It Was Always Going to Be a Cold War: Why Artificial Intelligence Could Not Have Been Fusion, and What the Nuclear Path Leaves Available

Travis Gilly
ORCID: 0009-0007-2954-6313
Real Safety AI Foundation
Illinois, United States

❧ Working paper, September 2026 (v0.4)

ABSTRACT

Artificial intelligence, it is often said, should have been developed as a worldwide cooperative pursuit from the start, on the model of fusion energy research, rather than as a competition between states and firms. This paper argues that the cooperative path was structurally unavailable, and that the reason it was unavailable is also the reason the failure to take it was foreseeable. The paper introduces two terms for the windows that govern whether a technology can be internationalized before it is contested. The symmetry window is the period during which no party holds a lead, so that an agreement to pool development costs each party nothing. The salience window is the period during which the technology's stakes are visible enough that governments will spend political capital on it. For fusion the two windows coincided for decades, because a technology with no weapon and no product never became salient, and cooperation stayed free. For fission the symmetry window closed at Trinity and the salience window opened at Hiroshima, three weeks later and in the wrong order; the Baruch Plan of 1946 was the worldwide pursuit proposed into that gap, and it failed because the party with the lead could not offer terms the party without one could accept. Artificial intelligence carries both a military gradient and a commercial gradient, so it closed its symmetry window faster than fission did, and it did so in full view of a published literature that had already modeled the race. The paper reconstructs the AI symmetry window as roughly 2014 to 2017, shows that it was closed by a national development plan rather than by a technical event, and reads the nonbinding declaration of 2023 as the Baruch outcome arriving on schedule. It then argues that open publication is the proliferation path rather than the cooperative one, drawing on the born secret doctrine of nuclear information law, and that the only completed precedent, nuclear arms control, ran in the order race, parity, verification. What that precedent leaves available for artificial intelligence is verification at the compute and evaluation layers, where the fissile material analogy holds, and not at the model layer, where it breaks. The declaration half of a compute layer instrument already exists in the European Union's Artificial Intelligence Act; the paper identifies the evaluation half as the part the genome precedent supplies, and answers two published positions that read the AI race as a trust dilemma or as an innovation race rather than an arms race.

Keywords: symmetry window; salience window; AI arms race; security dilemma; focusing events; Baruch Plan; non-proliferation; compute governance; export controls; precautionary governance; technology policy; harm blindness

I. Introduction

The complaint is familiar and it is usually stated as a regret. Artificial intelligence should have been a worldwide pursuit from the beginning, the way fusion energy has been, with the major powers pooling their research inside a shared institution instead of racing one another to deploy. Had that been done, the argument runs, the safety problem would be a collective engineering problem rather than a competitive one, and the present situation, in which laboratories and states treat every safety constraint as a concession to a rival, would not have arisen.

This paper takes the regret seriously and argues that it is misplaced in one direction and understated in another. It is misplaced because the fusion path was never open to artificial intelligence. The conditions that made fusion a cooperative enterprise for forty years were structural features of fusion, and artificial intelligence has the opposite features on every axis that matters. It is understated because the unavailability of the fusion path was knowable in advance, from a precedent that was already forty years old, and from a formal literature that modeled the race before the race had a national participant. The failure was foreseeable, it was foreseen in print, and it happened anyway. That combination, a structural trap with a published warning, is the subject of this paper.

Two terms carry the argument and are defined here before they are developed. The *symmetry window* is the period during which no party holds a lead in a technology, so that an agreement to develop it jointly costs each party nothing relative to developing it alone; inside that window an agreement asks no one to give anything up. The *salience window* is the period during which a technology's stakes are visible enough that governments will spend political capital on it; before that window opens, the technology has no place on any agenda, however well it has been modeled. The paper's central claim is that for a technology with a competitive gradient these two windows do not overlap, because the event that makes the technology salient is the same event that produces a leader, and that fusion is the exception only because it never became salient at all. Section II develops both terms and states the claim in full.

