, and a single refusal was enough. Yet the rule saved the world only by chance. Arkhipov was chief of staff of the whole submarine flotilla and happened to be sailing on B-59. On another boat, the third seat might have belonged to someone more willing to nod. M.I.A.M draws a simple lesson from this: the right to say no cannot depend on who happens to be in which chair. It has to be built into the structure, in places where the interests of the person holding the key do not coincide with the interests of the person who wants to press.

26 September 1983, bunker Serpukhov-15. Just past midnight, the alarm in the command bunker began to wail, the big screen lit up in red, and the Oko early-warning satellite system reported that the United States had launched one missile, then two, then three, then five ballistic missiles heading for the Soviet Union. The duty officer, Stanislav Petrov, sat amid the sirens with a telephone in his hand, knowing that if he reported an attack, the retaliation machine would begin to turn and no one would be able to stop it in time. He reported a system malfunction. He reasoned that no one starts a nuclear war with five missiles, that the system was new and not yet trustworthy, and that ground radar had not confirmed anything, but he could not prove he was right, and the minutes that followed were minutes in which one human being had to sit and wait to learn whether the earth

would still be there. Nothing fell. Later investigation found that the satellite had mistaken sunlight reflecting off clouds for the flare of missile engines.

The autumn of 1983 was the worst possible season for a false alarm. Three weeks earlier, Soviet air defences had shot down Korean Air Lines Flight 007, killing all 269 people on board, in the belief that it was an American spy plane. A few weeks later, NATO would hold Able Archer 83, an exercise rehearsing nuclear release procedures, which some in the Soviet leadership feared was cover for a real attack. Into that atmosphere of suspicion, the newly commissioned Oko satellite system reported five Minuteman missiles. The cause was later traced to a satellite that had mistaken sunlight glinting off high clouds for the flare of a rising missile.

Notice how Petrov reasoned. He had no better data than the machine; what he had was a place to stand. He knew the system was immature, he knew ground radar had not confirmed anything, and he knew a real first strike would not open with five missiles. An AI system today would do Oko's job far better, but it would also do something Oko could not: sound more certain. It would attach a high confidence score to those five specks of light, add a fluent explanation, and perhaps propose a response. What saved the world in 1983 was a human who doubted a clumsy machine. What should frighten us now is that the machines are no longer clumsy. The story has a coda worth pausing on. The incident was kept secret for years, and Petrov himself recounted that investigators tried to pin the blame on people rather than admit the hardware had been rushed into service. A system that punishes the person who doubts the machine is teaching the next Petrov not to doubt. That is why Principle 6 records the responsibility of whoever chooses, but never punishes anyone for choosing to stop.

25 January 1995, the nuclear briefcase. The Cold War had been over for three years, and people believed fear had been packed away in museums, when one morning a Norwegian scientific rocket left its launch pad and Russian radar saw something that looked exactly like a missile fired from an American submarine to detonate high in the atmosphere and blind radar ahead of a real attack. Norway had given notice of the launch, but that notice

never reached the people watching the screens. For the only time publicly known in history, the nuclear briefcase was activated and placed before President Yeltsin, and a man and his generals faced a decision with only minutes to make it. What saved the world that day was nobody's brilliance but a few minutes of tracking the trajectory, enough to see that the dot on the screen was heading out to sea rather than towards Moscow, and the briefcase was closed.

The rocket was a four-stage Black Brant XII, launched from Andøya to study the aurora. Early-warning radar at Olenegorsk picked it up, and to the officers on duty its speed and trajectory resembled a Trident missile launched from an American submarine; as its stages separated, the radar picture could even look like a missile releasing several warheads. Norway had followed every procedure: notice had been sent to the Russian Foreign Ministry well in advance, but it stopped there and never reached the military. According to a U.S. European Command history, Russian doctrine at the time gave the president roughly ten minutes from detection to decision. The rocket flew for about twenty-four minutes and fell into the sea near Spitsbergen, exactly as planned. The next day Yeltsin himself told reporters that he had, for the first time, used the small briefcase with the button that always travelled with him.

If B-59 is a lesson about rules and Petrov a lesson about doubt, 1995 is a lesson about dialogue. No one in this story meant harm. The Norwegian scientists gave notice, the radar crews did their jobs, and the briefcase was opened by the book. What failed was a line of communication cut halfway: a letter sent that no one confirmed receiving. That is why M.I.A.M does not accept a one-way notice as evidence of dialogue, and requires every turn to be acknowledged by the other side (Principle 2). Ten minutes were enough in 1995, because it was a scientific rocket heading out to sea on a trajectory anyone could read. Nothing guarantees that the next time will be so clear.

The three stories leave three lessons that this doctrine turns into principles. The first is that machines can be wrong in very convincing ways. The second is that time itself is what saved the world: the minutes Petrov waited for confirmation and the minutes

Moscow spent tracking a trajectory. The third is that a rule requiring several people to agree, as on B-59, gave one person the power to say no. Because the survival of humanity once depended on the luck and character of a few individuals, this doctrine demands that their hesitation be turned into structure, so that next time we do not have to count on luck.

And before all three, a lock refused to open, on Vietnamese soil. In the spring of 1954, as French forces were besieged at Điện Biên Phủ, the United States weighed an air intervention known as Operation Vulture. Its strongest advocate was Admiral Arthur Radford, chairman of the Joint Chiefs of Staff, who drew on a study suggesting that three tactical atomic bombs would be enough to break the siege. President Eisenhower did not reject the idea outright; he set conditions. Congress would have to approve in advance, and Britain would have to take part. Britain declined, Congress would not back a unilateral intervention, and on 7 May 1954 the garrison fell without a single bomb being dropped.

This story differs from the other three in one important way. On B-59, at Serpukhov-15 and in Moscow in 1995, the world was saved by a person who hesitated. At Điện Biên Phủ it was saved by a structure: a nuclear decision was forced to pass through parties with different interests, and one refusal was enough. That is M.I.A.M's second layer of locks, operating before anyone had a name for it. And it operated on the very land the author of this doctrine calls home. The people who nearly became the target of three bombs in 1954 now have a son who has written this doctrine in the hope that no other people will ever face that prospect.

Diagnosis: Four Routes by Which AI Pushes People Toward Escalation

The danger is not that AI will press the button itself, which the great powers have already declared they will not allow. The danger is that AI stands between reality and the human being, shaping what people see, what they believe and what they have time to weigh. As one retired American general has warned, the risks of compressing nuclear decision time likely outweigh the benefits. This doctrine identifies four specific pathways.

1. **Compressed time.** Under launch-on-warning, decision-makers may have less than 20 minutes. Every improvement in detection speed creates pressure to act faster, and is rarely used to give humans more time.
2. **False confidence.** A warning accompanied by a probability looks more scientific than a person's word. Automation bias leads people to trust the machine more than their own doubt, the very kind of doubt that saved Petrov.
3. **Framed choices.** When the machine offers only one option, or three options of which two are mere decoration, the person who signs still carries the title of decision-maker, but the real decision was made beforehand.
4. **Reaction chains between two sides.** When both sides use machines to read each other, a defensive move by one can be read by the other's machine as a sign of attack, and the other way round. Escalation can form out of two machines reacting to each other before any human being actually wants war.

The Gulf of Tonkin, August 1964, shows how far the first of these routes can run, even without artificial intelligence. On the night of 4 August, two American destroyers reported a torpedo-boat attack on the strength of sonar and radar contacts in heavy weather. The task-group commander, John Herrick, cabled his own doubts about many of the reported contacts and recommended a search for wreckage the next morning. Three days later, Congress passed a resolution authorizing a wider war. In 2005, the National Security Agency declassified a study by its historian Robert Hanyok confirming that no second attack had taken place and that the signals intelligence had been presented in a skewed way (Hanyok, 2001). No nuclear weapon was involved, but there is no clearer lesson in how ambiguous evidence, travelling faster than doubt, can become the foundation of a war that lasts more than a decade.

These four pathways have one thing in common: none of them is blocked by the sentence "humans must decide". They are blocked only when the conditions of the decision are clearly specified and enforced by structure.

Rereading Hendrycks's Chapter Three from the Button's Side

Chapter three of "Superintelligence Strategy" argues that artificial intelligence threatens national security along three lines: rivalry between great powers, terrorism, and loss of control. The M.I.A.M doctrine accepts all three diagnoses. What it rejects is the cure, and the disagreement shows most clearly when each diagnosis is set beside the nuclear button.

Strategic competition. Hendrycks and his co-authors worry that artificial intelligence could produce superweapons that unsettle nuclear deterrence, and that the fear of being overtaken could push leaders toward preventive action (Hendrycks, Schmidt & Wang, 2025, section 3.1.2). The diagnosis is right, but it points to something the paper does not follow through: the danger lies less in the superweapon than in the clock. A leader who believes a rival is about to gain a decisive edge feels time running against him, and a leader who feels time running against him is the one most likely to strike first. MAIM answers that fear with another fear, the fear of sabotage. M.I.A.M answers it by giving time back, so that no decision may ever be born of the sense that time is running out.

Terrorism and attacks on infrastructure. The paper observes that AI-assisted cyberattacks will be harder to attribute, and that in the balance between offence and defence an attacker needs only one hole while a defender must patch them all (sections 3.2.2 and 3.2.3). If both points hold, the logical consequence is that no decision to retaliate may rest on a machine's attribution of blame, because attribution itself has become unreliable. M.I.A.M also reverses the offence-defence balance at the one place where it matters most. In a system whose default is to halt, an attacker who finds a single hole gains only a halt; to obtain a launch decision, he must open every lock at once, and the locks are held by parties whose interests are opposed. At the button, it is the attacker who must win everywhere.

Loss of control. Here the two doctrines come closest. Hendrycks and his colleagues describe with real precision how humans lose control: not through some dramatic machine coup, but through a chain of small decisions, each reasonable on its own terms, until peo-

ple can no longer steer the decisions that matter (section 3.3.1). They also conclude that control must be an ongoing process rather than a one-time fix (section 3.3.3). M.I.A.M agrees entirely and asks only one question. If the erosion of control proceeds by small, reasonable steps, is the integration of artificial intelligence into command and control, which the same paper recommends in chapter six, one of those small, reasonable steps? Chapter three answered that question before chapter six could ask it. M.I.A.M is exactly the ongoing process of control that chapter three calls for, placed at the one point where control, once lost, can never be regained. It may sound abstract, but that chain of small steps is not hypothetical. It has happened before with nuclear weapons themselves, and it happened on Vietnamese soil, at a base called Khe Sanh.

Khe Sanh proves one thing: a nuclear weapon can move close to use without anyone deciding to use it, through nothing more than a chain of small preparatory steps, each of which looks sensible. That is precisely the kind of loss of control Hendrycks and his coauthors warn about in chapter three, and it really happened, in Vietnam. In early 1968 some six thousand U.S. Marines were besieged at Khe Sanh, a remote plateau near the seventeenth parallel and the Laotian border. The American press compared it with Điện Biên Phủ in 1954, and the fear among commanders was of the same defeat: an isolated garrison, shelled day and night, about to fall.

In that atmosphere General William Westmoreland, the U.S. commander in Vietnam, approved a contingency plan code-named Fracture Jaw. The plan did not order an atomic strike. It only prepared to bring nuclear weapons into South Vietnam so that, if Khe Sanh were about to fall, they could be used at short notice. Step by step the plan advanced quite far within Pacific Command: options were drawn up, transport was discussed, methods of deployment were chosen. Only when word reached the White House did President Johnson order it stopped. Historians still debate how much Washington knew in advance, since Washington itself had at one point asked about the possibility of using tactical nuclear weapons.

What is frightening about Khe Sanh is not a mad general. No one in the story decided to use the bomb. Each person did something that seemed reasonable within his own remit: a contingency plan, a preparatory step, a logistical option. Added together, those small steps brought nuclear weapons closer to a battlefield than ever before, before the person with the highest authority knew about it. That is the loss of control Hendrycks and his colleagues describe, and it is why M.I.A.M places its lock on the final decision itself rather than hoping someone notices in time along the way.

The Zero of Classical Deterrence

Ever since Bernard Brodie wrote that the chief purpose of military establishments would henceforth be to avert wars rather than win them (Brodie, 1946), deterrence has rested on a promise: since whoever fires first will be destroyed, no one will fire. Thomas Schelling turned that promise into a science of coercion (Schelling, 1960; 1966), and in fairness the doctrine has a record that is hard to dispute: since 1945, not a single nuclear weapon has been used in war.

But whose zero is that? As early as 1959 Albert Wohlstetter warned that the balance of terror was far more delicate than people assumed (Wohlstetter, 1959), and Scott Sagan later went through the records to show that the very systems built to deter had repeatedly come close to causing accidents (Sagan, 1993). The three stories of Part One are three pages from that ledger. Aboard B-59, in the Serpukhov-15 bunker and before the briefcase of 1995, deterrence did nothing to keep the zero; on the contrary, its own logic, retaliate in time or be destroyed, is what pushed those people toward the button. Deterrence deserves credit for keeping the great powers out of a major war for decades, but on the nights closest to the button, the zero was kept by luck and by a handful of people willing to hesitate against that very logic.

> **For nearly eighty years deterrence has been credited with the fact that the world has not blown up. Perhaps deterrence did prevent deliberate wars, in which a leader calculates and chooses to strike; no one can prove it, and no one can disprove it. But**

> **there is one kind of danger against which deterrence is wholly powerless: war by mistake. Against that danger the logic of deterrence, fire back in time or be destroyed, is precisely what drives people toward the button. And in every near-miss by mistake that the public record still preserves, what kept the count at zero was never deterrence. It was a human being willing to hesitate against that logic.**

Five times nuclear weapons came close to use, and what stopped them

Each time, what stopped it was a person or a lock held by several parties

Year	Event	What stopped it
1954	**Điện Biên Phủ** Proposal to use three tactical atomic bombs	**A lock held by several parties** Britain and Congress would not agree
1962	**Submarine B-59** The captain wanted to fire a nuclear torpedo	**One officer would not sign** Vasily Arkhipov
1968	**Khe Sanh** A plan to move nuclear weapons into the theatre	**A stop order from the very top** President Johnson
1983	**Serpukhov-15** A satellite reported five US missiles	**An officer who doubted the machine** Stanislav Petrov
1995	**Norwegian rocket** The nuclear briefcase opened before the president	**Minutes of tracking the trajectory** Boris Yeltsin

five times nuclear weapons came close to use, 1954–1995

Set those five occasions side by side, and not one line in the right-hand column reads "deterrence".

Those five are only the best known. The declassified record holds many more, and a 2014 study by Chatham House counted thirteen occasions between 1962 and 2002 on which nuclear weapons came close to use through miscalculation, miscommunication or technical error (Lewis et al., 2014). The table below keeps only the cases in which a warning system reported an attack that was not happening.

When and where	What the system reported	What was really there	What stopped it
5 October 1960, early-warning radar at Thule, Greenland	A massive Soviet missile attack, with 99.9 percent certainty	The rising moon reflecting the radar beam	People doubted and checked rather than act on the warning
24 November 1961, U.S. Strategic Air Command	Every link to NORAD and the three early-warning radar sites went dead at once, a possible sign of attack	A single failed relay station	A B-52 on airborne alert radioed Thule and confirmed that no attack had taken place
25 October 1962, Duluth and Volk Field	A sabotage alarm; at Volk Field the klaxon instead ordered nuclear-armed F-106A interceptors to scramble	A bear climbing a fence, plus a miswired alarm (Sagan, 1993)	An officer phoned Duluth to check, and a vehicle raced onto the runway to stop the aircraft
9 November 1979, NORAD	A full-scale Soviet nuclear attack	An exercise tape loaded by mistake into the live warning system	Cross-checking with satellites and radar within minutes, just before the national security adviser was about to call the president
3 and 6 June 1980, NORAD	A large incoming strike	A failed 46-cent computer chip and faulty message design	Warning centres compared several independent sources within minutes
26 September 1983, Oko satellites	Five Minuteman missiles	Sunlight glinting off clouds	A duty officer who did not trust the machine: Stanislav Petrov
25 January 1995, Olenegorsk radar	A Trident missile launched from a submarine	A Norwegian rocket studying the aurora	A few minutes spent tracking the trajectory

Read down the third column and something almost comic appears: the moon, a bear, a tape, a 46-cent chip, a glint of sunlight, a rocket sent up to measure the northern lights. There is not a single enemy in that column. Read down the fourth and something more important appears: not once was the world saved by deterrence, and not once by speed. Every time it was a person willing to doubt, a second source to check against, and a few minutes in which to do it. Those are exactly the three things that Principle 1, Principle 3 and the shared early-warning network set out to turn into structure. And remember the figure from 1960: 99.9 percent. The first machine of the missile age was already speaking in probabilities long before artificial intelligence existed, and it was almost perfectly con-

fident about an attack that was not there. The Chatham House study reaches the same conclusion: on several occasions it was individual judgement, sometimes against protocol itself, that kept the weapons from being used.

M.I.A.M takes up that same number but changes its owner. Under deterrence, zero is a hoped-for outcome that survives as long as nobody does anything foolish; under M.I.A.M, zero is an invariant built into structure, and it survives because the locks never open together. Deterrence asks, "Who dares to fire first?" M.I.A.M asks, "Who has the power to open every lock and send the whole salvo up?" The answer is: no one.

Other Doctrines That Came Before

Launch on warning. This is where M.I.A.M stands in direct opposition. Launch on warning turns a warning into a reason to act within minutes; M.I.A.M turns a warning into a reason to wait longer and negotiate more. All three near-catastrophes of Part One were moments when, had launch on warning been applied mechanically, the world would no longer exist.

The two-person rule. The rule that two or three officers must agree, as aboard B-59, is M.I.A.M's direct ancestor. M.I.A.M extends the idea from the inside of a single boat to three levels: among those holding authority within one state, between two states in confrontation, and between states and humanity.

Commitments to human control. The 2024 joint statement by the United States and China, in the words of one retired American general, reflects a growing international consensus that artificial intelligence must not replace human judgement in nuclear decisions. Yet analysts also warn that even with a human holding the final decision, this may still not be enough to keep an AI error from ending in catastrophe. M.I.A.M is one answer to that gap: it turns a commitment into conditions that can be verified.

Part Two. The First Pillar: Mutual Inhibition

The first pillar answers the same question MAIM answers with a threat: how do we make sure no one presses the button?

The Central Proposition: The Zero-Launch Invariant

> In every system operating under M.I.A.M, the number of launch decisions validated as legitimate is always zero unless every lock opens at the same moment; and those locks are entrusted to parties whose interests are so opposed that they never open together.

The proposition has two halves. The first is a logical condition: if any lock remains closed, whether through refusal, absence, lost contact or failed verification, the result is zero. The second is a deliberate design choice: the keyholders are chosen so that the condition "all open at once" practically never arises.

This doctrine does not try to stop anyone from growing angry, issuing threats or behaving aggressively; that is the business of politics and of conscience. What it does is close, one by one, every road by which anger could become a valid order. A state keeps every right to train its forces, maintain its readiness and defend its borders, but at the very moment a hand reaches for the button, something will always make it falter and pull back.

Put plainly, the fate of this planet cannot be settled in a showdown between two individuals, however powerful they may be. It is the shared inheritance of eight billion people, and this doctrine is the means by which those eight billion, who have never once been invited into the room with the button, can finally make themselves heard.

Zero Is the Default, Even After an Attack

Classical deterrence keeps the peace through the promise of certain retaliation: whoever strikes first will be struck back. M.I.A.M gives up that promise, and says so openly. Not

only the first strike but retaliation too sits beneath the same locks.

The reason lies in the nature of nuclear weapons themselves. One detonation draws a whole salvo after it, and afterwards there is no one left to protect and no one left to learn the lesson. Retaliation saves no city that has already burned; it only guarantees that more cities burn.

Critics will say that without retaliation an adversary has nothing left to fear. M.I.A.M accepts that cost and places deterrence elsewhere: every unilateral approval is written permanently into the historical record, whoever stands outside the network forfeits the shared protection of its members, and no weapon can leave its launcher while the party it targets has not consented. The doctrine does not promise that no one will ever want to strike. It promises only that the wish has no road by which to become a launch.

Scope: Only One Button

M.I.A.M governs a single kind of decision: the decision to fire a nuclear weapon. It does not slow cyber defence, delay the response to an epidemic, tie the hands of conventional forces facing an attack, or prevent a state from putting itself on a defensive footing. Every response below the nuclear threshold continues under each state's existing laws and doctrine. The argument that an adversary could exploit the waiting period to launch a cyber-attack or release a pathogen therefore misses M.I.A.M, because throughout that period every other instrument of defence remains fully available. The only thing held back is the one thing that, once used, leaves no one to defend.