The argument has three parts. Section II sets out the mechanism: two windows, one of symmetry and one of salience, that must overlap for a technology to be internationalized before it is contested, and a taxonomy of three technologies (fusion, fission, artificial intelligence) showing why they overlapped for one, missed for the second, and missed faster for the third. Section III reads the Baruch Plan of 1946 as the historical instance of a worldwide pursuit proposed after the symmetry window had closed, and identifies the sequencing problem that made it unacceptable to the side without the lead. Section IV brings the mechanism to artificial intelligence: the symmetry window can be dated, the race was modeled inside it, and the window was closed by a policy document rather than by a discovery. Section V resolves the tension the dating creates, between a window that was open and a path that was unavailable, and shows that the resolution is a harm blindness diagnosis. Section VI addresses the reflex that treats open publication as the cooperative alternative and shows, with the help of nuclear information law, that it is the proliferation alternative. Section VII asks what the one completed precedent, nuclear arms control, leaves available, answers that it leaves verification at two layers and forbids it at a third, and locates the first half of that instrument in legislation already in force. Section VIII states five limits of the argument and replies to two published positions that read the race

differently; Section IX names the work the limits open. The conclusion returns to the regret and restates it as a diagnosis.

A word on what the paper does not claim. It does not claim that cooperation is impossible now, or that arms control for artificial intelligence is futile. It claims that cooperation of the fusion kind, pooled development before contest, was never available, and that the arms control which is available comes in a fixed order and at a fixed layer. Nor does it claim that the actors who declined the symmetry window were irrational. On the contrary: the mechanism works because each actor's choice was locally rational, which is what makes it a trap rather than a mistake.

II. Two Windows and Three Technologies

A. The Symmetry Window

Call the symmetry window the period during which no party holds a lead in a technology, so that an agreement to develop it jointly costs each party nothing relative to developing it alone. Inside the window the game is not a prisoner's dilemma, because there is no advantage to defect toward. Outside it, the leader gives up an advantage by agreeing, the follower gives up the chance to catch up by agreeing, and each side's preferred sequencing is the other side's loss.

The security dilemma literature already states the conditions under which cooperation survives anarchy. Jervis (1978) showed that cooperation is more likely when the cost of being exploited and the gain from exploiting others are both low, when the gains from mutual cooperation and the costs of mutual defection are both high, and when each side expects the other to cooperate. The symmetry window is the period in which the first pair of conditions holds by construction: there is nothing to exploit and nothing to be exploited for. It is the only period in which an agreement can be reached without either side accepting an asymmetric loss.

B. The Salience Window

Call the salience window the period during which a technology's stakes are visible enough that governments will spend political capital on it. Agenda setting runs on focusing events rather than on foresight: the sudden, striking occurrences that move an issue onto the agenda by making its stakes undeniable (Birkland and Warnement, 2017). Not every event that could focus attention does; the events that reach the agenda are disproportionately manmade, large in visible consequence, and close to the deciding body (Alexandrova, 2015). A technology that has not yet produced a focusing event is, for agenda purposes, a technology that does not exist.

C. The Non-Overlap Claim

The claim on which the rest of the paper rests is that for any technology with a competitive gradient, the two windows do not overlap, and that the reason is not contingent. The event that makes a technology salient, a demonstration that it works and that it matters, is the same event that produces a leader, because someone performed the demonstration. Salience and asymmetry arrive together. By the time governments are willing

to spend on an agreement, there is already a party for whom the agreement is a loss.

There is one exception, and it defines the class of technologies for which the fusion model works: a technology that never produces a focusing event, because it has no weapon and no product, never closes its symmetry window, and cooperation stays free indefinitely.

D. Three Technologies

Fusion is the exception. It was set in motion as a cooperative enterprise at the Geneva summit of November 1985, when Gorbachev proposed a joint project to Reagan, and the resulting ITER Agreement was not signed until November 2006, by seven Members who then built the machine largely through in kind contribution (ITER Organization, n.d.; Ikeda, 2010). The 2009 reference schedule aimed at first plasma in 2018 (Ikeda, 2010); the machine has not yet produced it. Forty years of cooperation passed without a defection, and the reason is visible in the schedule: at no point in those forty years could any party have sold a fusion reactor next quarter or fielded one as a weapon. The gain from exploiting the collaboration was zero because there was nothing to exploit. Fusion stayed inside its symmetry window because it never left the pre-salience state.