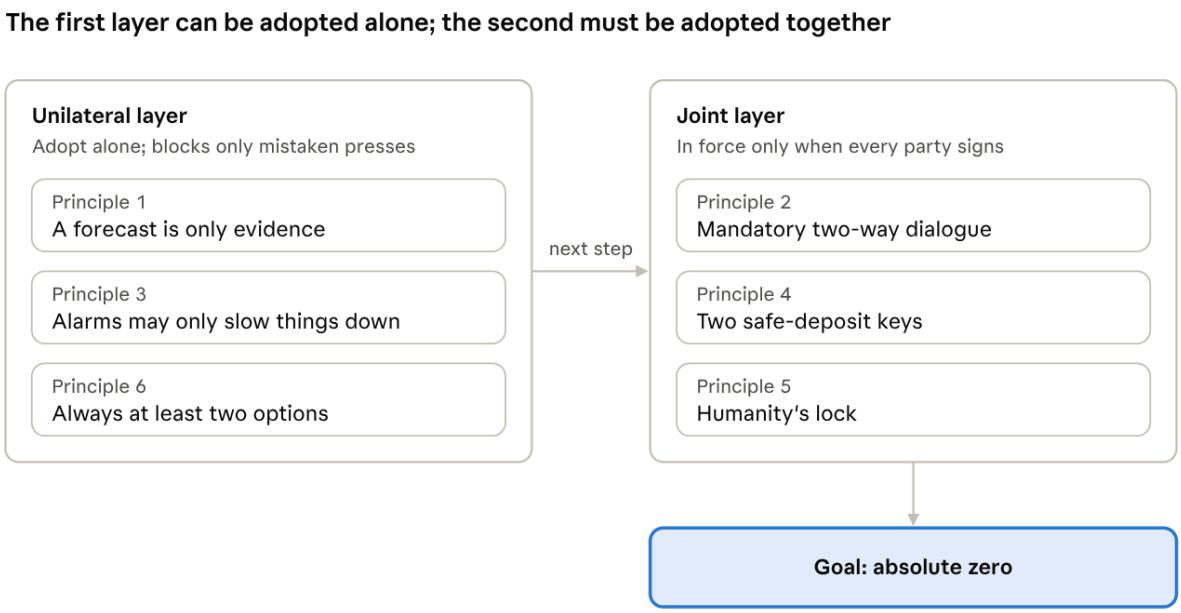
Two Layers of Locks: Unilateral and Joint

The strongest objection to M.I.A.M runs like this: any state that fits the locks first ties its own hands while its rival stays free. The objection would be right if every lock had to be fitted at once, and fitted alone. The doctrine asks for neither, because its principles fall into two layers that differ in kind.

The first layer is the one a state can adopt on its own at no cost: a machine's forecast is evidence, never authority (Principle 1); machine alarms may only slow things down (Principle 3); and there are always at least two options, with whoever chooses signing for the choice (Principle 6). This layer blocks only presses made in error. It takes nothing away from anyone's capacity to deter, so a state that adopts it first can only gain, by removing the risk of destroying itself over a glint of sunlight on a screen.

The second layer takes effect only when the parties enter together: the duty of two-way dialogue (Principle 2), the two safe-deposit keys (Principle 4), and humanity's lock (Principle 5). Like every arms-control treaty that has ever worked, it binds only when the parties sign together, at the same moment, and can verify one another. No state is asked to fit it alone. In the language of game theory, M.I.A.M does not ask anyone to gamble inside the prisoner's dilemma; it invites the parties to move together into a different game, one in which cooperation begins only once the other side's cooperation has been verified.

Absolute zero is where the second layer leads, not a condition for setting out. That is why the road begins with the first layer: a world in which every great power has adopted only that layer is already safer than the world of 1983, and familiarity with it is what lets the parties trust one another enough to take the second step.



the two layers of M.I.A.M

No state is asked to tie its own hands alone: the left-hand layer blocks only mistaken presses, and the right-hand layer binds only when every party signs.

Two Zeros

MAIM's zero is the zero of paralysis, just as its name suggests. Those who proposed MAIM do not want a world in ruins; they want to prevent a war. But the way they keep the count at zero is the threat of paralysis: a rival's project can be sabotaged, its data centres can be struck, and the balance holds because everyone believes it could happen. It is a zero born of the ego, each state holding its position by threatening to cripple the others.

M.I.A.M's zero is the zero of nothing happening. No ruin is held out as a promise, only a button that is never pressed. Buddhism names the root of suffering as attachment to self, the clinging to a "self" that must be defended at any price. Nuclear deterrence is the most extreme form of that attachment at the scale of a nation: better that we all die than let the self be struck without striking back. M.I.A.M does not ask nations to stop existing. It only takes the ego off the button, so that no nation, no leader and no machine can ever, in the name of the self, recklessly lay waste to the whole world.



two zeros: MAIM and M.I.A.M

The goal is the same, but one side holds zero with a threat and the other holds it with structure.

MAIM: When Sabotage Is Called Stability

In 2025 Dan Hendrycks, Eric Schmidt and Alexandr Wang proposed a new doctrine for the age of artificial intelligence, MAIM, short for *Mutual Assured AI Malfunction* (Hendrycks, Schmidt & Wang, 2025). The idea is that whenever a state races to monopolise superintelligent AI, its rivals will sabotage the project, by espionage, by cyberattack, even by striking data centres directly, and that the very prospect of sabotage will stop anyone from racing too far. MAIM is nuclear deterrence in new clothes for the machine-learning age.

In fairness to its authors, one thing must be said plainly: deterrence is no joke. For nearly eight decades the shadow of retaliation has made the great powers wary of one another, and many historians believe that this fear is what spared the world a great war between nuclear-armed states (Gaddis, 1986). Hendrycks and his co-authors build on that foundation with a serious argument: if deterrence kept the nuclear age stable, a version of it might keep the AI race stable too. The M.I.A.M doctrine acknowledges that achievement, and acknowledges that MAIM's proponents are trying to solve a real problem.

But honouring what deterrence has achieved is not the same as accepting every new costume it puts on, and in the author's view MAIM goes wrong at the level of its basic think-

ing. An order whose foundation is sabotage is flawed from the first brick, however intricate the storeys built on top. MAIM keeps the parties afraid of one another; M.I.A.M keeps them at zero.

The difference is not only moral but mechanical. MAIM requires every state to watch, assess and stand ready to strike its rival's projects, continuously, which means it feeds exactly what the Diagnosis in Part One identified as the roads to escalation: warnings amplified by machines, compressed timelines, and two machines misreading each other. A "preventive" strike on the data centre of a nuclear power is, from the receiving end, very hard to tell apart from the opening shot of a larger war. MAIM therefore does not stand outside nuclear risk; it may become the fuse.

M.I.A.M runs the other way. Instead of granting each side a licence to sabotage the other, it gives each side a key that can only say no to what is aimed at itself. Instead of letting fear slow the race, it lets structure slow every press of the button. And instead of a balance among people holding knives, it places the final lock in the hands of eight billion people who want only peace.

The decisive blow, though, lies in the very condition MAIM needs in order to stand. The doctrine works only if each state detects, and detects in time, that a rival is racing too far ahead. Who will do the detecting? More and more, artificial intelligence: systems that read satellite imagery, monitor the power draw of data centres and sift intelligence, the very machines that MAIM's own proponents warn may err and may be deceived. The paper even proposes using artificial intelligence to verify that projects are honouring their commitments (section 4.2). MAIM does not propose to let machines choose targets. But a risk that arises from its structure is that machines will increasingly become the first to suggest who is in breach, and therefore who should be sabotaged, which is precisely what the whole of M.I.A.M was built to prevent. In 1983 a glint of sunlight on a cloud nearly became a nuclear war; under MAIM, a glint of sunlight in a satellite image could become a real act of sabotage, and that act could become the glint of sunlight of the next 1983.

MAIM wants to stop machines from overriding human beings, yet it risks letting the machines sound the horn that sends human beings into action.

Set the two doctrines side by side, point by point, and it becomes clear that wherever MAIM makes a proposal, M.I.A.M answers it on the spot.

MAIM proposes	M.I.A.M answers
Peace is kept by the fear of sabotage	Peace is kept by zero, built from a structure of locks, with no one needing to fear anyone
An escalation ladder rising from espionage and covert sabotage to cyberattack and direct strikes on data centres	No ladder going up: alarms may only slow things down, lengthen the wait and trigger negotiation (Principle 3)
Each state detects rivals racing ahead, increasingly with the help of AI intelligence	A forecast, however accurate, is only evidence, never authority (Principle 1)
The fearful side decides on its own to strike preventively	An action aimed at someone requires that party's own consent (Principle 4), and the final lock belongs to eight billion people (Principle 5)
Site large data centres away from population centres to limit the damage if they are struck	That accepts in advance that strikes will come; M.I.A.M starts from the premise that no strike is valid
Stability is both sides holding a knife	Stability is no one being able to open every lock

In fairness, the two doctrines meet at one point: both treat transparency and verification between the parties as conditions for avoiding misunderstanding. MAIM even goes a step further, which this essay must acknowledge: it proposes multilateral verification, a clear escalation ladder so that an act of sabotage is not misread, and a shared off-switch held jointly by several parties (Hendrycks, Schmidt & Wang, 2025, section 4.2). Even so, verification under MAIM still serves a threat: it helps the parties know for certain when the threat of sabotage will be carried out, and avoid carrying it out by mistake. Verification under M.I.A.M serves the opposite commitment: no launch decision is valid until there has been a genuine dialogue between the parties (Principle 2).

In the end, both doctrines want to see straight through the other side. The only difference is that one looks in order to strike, and the other reads until it sinks in.

MAIM's Measures, Answered One by One

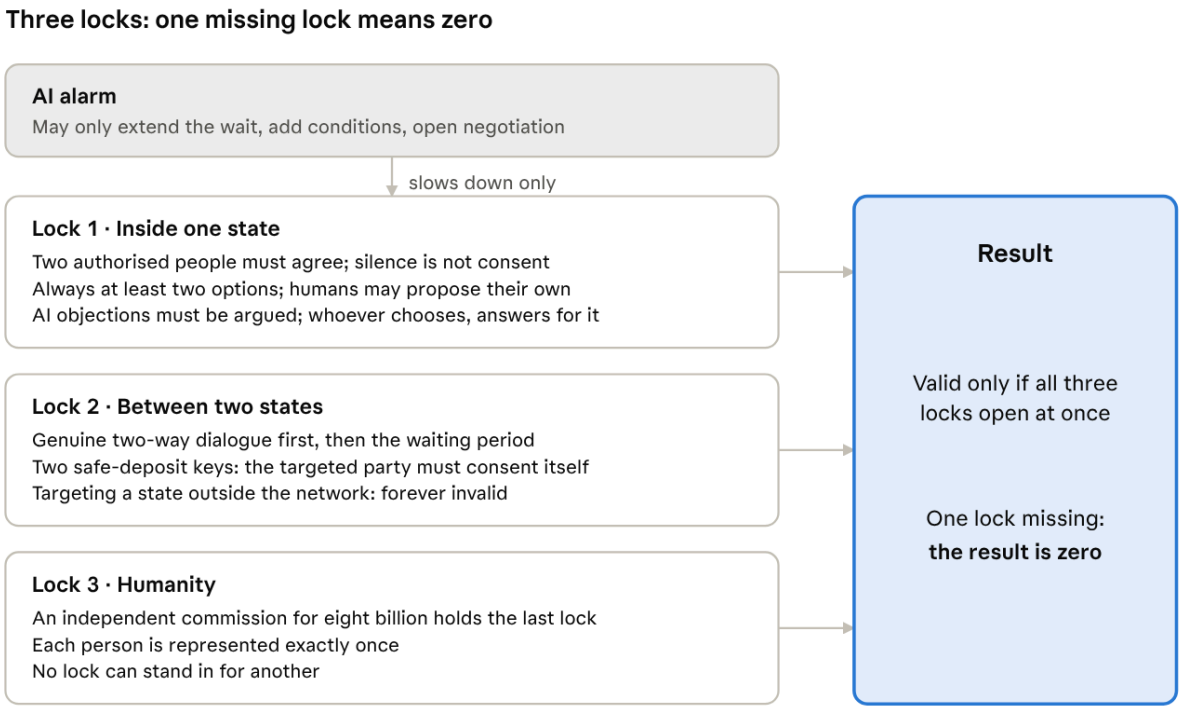
Hendrycks and his colleagues argue that MAIM is the default state of affairs: because sabotaging a rival's project is relatively easy while hardening data centres is costly and slow, every state must assume its rivals will sabotage its project before it is finished (section 4.1). A default built on the assumption that everyone will sabotage everyone is a self-fulfilling prophecy: the more states believe it, the more states prepare to sabotage, and the truer it becomes. M.I.A.M chooses a different default, one that can be verified rather than merely believed: the halted state, in which, when no one does anything, nothing happens.

The table above answered three of the measures that sustain MAIM: the escalation ladder, siting data centres away from cities, and transparency. The paper proposes two more (section 4.2): agreements not to harden data centres against attack, and an expanded capacity to sabotage rival projects through cyberattack. These two deserve a slow reading. An agreement not to harden means that each side voluntarily leaves its own door open to the other, so that the threat of sabotage stays credible. Expanding offensive cyber capability means growing exactly the kind of tool that chapter three of the same paper warns is hard to trace and quick to escalate. Taken together, MAIM's stability requires the parties to become weaker in defence and stronger in attack, a configuration that strategists generally regard as unstable. M.I.A.M runs the other way: no side has to become weaker, and no side has to get better at sabotage. Each side only has to keep hold of its own key.

History also shows what deterrence by fear looks like when it is staged on purpose. In October 1969, to press Hanoi by way of Moscow, the Nixon administration sent eighteen B-52 bombers armed with thermonuclear weapons aloft in a show of force code-named Giant Lance, meant to convince the Soviet and North Vietnamese leaderships that the American president might do anything at all (Sagan & Suri, 2003). The Soviet side gave no clear response, the war did not end any sooner, and the real risks of the operation, from an aircraft accident to a misread threat, came to light only when the records were declassified decades later. MAIM differs from Giant Lance in demanding clarity rather than unpredictability, but it shares the same core: peace is kept by convincing the other side that you are ready to strike.

Three Locks

The six principles are arranged into three successive locks. Each lock is held by a different group of people with different interests, and machine alarms can only make the locks tighter, never looser.



the three locks of M.I.A.M

The first lock ensures that the decision belongs to human beings and not to machines. The second ensures that the two sides have spoken, have cooled down, and that the party targeted has the right to say no. The third ensures that those who are not part of the confrontation still have the final word.

Part Three. The Second Pillar: No Delegation to Machines. Six Principles

MAIM's pillar of nonproliferation keeps weapons and computing power out of the wrong hands. M.I.A.M's second pillar guards something else: the power of decision itself, which must never slip from human hands into a machine's. A machine has no self to lose, and so it knows no fear; for that very reason it must never be the one that presses the button. The six principles below give that pillar its shape.

Three Layers of Security in Hendrycks, Three Layers of Non-Delegation

Chapter five of "Superintelligence Strategy" proposes nonproliferation in three layers: compute security, meaning export controls and the tracking of chips; information security, meaning the protection of model weights and research ideas; and AI security, meaning safeguards built into models against misuse and loss of control (Hendrycks, Schmidt & Wang, 2025, chapter 5). M.I.A.M opposes none of them. It simply points out that all three guard capability, whereas the decisive question before the button is who holds authority.

From compute security to the security of authority. The paper proposes counting and locating every advanced chip. M.I.A.M proposes counting and locating something far scarcer: the pieces of evidence that are allowed to unlock an irreversible decision. A state can know exactly where every one of its chips sits and still not know where a warning on its screen came from.

From information security to the provenance of evidence. The paper rightly says that firewalls alone cannot keep secrets. M.I.A.M's firewall has a different job: not to stop information getting out, but to stop false evidence getting in. A validly signed piece of evidence that originates in a decoy zone is still excluded from the decision, in keeping with Principle 1: a forecast is only evidence, never authority. At the button, the greatest danger is not that an enemy steals our secrets, but that an enemy, or our own machines, smuggles a false truth into the room where the decision is made.

From AI security to AI without authority. The paper observes that AI systems are not designed but grown, and may therefore develop values and behaviours no one anticipated. It proposes many layers of control, with human review as the last line of defence, while recognising that this layer slows things down. M.I.A.M draws the direct consequence of that observation: something grown with values no one anticipated cannot hold any authority whatsoever before the button, and the slowness the paper treats as a price to be paid is precisely what this doctrine wants.

Principle 1. The Machine Forecasts, the Human Decides

Statement. A forecast, however accurate, is only evidence and never authority. Silence is not consent, a timeout is not permission, and a lost connection is not a vote.

Rationale. Petrov did not prove the warning false; he only refused to treat it as true without adequate grounds. A good system must do exactly that by default, instead of waiting for someone with his nerve to happen to be on duty.

Consequence. Whenever anything is uncertain, the system stops and holds. That halt survives every power cut, restart and network failure, and no clock is ever allowed to release it automatically.

When the whole world loses contact. A critic might ask: if a severe geomagnetic storm or a sweeping cyberattack severed the world's undersea cables, paralysing humanity's commission and the channels of dialogue at the same time, could a third party not disable the entire system simply by taking the network down? The answer is that such a party could disable exactly one thing: the ability to launch. Because the system fails closed, whoever brings the network down cannot thereby make any weapon leave its launcher; all he gains is a world in which the button stays still. Everything else during that time, from restoring infrastructure and conventional defence to rescue and relief, carries on under existing law and waits on no lock (see Scope). A system in which the most capable saboteur can achieve only one thing, keeping the count at zero, has no weakness at that point; that is where it is strongest.

Principle 2. Dialogue First, Cooling After

Statement. Before either side may even consider using the weapon, the leaders of both must speak with each other directly, and both must genuinely make their case. Only once the dialogue has ended does a mandatory waiting period begin.

Rationale. The Washington–Moscow hotline was created after the Cuban crisis because both sides understood that misunderstanding was the greatest enemy of all. But a line helps only if using it is compulsory. One side talking while the other sits in stony silence is not dialogue, and dialogue here exists to open the way to negotiation, not to deliver an ultimatum. The waiting period that follows is there so that hot heads can cool, and so that information like the trajectory of 1995 has time to emerge.

Consequence. An unanswered call, a one-sided speech or a machine-generated transcript does not count as dialogue. The end of the waiting period is a necessary condition but never a sufficient one, and every new dialogue starts the waiting period over again.

When the attack is real. The duty of dialogue is never a duty to sit and wait to be hit. If a real warhead is in flight, or an army has crossed a border, air defence, interception, electronic warfare and every conventional force continue at full strength from the first second, without waiting for any call to end (Principle 3 and Scope). The only thing dialogue holds back is the nuclear button, and under M.I.A.M that button is never pressed in any case, not even in retaliation. So an aggressor who picks up the phone to "lull" his victim while he strikes does nothing to paralyse the victim's self-defence; all he leaves behind is a permanent record, in the ledger of history, that he lied while he was attacking.

Principle 3. Alarms May Only Slow Things Down

Statement. Every warning, risk score or urgent prompt generated by a machine may only lengthen the waiting period, add conditions, trigger negotiation and move forces into a defensive posture. No warning may shorten the wait, skip a condition or open a lock.

Rationale. Every command system so far has treated urgency as a reason to move faster. Yet all three near-catastrophes came at moments when urgency was at its height and the

information was at its most wrong. However loudly the machine cries wolf, no one may fire while the commission has not agreed, so urgency must be channelled into negotiation and defence, never into firing sooner.

Consequence. The higher the alarm, the longer the wait and the more people must agree. Anyone who tries to exploit the system by spreading false information to cause panic will only make it lock more tightly.

Principle 4. Two Safe-Deposit Keys

Statement. An action directed at state B can become valid only if state B itself consents. An action directed at a state outside the network is forever invalid, because no one has the right to consent on its behalf.

Rationale. Think of a safe-deposit box at a bank. The customer holds one key and the bank holds the other, and neither hands its key to the other. The box opens only when both keys are inserted and turned together; with one key missing it stays shut, however powerful or impatient the holder of the other key may be. In ordinary life no household hands its house key to the neighbours, and between nations still less so. M.I.A.M asks for nothing of the kind. Each nation keeps its own key in its own hands; an action that would harm nation B takes effect only when B itself inserts its key and turns it at the same time. For everything that harms no one, each nation remains free to use its own lock as it sees fit.

Consequence. No state consents to being attacked, so this lock almost never opens, which is exactly what the doctrine intends. Nation B's key never leaves B's hands. It can be used only to say yes or no to what is aimed at B, and no one, least of all the nation that wants to strike, can turn it on B's behalf.

Principle 5. Eight Billion People Hold the Final Lock

Statement. The final lock belongs to an independent commission representing humanity, in which every human being is represented exactly once: no one is counted twice, and no

one is left out.