Fission is the first miss. The Trinity test on July 16, 1945 closed the symmetry window; the bombing of Hiroshima on August 6 opened the salience window (Office of Scientific and Technical Information, n.d.-a, n.d.-b). The two events came three weeks apart and in the wrong order. Fission had a weapon and no product, and the weapon alone was enough. Section III takes up what happened when a worldwide pursuit was proposed into that gap.

Artificial intelligence is the second miss, and it is faster. It has a weapon gradient, because the same systems that write code and plan logistics can be turned to cyber operations and target selection, and it has a commercial gradient, because the model that a laboratory trains is the product that it sells. Fission had one gradient and needed three years to run from the Baruch Plan to the first Soviet test on August 29, 1949, which ended the American monopoly (Alario and Freudenburg, 2007). Artificial intelligence had two, and Section IV shows the corresponding interval running in less time and with less friction.

III. The Baruch Plan as the Rejected Worldwide Pursuit

The proposal that artificial intelligence should have been a worldwide pursuit from the start has been made before, for the last technology with a comparable profile, by the party holding the lead. On June 14, 1946, Bernard Baruch presented to the United Nations Atomic Energy Commission a plan under which all atomic development would be placed under an international authority with rights of inspection and no Security Council veto over its enforcement, after which the United States would dispose of its weapons (Gerber, 1982). It is the closest thing in the historical record to the institution the regret imagines: pooled development, international ownership, inspection as the price of membership.

The Soviet Union refused. The reason it refused is the reason the plan could not have succeeded, and it is a sequencing reason rather than an ideological one. Accepting inspection while the other side held the only weapons meant freezing the lead in place; the Soviet counterproposal reversed the order, disarmament

first and inspection afterward, which the United States could not accept because it meant giving up the lead before verification existed (Gerber, 1982; Mueller, 2015). Gerber, quoting Nogee and Spanier, sets out the trap in the terms that matter here: had the Soviets agreed, they would have accepted permanent military and therefore political inferiority; when they declined, they assumed responsibility for the Cold War (Gerber, 1982). Ulam's observation, quoted in the same article, generalizes it: even had atomic energy carried no military significance at all, the Soviet state could not have accepted an international agency sending inspectors into its economy (Gerber, 1982). The dispute over inspection then outlived the plan; the disarmament proposals of the following decade failed on the same ground, and in 1955 the Soviet side rejected on site inspection paired with a production pause precisely because the pause would have preserved the American quantitative lead (Mueller, 2015).

Two features of the episode carry into the argument. First, the worldwide pursuit was offered by the leader, in good faith by at least some of its authors, and it still failed, because good faith does not change the payoff structure. Second, the failure was not a failure to imagine cooperation. The institution was designed, tabled, and debated. It failed because it was tabled after the symmetry window had closed, and no sequencing exists that lets a leader and a follower enter a pooled arrangement at the same time without one of them accepting a loss. That is the whole mechanism, and it was complete by 1947.

IV. The Artificial Intelligence Symmetry Window

If the mechanism is right, the question for artificial intelligence is not whether there was a symmetry window but when it opened, when it closed, and what was on the table while it was open.

A. Opening

The window could not open before the stakes were arguable in print, because before that there was nothing to agree about. That condition was met in 2014, when Bostrom's *Superintelligence* (2014) put the governance stakes of advanced systems into print at book length, and the formal model followed: Armstrong, Bostrom and Shulman (2016) built a game theoretic model of an artificial intelligence development race in which competing teams trade safety against speed, with the outcome depending on the number of competitors and on how much each knows about the others' capability. Bostrom (2017) then drew the strategic consequence for openness directly: the larger the pool of competitors, the harder it becomes for all of them to coordinate on avoiding a risk race to the bottom, and in a close race a team that unilaterally accepts a safety handicap may simply forfeit the lead. The security dilemma was written down for artificial intelligence, in the technical literature, before any state had adopted the technology as a national objective.

At the same moment, the institutional experiment that the regret imagines was run. OpenAI was founded in December 2015 as a nonprofit research company whose goal, in its own words, was to advance digital intelligence in the way most likely to benefit humanity as a whole, unconstrained by a need to generate financial return (OpenAI, 2015). That is the fusion model: pooled, open, and built so that no single actor would own the result. It is the closest thing artificial intelligence has to ITER, and it was founded inside the symmetry window.