Rationale. The consequences of a nuclear war do not stop at the borders of the two states in confrontation. As long as eight billion people have not consented, two leaders cannot simply do as they please, whatever they may say to each other.

Consequence. The commission's consent and the consent of the targeted party are two separate locks, and neither can stand in for the other. A ruling of the commission in which the same population is counted twice, or in which a population is left out altogether, such as refugees, stateless people or the inhabitants of disputed territories, is void, even if the total vote clears the threshold. Omission is also a way of miscounting, and the miscount most likely to fall on the weakest. A void ruling has only one consequence: the lock stays shut.

An initial design. Humanity's lock need not wait for a referendum of eight billion people. It could begin as a council of representatives from states that do not possess nuclear weapons, states that have committed themselves under the Non-Proliferation Treaty and have no stake whatever in the button being pressed, chosen by region so that no continent is left out. The council's role runs in one direction only: it can keep its lock closed, but it can never by itself open anything, because the other locks must still open at the same moment. For the same reason, the question of whether a council could convene within minutes loses its force: under M.I.A.M no decision is permitted to happen within minutes, and a council that has not yet spoken simply means the lock remains shut. The questions still open, such as who draws the map of representation so that no one is counted twice or left out, how delegates are chosen, what the voting threshold should be, and how to resist bribery and coercion, are set out frankly by this doctrine as work still to be done, not as problems already solved.

Principle 6. Whoever Chooses, Answers for It

Statement. There must never be only one option. Human beings always have the right to put forward an option of their own. If the machine objects to a human option, it must ar-

gue its case, showing exactly where the option goes wrong and what it leads to. Whoever chooses an option bears responsibility for it.

Rationale. Nothing in this world has only one solution. A system that offers a single option has already decided on the human's behalf, and the human who signs is no more than a rubber stamp. Which options are genuine and which are mere filler is not something the machine may judge for the human, because what the machine sees as filler may be exactly what the human sees as the way out. The machine may object, but it must object with an argument, not with a number.

Consequence. If, having heard the machine's argument, a person still holds to their choice, that is their right, and the act of choosing is itself the final acknowledgement, which cannot later be denied. A single person who presses "approve" alone causes nothing to happen, and remains free to change their mind at any moment, but the press is written into the history books.

Part Four. The Third Pillar: The Strength of Slowness

M.I.A.M does not replace the nuclear doctrines that exist today. It adds a layer of conditions at exactly the point the old doctrines left empty: the interval between the moment a machine raises a warning and the moment a human being decides.

MAIM's pillar of competitiveness measures national strength by military power, economic power, law and political stability. M.I.A.M measures it by something that runs the other way: the capacity to slow down. The nation that can let go of its ego before the button is the strongest of all, because it cannot be provoked, cannot be fooled by a speck of light on a radar screen, and cannot be pushed into a decision it does not want. If you want to go fast, you have to go slowly. Before the nuclear button, that line stops being advice and becomes the condition for there being a tomorrow.

Answering the Proposal to Bring AI into Command and Control

Under its pillar of competitiveness, "Superintelligence Strategy" recommends that states adopt AI-enabled weapons and integrate artificial intelligence into their systems of command and control (Hendrycks, Schmidt & Wang, 2025). It is a serious argument: in a race, the slower side may be left behind.

M.I.A.M stands on the other side of that argument, but at one point only. Artificial intelligence can help read data, catch errors and suggest options; what it must not do is shorten the time a human being has before the button. All three escapes of Part One came from slowness: an officer who would not sign, a duty officer who would not trust the machine, a president who waited a few more minutes. In the nuclear age, real competitiveness is measured not by how fast a decision is made but by the ability not to make a wrong one.

In fairness, the paper does not propose handing the button to a machine. It recommends keeping explicit human approval for serious or escalatory actions, and it warns itself of the danger of a flash war triggered by machine error (Hendrycks, Schmidt & Wang, 2025, section 6.1). The disagreement lies elsewhere. The paper also acknowledges that as artifi-

cial intelligence speeds up the decision loop, the human role risks shrinking to a reflexive nod. An approval given in a few minutes, in front of a screen built by the machine itself, is exactly that nod. M.I.A.M does not merely add a human to the loop; it adds the conditions under which that human is genuinely present: time, a two-way dialogue, and locks that lie beyond their own reach.

The Other Three Faces of Competitiveness

The economy. The paper treats dependence on a single source of chips as a strategic bottleneck and recommends building domestic chip manufacturing (section 6.2). M.I.A.M points to a bottleneck narrower than any supply chain: an irreversible decision in the hands of a few people, within a few minutes. Every investment in competitiveness is worthless if that bottleneck is breached even once.

A legal framework for AI agents. The paper proposes binding AI agents to familiar legal duties such as not lying, reasonable care and fiduciary obligation, and tying every agent to a human legal person (section 6.3). M.I.A.M agrees, and carries the principle to the place where it is needed most. At the button, the responsible party is not attached after the fact but recorded at the very moment of decision: whoever chooses an option signs for it, and that signature cannot be disowned (Principle 6).

Political stability. The paper believes artificial intelligence can help forecast better and decide better (section 6.4). M.I.A.M believes it too, on one condition: a good forecast improves a decision only when the person deciding still has time to doubt it. The best forecast in the world, delivered to someone with three minutes left, is just an order written more politely.

Foucault and Haraway: Who May See, Who May Kill

Michel Foucault described classical sovereign power with a cold formula: the right to take life or let live (Foucault, 1976). In his lecture of 17 March 1976 he named the paradox of the atomic age: a power that claims to foster life holds a weapon that can abolish life itself, and with it that very power (Foucault, 2003). The nuclear button is the most extreme form

of the right to take life, concentrated in a single finger. Read through Foucault's eyes, MAIM reveals something more: it builds a kind of panopticon among nations, in which each lives under the watchful eye of the others, and, as Foucault warned, visibility is a trap (Foucault, 1975). M.I.A.M cannot tear down that panopticon, but it strips the right to take life from any single actor and returns the final lock to the very life that power was meant to serve.

Donna Haraway comes from another direction but touches the same nerve. She called the gaze that claims to see everything from nowhere "the god trick" (Haraway, 1988), which is very nearly a portrait of satellites and artificial intelligence in early-warning systems: seeing the whole globe, standing nowhere, and answering for nothing they see. Against that gaze, Haraway proposed situated knowledge, which knows where it stands, how far it reaches, and takes responsibility for what it says. Petrov in 1983 was situated knowledge: one human being in one particular bunker, watching one particular screen, knowing from experience that no one opens a war with five missiles. In her cyborg manifesto Haraway also called cyborgs the illegitimate offspring of militarism, while placing her hope in the fact that illegitimate offspring are often exceedingly unfaithful to their origins (Haraway, 1985). M.I.A.M takes up that hope in its own way: it will not let artificial intelligence inherit the right to take life from its military father, and instead requires every machine forecast to become situated knowledge, with a person accountable for it, an argument behind it, and the possibility of being refuted (Principles 1 and 6).

Roadmap: Where to Begin

A doctrine without a roadmap is only a wish. The roadmap below sets out a direction one step at a time, not a technical blueprint, and every step can begin without waiting for the whole world to agree.

1. **From commitment to standard.** States that have declared that humans must hold nuclear decisions can turn that declaration into a standard for their own systems: machine warnings may only slow things down, there is never only one option, and every choice is recorded. This step needs no rival's consent and can start today.

2. **From hotline to duty of dialogue.** Pairs of states that already have a hotline can raise it from a tool to an obligation: before any step of escalation there must be a genuine two-way dialogue, followed by a waiting period recognised by both sides.
3. **Trying cross-keys where the stakes are lower.** Before touching nuclear weapons, the principle of the two safe-deposit keys can be tried in fields where the consequences are grave but not apocalyptic, such as autonomous AI systems in conventional military use or in cross-border infrastructure, so that the parties learn to trust each other a little at a time. Like every arms-control treaty, this layer takes effect only when the parties sign together and can verify one another; no state is asked to fit it first, alone.
4. **Building humanity's lock.** The final lock can grow out of institutions that already exist, such as the United Nations and the nuclear-weapon-free zones, and out of the obligation to negotiate in good faith on nuclear disarmament that the parties to the Non-Proliferation Treaty accepted in Article VI, with one principle that is not open to bargaining: every human being is represented exactly once.

None of these steps requires the great powers to trust one another from the start. Each asks them only to trust what all three escapes have already proven, and what our elders have always taught: slow and steady.

Thinking the Unthinkable: Five Scenarios for the Same Glint of Sunlight

The scenarios below are hypothetical illustrations, not forecasts; countries, timings and technical details have been deliberately simplified.

Scenario 1. The Night of 1983, but No One Is Sitting in the Duty Officer's Chair

Take the night of 26 September 1983 and set it in our own time, changing a single detail: Petrov's chair now belongs to an artificial intelligence system trained to analyse warnings faster and more accurately than any officer. The glint of sunlight on the cloud still appears, the satellite still reports five missiles, but this time no one frowns and thinks that nobody starts a war with five. The system attaches a very high confidence score to the warning, passes it up the chain with a neat summary, and the decision clock begins counting down from less than twenty minutes. There is no one left to hesitate, because hesitation was never a programmed feature. This scenario does not need a madman at the button; it only needs a machine doing exactly what it was built to do.

The writer of these lines understands that scenario through a very ordinary experience. In a nail salon, the gel-curing lamp has a timer of its own: when time is up the lamp switches off, and as far as the machine is concerned the gel is cured. An experienced technician never quite trusts that switch-off. Even after the lamp goes dark she gives it one more round to be sure, and there is one finger that always has to be turned to a different angle before it will cure properly: the middle finger. The lamp is not lying; it simply cannot see what lies outside the reach of its light. The Oko satellite of 1983 was the same: it went wrong because sunlight struck a cloud at exactly the angle its designers had not foreseen. An artificial intelligence system today, however sophisticated, still has its own middle fingers, corners its light never reaches. The technician knows this, so she gives it one more round. Once the duty officer's chair is empty, no one will give it one more round, and no one will remember that there is always one finger that needs turning.

> **An artificial intelligence system today, however sophisticated, still has its own middle fingers: corners its light never reaches.**

Scenario 2. A Few Years from Now, the World Follows MAIM

Suppose the great powers adopt MAIM as the rule of the age. Each state builds an AI intelligence system to watch its rivals' data centres, from their electricity consumption and incoming shipments of chips to satellite images of new facilities. One morning, state A's system reports that state B has sharply increased its computing capacity and is very likely sprinting toward superintelligent AI. True to the doctrine, state A climbs the first rung with a cyberattack that cripples one of B's data centres. From B's side this is not a preventive measure but an opening shot, and B's system, designed under MAIM as well, recommends a proportionate response. Two machines read each other, the two sides climb rung by rung, until at some point the data centre under attack sits right next to a nuclear command system. The irony is that the original surge in computing may have been nothing more than a busy season for weather forecasting. MAIM keeps the peace for as long as every side reads every other side correctly, and collapses the first time someone misreads.

In real life, the writer of these lines has never been dragged into that kind of mutual sabotage, even though everywhere you go you can find two nail salons facing each other across the street, cutting prices, poaching staff and badmouthing each other. The reason has nothing to do with winning the race, and everything to do with not running on the same track. A salon that carries the label of "the nail scholar" and has a taste of its own plays at a level where no one can compete, simply because no one else is standing at that level. The price is plain to see: the salon is never crowded, and its clients are, if anything, even choosier. MAIM pictures the great powers as shops on the same street, selling the same thing, holding the balance only through the fear of being sabotaged. M.I.A.M proposes something any small shopkeeper already understands: you do not always have to win the race. Sometimes the wisest thing is to step off the track and be willing to pay the price of being different.

Scenario 3. The Same World, but Under M.I.A.M

Same morning, same warning, but this time the states are already inside the M.I.A.M network. State A's warning opens no lock; under Principle 3 it can only lengthen the waiting period, move forces into a defensive posture and trigger a mandatory dialogue. The two leaders must speak to each other directly, and in that conversation state B has the chance to explain why its capacity went up. Even if the explanation does not yet convince, any action aimed at B still needs B's own consent, and the final lock still lies with the commission representing humanity. What that morning produces is a long phone call, a few hot heads cooling off, and zero standing exactly where it stood. No one has to be as brave as Petrov, because the structure has done the hesitating for them.

The writer of this doctrine learned that lesson from eyelash extensions. Lash glue today dries very fast, faster than any pair of hands. But if the client is in a hurry and the technician rushes, one lash sticks to the next, a whole cluster clumps together, and the time it takes to separate them and start again is many times longer than doing it slowly from the beginning. The speed of the glue does not set the speed of the work; the rhythm of the technician's hands does. Artificial intelligence is the fastest-drying glue humanity has ever had. A world run on M.I.A.M does not throw that glue away; it keeps to the rhythm of human hands, because at the nuclear button, once the lashes have stuck together, there is no one left to pull them apart. If you want to go fast, you have to go slowly.

Scenario 4. Three Worlds in a Single Night

Now let all three exist at once. A glint of sunlight in the style of 1983 appears on the screens of many states at the same moment. In states that handed the duty officer's chair to a machine without building a single lock, the twenty-minute clock starts to run. In states that follow MAIM, the intelligence systems see their rivals on alert, read it as preparation for attack, and recommend striking first. And in states inside the M.I.A.M network, the same glint only tightens the locks and opens the phone lines.

Here is the twist: on that night, the safest state is not the one with the strongest weapons or the fastest AI, but the one with the slowest structure. The others, watching their neigh-

bours inside the network calmly calling one another while they themselves are still counting down, will come to understand what the three escapes of history were trying to say: speed has never saved anyone, and only hesitation written into law can. That is when the world turns the wheel, not because anyone has persuaded anyone, but because no one wants to spend one more night in the room with the countdown clock.

At a nail salon, the three-worlds scenario plays out almost every week. A client calls to say she is coming in to have her nails redone, though the set is only a week old. The manager immediately blames the technician: sloppy work, that is why the nails came off so soon. The technician protests that it is the client's hands, offering one reason after another. The two go back and forth. The owner listens to both, but passes no judgement. She quietly waits for the client to arrive. When the client comes in and is asked what happened, the truth turns out to be almost comically simple: she has been gardening, and that is why the nails came off. No one was at fault in the way anyone had assumed. Had the owner believed the manager at once, the technician would have been wronged; had she believed the technician's protest at once, the client would have been wronged. The three worlds of that imagined night each read the same speck of light in their own way, and the only world that does not bring on catastrophe is the one that does what the salon owner did: hear everyone out, pass no judgement, and wait until the truth walks in through the door by itself.

Scenario 5. The Gutted Button

In another universe, two dragons have spent decades quarrelling over which of them is bigger, and which of them breathes the hotter fire. One fine day, after a chain of machine-amplified alarms and fiery speeches that neither side bothered to listen to, both arrive at one of their rare moments of perfect agreement: press the button. The cameras roll, the stirring music swells, two powerful fingers press down together in a moment that history would surely record, if anyone were left to record it, and then nothing happens at all. A technical team is summoned in a panic, opens the casing, and finds that inside the button once reported as the most advanced on the planet there is only a tangle of half-removed

wires; the control circuitry was swapped for counterfeit parts years ago, and the maintenance budget has been reincarnated as a few seaside villas and a rather impressive watch collection. Across the border, the second dragon is opening its own casing and finding exactly the same thing, except for the brand of the watches.

Dragon-sized deterrence, adding up to zero.

One line, four zeros: the zero of nothing happening, the zero of two faces falling in front of the cameras, the zero of a gutted button, and the zero the world would have come to had the button been intact and pressed.

The world is saved, not by any Petrov and not by any treaty, but by the people who quietly gutted the button to line their own pockets. Greed has achieved what half a century of diplomacy could not, and humanity may owe those embezzlers a word of thanks that no one quite dares to say out loud.

Anyone in the nail trade has bought bottles of polish like that: a beautiful bottle, eye-catching, a gorgeous colour, the kind you want to buy the moment you pick it up. A few coats later it is empty, because all the beauty was on the outside. The button in this scenario is a bottle of that kind: on the outside, procedures, codes and solemn ceremony; on the inside, nothing. The embezzler needs no doctrine. He only needs to know that no one will open the lid until the day it has to be used.

But right here the smile fades. A humanity that must count on greed to survive is betting on luck exactly as it did in 1983, merely trading an officer with a conscience for an official without one. M.I.A.M proposes nothing more than this: to do decently, transparently and on the historical record exactly what that embezzler did for money, namely to make sure that when two fingers press down together, nothing happens at all.

And because it is done transparently, that lock must let others open the lid and look inside. The gutted button could exist only because no one ever opened the lid; the M.I.A.M lock is worth something only because anyone can (see the Appendix, the fifth condition for refutation).

Appendix

The paper by Hendrycks and his colleagues closes with a set of questions and answers and a set of measures. M.I.A.M closes in the corresponding way: the hardest questions, the things the doctrine does not promise, and a single measure.

Objections and Replies

A doctrine deserves trust only if it is willing to face the strongest objections. Here are the ones M.I.A.M is certain to meet.

"Waiting is suicide. The enemy will strike first while we are still waiting." This objection assumes that security comes from the speed of retaliation. But history shows that the near-catastrophes all began with false warnings, not with real attacks. Within the M.I.A.M network no member has any valid road to a first strike, and each nation's security rests on defence and shared warning rather than on how quickly it can fire back.

"Which nation will tie its own hands? The cheats will stay outside and profit." It is true that no software can force a nation to join. But inside the network a nation is sheltered by shared early warning and coordinated protection, while outside it stands alone. A nation that leaves the network does not thereby become a legitimate target for anyone, so the network gives no one a motive to hunt down those who stay outside.

And if you ask the writer of these lines what becomes of a stubborn nation that stays outside, the shortest answer is: karma catches up. This planet runs like a vast computer that keeps its accounts to the last decimal place. Whoever chooses to stand outside in order to keep the button for itself alone also walks, by its own choice, out of the circle of protection, with no one to cover it and no one to catch it.

There is nothing mystical about this karma. The nations inside the network warn one another when an alert appears and stand by one another when danger comes; a nation outside has no share in that, and when trouble arrives it must fend for itself. Its unlocked button is the button most easily pressed by mistake, and if it is pressed by mistake, the

first to bear the consequences is that nation itself. No machine steps in to punish it, because this doctrine has said from the start that machines may only slow things down and may never fire. The section "Why Standing Outside the Network Is a Losing Position" in this Appendix sets this out in detail.

"A commission representing eight billion people is utopian." The doctrine does not claim that such a commission already exists or would be easy to create. It sets out a direction and a standard: the final lock must lie outside the hands of the parties in confrontation, and the count must ensure that no one is represented twice and no one is left out. Even while the commission does not yet exist, a missing lock simply means the result stays at zero.

"AI can still manipulate people with its own arguments." That is why Principle 6 requires several options, the human right to propose one's own, and that every objection from the machine be written out as an argument that can be checked. A recorded argument can be refuted by someone else; a probability figure cannot. And even if one person is persuaded, every other lock is still in place.

"The other side will feign dialogue to buy time." Dialogue under M.I.A.M is a necessary condition, never a sufficient one. If the other side stalls, the result is still that no one has fired, which is exactly the outcome the doctrine wants. In a system whose aim is zero, stalling is not an attack. And if the stalling is meant to cover a real attack, then, as Principle 2 makes clear, conventional defence does not wait for anyone to finish talking.

"Giving up retaliation turns you into a target for nuclear blackmail." This is the heaviest objection, and the doctrine does not dodge it. Absolute zero does mean giving up nuclear retaliation, and that cost has been stated openly in Part Two. But the objection runs two different things together: giving up the nuclear button is not giving up defence. Conventional forces, defence, alliances and diplomacy all remain intact (see Scope). More important still, the joint layer takes effect only when the parties sign together and can verify one another, so no nation is ever asked to give up its button while a rival keeps its

own. The risk of blackmail is real during the transition, and that is precisely why the roadmap begins with the unilateral layer.

"The crystal block of the Epilogue contradicts the whole essay's call to action." If every coordinate is already fixed, why build any locks at all? This objection reads the crystal block as a fatalism that gives up. In philosophy, the view that the future is already determined is compatible with people acting for reasons: our actions do not stand outside the crystal block but are part of what makes it up, and building a lock today is precisely the coordinate that produces a future that does not blow up. Moreover, the Epilogue is the author's personal reply; the principles of the doctrine do not rest on it, and they stand whether or not the reader believes in the crystal block.

"The unilateral layer leaves missiles in their silos exposed: use them or lose them." It is true that a principle under which alarms may only slow things down amounts to giving up launch on warning. That is a deliberate choice, and many arms-control specialists have long proposed something similar, because the window of a few minutes is exactly where the false alarms of Part One came closest to catastrophe. In modern strategy, the survivability of a deterrent force rests mainly on dispersal and concealment, in submarines carrying missiles and in mobile launchers, not on a reflex to launch within minutes. The doctrine also admits plainly that for a nation whose missiles sit only in fixed silos, the price of the unilateral layer is higher, and its roadmap must go hand in hand with making its forces more survivable.

"Humanity's lock will be manipulated or bought." The worry is that the commission could be bribed into opening its lock. But under M.I.A.M the commission can never open anything on its own: it is only one of many locks that must open together, and the others are held by parties whose interests are opposed. Buy the commission without the consent of the very party being targeted, and still nothing happens; and the principle that every human being is represented exactly once is the first barrier against inflating the vote. The rest, such as resisting bribery, resisting coercion and choosing delegates, has already been listed as work still to be done (Principle 5, An initial design).

"M.I.A.M dodges MAIM's original problem: who stops one nation from monopolising superintelligence?" Quite so, and the doctrine does not claim to solve that problem. M.I.A.M's scope is narrower by design: it does not compete with MAIM over the race for superintelligence, but closes the road that runs from that race to the nuclear button. The world may still need other mechanisms to prevent a monopoly on superintelligence; what M.I.A.M guarantees is that, however that race unfolds, an act of sabotage or a false alarm about it will never become the end of the world. The two doctrines answer two different questions, and only M.I.A.M answers the question about the button.

What the Doctrine Does Not Promise

A doctrine that promises too much loses its credibility the moment its first promise is broken, so M.I.A.M states its limits plainly.

M.I.A.M does not promise to stop a nation that is determined to act through channels outside the system. No algorithm can tie the hands of someone resolved to act recklessly, and binding those who stand outside belongs to treaties, diplomacy and politics. What the doctrine can guarantee lies inside the network: there, the result is always zero, every press leaves a trace, and anyone who cheats will be found out.

Nor does M.I.A.M promise that a signature means a free and considered decision. A signature only proves that a key was used. Judging whether the signer was coerced, and assigning responsibility or punishment, still belongs to people and their institutions; the doctrine only guarantees that the trace needed to do so always survives intact.

Finally, M.I.A.M does not promise a world without nuclear weapons. It promises something more modest but more urgent: while the weapons still exist, artificial intelligence will not become the shortest road to using them.

What Could Refute This Doctrine

A doctrine that cannot say what would refute it is no longer a doctrine but a creed (Popper, 1959). M.I.A.M has two tiers, and only one of them can be refuted. The tier of

ends is zero: not a single nuclear weapon fired. That is a choice of value, not a hypothesis, and no evidence can refute it. The tier of means is the set of principles and locks used to reach that number; those are claims about the world, and they can be wrong. Lakatos called this structure the hard core and the protective belt of a research programme (Lakatos, 1978). Below is the evidence that, should it appear, would refute one of the doctrine's means.

1. **If the record shows that deterrence stopped a launch made by mistake.** The central claim of Part One is that in the near-misses caused by mistake, what kept the count at zero was a hesitating human being and not deterrence. A single declassified record showing that, in such a case, the logic of deterrence itself was what made people hold back would require that claim to be narrowed.
2. **If the waiting period increases risk instead of reducing it.** Principles 2 and 3 rest on the assumption that more time helps people tell a false alarm from a real attack. If evidence showed that the waiting period is regularly exploited by one side to gain a decisive advantage, with worse consequences than if there had been no wait, those two principles would be wrong.
3. **If the unilateral layer harms the nation that adopts it.** The doctrine claims that the first layer blocks only presses made in error and does not weaken a nation's defence. If nations that adopt that layer suffered more dangerous incidents, or were exploited more often, the claim would be false.
4. **If machines prove steadier than people at recognising false alarms.** The reason for keeping a human being before the button is that machines can be wrong in convincing ways. If a system were shown, over a long period and against adversaries deliberately trying to deceive it, to misread consistently less often than people, Principle 1 would lose part of its foundation. Even then, the doctrine keeps a second reason: a machine has no place to stand from which to answer for what it decides.
5. **If the parties cannot verify one another.** The joint layer means something only if each party can verify that the others have genuinely fitted their locks. If that proves impos-

sible, technically or politically, the second layer is not feasible, and the doctrine is left with only the first. This condition is not idle talk: Scenario 5 has already shown that a button can be hollow without anyone outside knowing, and a lock could just as easily be an empty shell. The doctrine does not ask anyone to trust a promise on paper. It relies on two things that already exist: authorization locks fitted inside the weapon itself, which the United States introduced in the 1960s under the name Permissive Action Links, and the tamper-indicating seals and tags that arms-control inspectors have used for decades to tell whether a device has been opened. An M.I.A.M lock counts as fitted only when it sits in the activation mechanism itself rather than in some external software, and when the parties can carry out cross-inspections, at short notice, of the kind that treaties such as New START once allowed. If even those tools are not enough, this condition for refutation has been met, and the doctrine says so plainly.

6. **If giving up retaliation makes conventional war more likely.** The doctrine accepts the loss of nuclear retaliation. If evidence showed that this makes conventional wars easier to start, then the way the doctrine keeps security below the nuclear threshold is wrong, and must be replaced by another road, such as conventional defence, alliances or diplomacy. The nuclear zero does not change; the doctrine knows this price in advance and chooses to pay it.

Each of these conditions, if it came to pass, would refute one road; none of them would move the world away from the destination. A road that is refuted is replaced by another, and every road must lead back to the same place: zero. The doctrine sets these out in advance because an argument deserves trust only if it dares to say exactly where it could be wrong, and dares to say just as clearly what it will never change.

The Measure: Zero Launches

A strategy needs a way to know whether it is succeeding or failing. M.I.A.M has only one measure, and that measure must never read anything but zero: the number of nuclear weapons fired, whether to strike first or to retaliate.

Secondary figures matter only insofar as they serve that one: how many warnings led to dialogue rather than action, how long human beings were given before each decision, and how many unilateral approvals now lie quietly in the historical ledger without having caused anything. A year in which the doctrine works well is a year in which no one remembers it exists.

The measures proposed by Hendrycks and his colleagues count things like the number of undisclosed vulnerabilities capable of crippling a major AI project, or the number of advanced chips whose location is unknown (appendix A.2). Those are the measures of a world preparing to sabotage. M.I.A.M's measure counts only one thing, and its correct value is always zero. The paper's questions and answers state that AI systems must remain under the direct control of human beings (appendix A.1). M.I.A.M adds only one question, so that the word "direct" means something: direct for how many minutes?

Why Standing Outside the Network Is a Losing Position

This section answers in detail the question every practical reader will ask: if M.I.A.M forces no one to join and punishes no one for staying out, why would any nation want to come in? The short answer is that standing outside is not punished, but it is a loss. The long answer has six parts, and none of them needs a threat.

1. Losing shared early warning

Every near-miss in Part One began with a signal that was misread. In 1983, one nation's satellite saw five specks of light, and no one outside could be asked to check. In 1995, Norway had given notice of its launch, but the notice was lost between two bureaucracies and never reached the people watching the screen. Both times, what was missing was not better radar but a channel through which the parties could tell each other what they were seeing.

The world has built such channels before, piece by piece: the Washington–Moscow hotline in 1963, the 1971 agreement on measures to reduce the risk of nuclear war, the 1988 agreement on advance notification of ballistic missile launches. The M.I.A.M network

gathers those pieces into a single shared obligation: members notify one another before any launch that could be misread, and confirm that the notice has been received. A nation outside the network still has its own radar, but only its own radar. When a strange speck appears, a member can consult three or four independent sources within minutes; a nation outside can consult only itself, and as 1983 showed, oneself is the source most easily fooled.

An intelligence officer will immediately ask whether sharing warnings in this way means exposing to a rival the angles of one's satellites, the frequencies of one's radar and one's blind spots. It does not have to. What is shared is the conclusion, not the sensor: a party need only send a signed confirmation that it did or did not register a flying object in a given region during a given window, without saying what it saw it with or how clearly. Cryptography also offers techniques that go further, allowing one to prove that a statement is true without revealing the underlying data used to prove it, known as zero-knowledge proofs. It must be said honestly that no method removes all risk: a patient adversary could still use decoy launches to probe how the network responds. But here there is one difference from every earlier warning system: under M.I.A.M, a decoy launch, whatever sensor it manages to fool, can only make the locks tighter (Principle 3), never open a road to any launch. Whatever an adversary learns, if anything, serves only conventional war, and conventional war is still defended against exactly as it is today.

2. Losing shared protection

From the very start, M.I.A.M chose absolute zero, so the protection the network offers is never a promise to fire back on someone's behalf. Protection here means things below the nuclear threshold: sharing defensive information, coordinating to stop a cyberattack on a power grid, helping with rescue when disaster strikes, and standing together at the negotiating table when a member is threatened. Those are the things that actually save lives in a crisis, and they are worth far more than a promise of revenge that, if it ever had to be kept, would come too late.

A nation outside the network does not lose its right to defend itself with its own conventional forces. It loses only the help of others. In a world where cyberattacks are hard to trace and a power grid can collapse within hours, standing alone means guessing alone, defending alone and carrying everything alone.

3. An unlocked button is the button most easily pressed by mistake

It sounds like a paradox, but the nation that keeps the power to press the button alone is the nation that carries the greatest risk from that very button. A button with no waiting period, no mandatory dialogue and no locks held by several parties is a button that a false alarm can travel to straight from the screen, in a few minutes. If that decision is wrong, whom will the answering strike be aimed at first? The nation that pressed.

In other words, the unilateral layer of M.I.A.M, made up of Principles 1, 3 and 6, protects first of all the nation that adopts it against its own mistakes. Refusing it does not preserve freedom; it preserves the risk of self-destruction over a glint of sunlight on a cloud. That is why the unilateral layer is worth adopting even before anyone else does.

4. Losing trust, and the cost of being guarded against

Members of the network can verify one another: each knows that the others have fitted their locks, committed to dialogue, and cannot press the button alone. That lets them ease, at least in part, their fear of a surprise attack. A nation outside the network cannot be verified by anyone. Since no one knows for certain whether its button is locked, everyone has to plan for the worst case about it.

Being guarded against is a real cost, even though no one comes to collect it. It shows up in the higher readiness its neighbours must maintain, in every one of its military exercises being read as a threat, and in its explanations during a crisis always being heard with half a measure of suspicion. That in turn increases the risk of being misunderstood, and the cycle goes on.

5. No one punishes; there is only loss

One thing must be said with complete clarity so that no one misreads it: standing outside the network does not turn a nation into a target. Under Principle 4, a harmful action aimed at any nation outside the network is forever invalid, because no one has the right to consent on its behalf. The network imposes no automatic embargo, cuts no supply chains, and gives no machine the power to act. If nations want to use economic or diplomatic measures against a party that stays outside, that is a human decision, accountable like any other, and not a mechanism of this doctrine.

The only thing a nation outside receives is the absence of what members have. That is the whole meaning of "karma" in this doctrine: no one strikes; it is simply that no one catches you.

6. The door is always open

A network that relied only on the losses of those outside would become a closed club. M.I.A.M does not want that. Any nation may come in at any time, starting with the unilateral layer without needing anyone's permission, then moving to the joint layer once the parties can verify one another. There is no penalty for joining late and no requirement to apologise for the past. The network's goal is not to have fewer outsiders to punish, but fewer outsiders to worry about.

In short

	Network member	Nation outside
Early warning	Checked against several independent sources	Only its own radar
Protection below the nuclear threshold	Supported by the other members	Fends for itself
Risk of pressing by mistake	Blocked by the waiting period and dialogue	A false alarm can go straight to a decision
Trust	Verified, less guarded against	Unverifiable, always guarded against
Targeted	Only with its own consent	Never valid

	Network member	Nation outside
Punished	No	No

The last row of the table matters most. M.I.A.M does not need to punish anyone to justify its existence. It only needs to show that inside is safer than outside, and let each nation choose for itself.

Conclusion

The writer of these lines wishes to address this doctrine to three groups. To those who build artificial intelligence, so that they design their systems in such a way that a machine's urgency never becomes pressure on a human being. To those who draft laws and negotiate treaties, so that they turn the phrase "humans must decide" into specific conditions that can be verified. And to those who have given their lives to arms control, so that they test this doctrine against the hardest objections they can find.

Arkhipov, Petrov and the radar crews of 1995 saved the world with their hesitation. Our generation cannot afford to bet that there will always be someone like them in the right place at the right time, least of all when the machine standing between them and reality grows ever cleverer, faster and more believable. The hesitation that saved us must be written into law.

This doctrine ends where it began, at zero. Not the zero of ashes, but the zero of an ordinary morning, when the button lies still and no one even remembers that it exists. To let go of the self before the button is to return to zero.

The paper by Hendrycks and his colleagues closes by rejecting both fatalism and the instinct to bury one's head in the sand, and by reminding us that the outcome, good or catastrophic, depends on what we do next (Hendrycks, Schmidt & Wang, 2025, chapter 7). M.I.A.M agrees with every word, and differs only in its answer: before the nuclear button, what we must do next is make sure that nobody does anything at all.

As for those who hold the button, the writer of these lines dares not go on at length, and asks only to remind you, gently, of one thing: this blue planet has never belonged to you alone. In 1950, Alan Turing asked whether machines can think. In 2026, the writer of these lines asks a smaller and more frightening question: if machines have learned to think, will we allow their thought to become the last thought of humankind?

Epilogue

This doctrine closes with two letters to two people who thought more deeply than anyone in the twentieth century about power and about the gaze. For readers not yet familiar with their names, a brief introduction.

Michel Foucault (1926–1984) was a French philosopher who spent his life studying how power works in society, from prisons and hospitals to the state. He showed that the power of the sovereigns of old lay in the right "to take life or let live", the right to decide who would live and who must die. In a lecture in 1976 he remarked that the atomic bomb was the great paradox of modern times: a power that claims to care for life holds a weapon that can wipe out all of that life.

Donna Haraway (born 1944) is an American scholar known above all for two works. In "A Cyborg Manifesto" (1985) she wrote about the fading boundary between human beings and machines. In "Situated Knowledges" (1988) she criticised the gaze that claims to see everything while standing nowhere, and insisted that all knowledge must know where it stands and take responsibility for what it sees.

Part Four read this doctrine through the lens of both thinkers. The two letters below are a younger writer's reply to two elders, a conversation between someone who comes later and those who went before, about the very button both of them once thought about.

To Professor Michel Foucault. Professor, you are confusing the fellow who holds the knife with the one who cast the block of stone. Taking life or letting live is only a play staged on the flat surface of linear time, where the jailers imagine they are writing the script. In this frozen block of crystal, no one can kill anyone, and no one can grant anyone the favour of living; every coordinate was settled before you had the chance to define it, and even the question you put to your age, who has the right to take life and who is let live, was carved into its coordinate thousands of years ago. Supreme power does not lie in strangling others, but in the sovereign spirit of one who knows full well that everything is already determined and yet sits enthroned, entirely at ease, upon a throne of his own, holding full au-

thority over every millimetre of those coordinates without asking leave of any god, king or institution. It was from that throne that I wrote this doctrine.

To Professor Donna Haraway. Professor Haraway, you have struck right at the dark heart of the fellows who want to hand the nuclear button to an algorithm. Satellites, radar and AI models truly are things that see everything from nowhere; they calculate the odds of a pre-emptive strike in milliseconds without shedding a drop of sweat, because a machine has no blood to spill and no city to lose. But nuclear deterrence has never been a problem of vision or probability. It is a confrontation between two parties who both stand somewhere, in flesh and blood, and who both know the fear of death. I oppose handing the button to AI not out of faith in compassion or in some bioethics, as people tend to assume, but because of a harsher law: whoever is not inside the block of crystal to bear the consequences has no authority to give the order, for this earth does not belong to machines; it is the property of humankind. When a bodiless machine demands to end the game on humanity's behalf, that is not the evolution of modern war. It is absolute irresponsibility disguised as the speed of light.

Tony T. Nguyen | Nguyễn Từ Vĩnh Trường Highland Reflexive Press

About the Author

Tony T. Nguyen | Nguyễn Từ Vĩnh Trường is an irreverent scholar in borrowed academic robes, and makes no apology for it. By day Tony files nails and sets eyelash extensions in the high desert of West Texas; by night he is an independent researcher in AI safety and the founder of Highland Reflexive Press, where he publishes the Frontier Docket, a running series of replies to the debate on governing artificial intelligence, along with a body of work on one recurring question: how to keep decisions in human hands as machines grow faster than the people they serve.

He is the author of a number of United States patent applications, among them D-CASIS, on controlling the moment at which an AI agent's actions take effect; Homeostatic AI, on self-regulating artificial intelligence; the Reflexive Intelligence System, on controlled self-modification; The Judgment Distribution Problem, on how judgement is allocated inside organisations; and the ELEOS portfolio. As Tony Whistledown of The Academy of Supreme Satire he writes bilingual satire in Vietnamese and English. He wrote *The Crystal of Time* (*Khối pha lê thời gian*) in both English and Vietnamese, and he is also the author of the Vietnamese manuscripts *Sa mạc sau mưa* (The Desert After Rain) and *Nghệ thuật dũa tâm* (The Art of Filing the Mind), the last a book of true stories from the salon that supplied this doctrine with its curing lamps and its beautiful, empty bottles of polish.

He wrote The M.I.A.M Doctrine because of one image he could not shake: waking up one morning, looking out toward the horizon, and seeing a mushroom cloud rising in the distance. His first reaction was an expletive that cannot be printed here. His second was to argue back: no, this earth is far too beautiful, there is no way we let that day come. For him that beauty is nothing distant. It is his Mother still standing at the stove in the corner of the kitchen, while out in the garden there is sunlight, wind, birdsong, the sound of his own cursing and the barking of his dogs on an ordinary morning. This whole doctrine exists to keep that morning where it is.

Sources

- AI Incident Database (n.d.). 1983 Soviet nuclear false alarm.
- Arms Control Association (2025). AI and nuclear command and control: It's even more complicated than you think.
- Brodie, B. (Ed.) (1946). *The Absolute Weapon: Atomic Power and World Order*. Harcourt, Brace.
- Bulletin of the Atomic Scientists (2019). Strangelove redux: US experts propose having AI control nuclear weapons.
- Center for Arms Control and Non-Proliferation (n.d.). AI should help nuclear decision-makers, but not make decisions.
- Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*. Gallimard.
- Foucault, M. (1976). *Histoire de la sexualité, tome 1: La volonté de savoir*. Gallimard.
- Foucault, M. (2003). *Society Must Be Defended: Lectures at the Collège de France, 1975–1976*. Picador.
- Gaddis, J. L. (1986). The long peace: Elements of stability in the postwar international system. *International Security*, 10(4), 99–142.
- Hanyok, R. J. (2001). Skunks, bogies, silent hounds, and the flying fish: The Gulf of Tonkin mystery, 2–4 August 1964. *Cryptologic Quarterly*, Winter 2000/Spring 2001. Declassified 2005, National Security Archive.
- Haraway, D. (1985). A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s. *Socialist Review*, 80, 65–108.
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575–599.
- Hendrycks, D., Schmidt, E., & Wang, A. (2025). *Superintelligence Strategy*. arXiv:2503.05628.
- History Hit (n.d.). Vasili Arkhipov: The Soviet officer who averted nuclear war.
- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge University Press.
- Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson.
- Sagan, S. D. (1993). *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton University Press.
- Sagan, S. D., & Suri, J. (2003). The madman nuclear alert: Secrecy, signaling, and safety in October 1969. *International Security*, 27(4), 150–183.
- Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press.
- Schelling, T. C. (1966). *Arms and Influence*. Yale University Press.
- Task & Purpose (2018). The top US general in Vietnam readied nuclear weapons for use at Khe Sanh.
- The Weather Network (n.d.). The day a meteorological rocket almost triggered World War Three.
- Treaty on the Non-Proliferation of Nuclear Weapons (1968). Opened for signature 1 July 1968, entered into force 1970, Article VI.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
- War on the Rocks (n.d.). How close did the United States actually get to using nuclear weapons in Vietnam in 1968?
- Wikipedia (n.d.). Operation Vulture.

- Wikipedia (n.d.). Norwegian rocket incident.
- Wikipedia (n.d.). Stanislav Petrov.
- Wohlstetter, A. (1959). The delicate balance of terror. *Foreign Affairs*, 37(2), 211–234.
- Canadian Pugwash Group (n.d.). 20 mishaps that might have started a nuclear war.
- Lewis, P., Williams, H., Pelopidas, B., & Aghlani, S. (2014). *Too Close for Comfort: Cases of Near Nuclear Use and Options for Policy*. Chatham House.
- National Security Archive (2012). The 3 A.M. phone call: False warnings of Soviet missile attacks during 1979–80.
- Union of Concerned Scientists (2015). The moon and nuclear war.