B. What Was on the Table

In January 2017 the Asilomar AI Principles, signed by researchers across the major laboratories, included a race avoidance principle stating that teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards (Future of Life Institute, 2017). The commitment was explicit, the signatories were the people who would later run the race, and the document had no enforcement mechanism of any kind. It was a Baruch Plan without the inspection clause, which is to say a statement of what everyone would prefer if no one had a lead. The research community read the situation the same way at the time: Cave and ÓhÉigeartaigh (2018) assessed the race rhetoric then spreading through industry and government, noted that China had announced ambitious goals for global leadership, and identified corner cutting on safety and governance as the risk a real race would carry.

C. Closing

The window closed in the summer of 2017, and it was closed by a policy document rather than by a technical event. In July 2017 the State Council of the People's Republic of China issued the New Generation Artificial Intelligence Development Plan, setting staged goals through 2030 under which some technologies and applications would lead the world and artificial intelligence would become the main driver of the country's industrial upgrading (State Council, 2017). From that date artificial intelligence was a declared national objective of a major power, and any pooling arrangement would have required that power to accept constraints on an objective it had just announced. The technical event usually credited with starting the modern era, the transformer architecture, was posted in June of the same summer (Vaswani et al., 2017), but the transformer did not close the window; the plan did. A technology becomes contested when a state says it is contested.

The institutional experiment closed on the same schedule. In March 2019, less than four years after its founding, OpenAI created OpenAI LP, a hybrid it described as a capped profit company in which investors and employees could receive a capped return (OpenAI, 2019), and it became a race participant; the pooled model slid down the commercial gradient without anyone having to betray it, because the structure contained no wall against the gradient. Fission took three years from the Baruch Plan to the first Soviet test. Artificial intelligence, carrying two gradients, took about the same time to go from a pooled nonprofit and a signed race avoidance principle to a national development plan and a converted laboratory.

D. The Baruch Outcome, on Schedule

What arrives after the symmetry window closes is predictable from 1946, and it arrived. The first state level agreement on frontier artificial intelligence, the Bletchley Declaration of November 2023, was signed by 28 countries and the European Union, including the United States and China, and it committed its signatories to nothing binding: an expression of shared concern about serious and even catastrophic harm from frontier systems, and an intention to meet again (United Kingdom Government, 2023). It came a year after the public demonstration that made the technology salient to governments, the release of ChatGPT on November 30, 2022 (OpenAI, 2022), and six years after the lead had been declared. It is the Baruch outcome without the Baruch Plan: the parties gathered in the room where a worldwide pursuit might have been agreed, found that

one of them held the lead, and produced a declaration.

V. The Failure Was the Inevitability

The two preceding sections pull against each other, and the tension has to be resolved rather than managed. If the symmetry window was open from 2014 to 2017, and an agreement inside it would have cost nobody anything, then the cooperative path was available and its non-adoption was a failure. If the mechanism of Section II is right, the cooperative path was never available and no one failed. Both cannot be true as stated.

The resolution is that the mechanism does not make agreement impossible inside the symmetry window; it makes agreement impossible inside the symmetry window *on the basis of salience*. An agreement reached in 2015 would have had to be reached on prediction alone, because nothing had happened yet. The models existed; the focusing event did not. The agenda literature is clear about what that means: an issue that has not produced a focusing event does not reach the agenda, however well it has been modeled (Birkland and Warnement, 2017). So the structural claim and the failure claim are the same claim. The structure guarantees that the salience window opens after the symmetry window closes, unless a government acts on prediction alone, and acting on prediction alone is the one thing the agenda process is built not to do. The failure is the inevitability. It is inevitable because governments do not act on models, and it is a failure because the models were right and were available.

This is a harm blindness diagnosis in a game theoretic setting. The Harm Blindness Framework identifies a recurring sequence in which harm is recognized inside an institution before deployment, the recognition is documented, and the recognition fails to convert to action because the people who would bear the cost of preventing the harm cannot yet see it. The artificial intelligence symmetry window is that sequence at the scale of states. The race dynamic was recognized in the formal literature, the recognition was documented in a signed principle, and nothing converted, because the harm the models described was not yet visible to any government that would have had to pay for preventing it. When it became visible, with the ChatGPT release of November 2022 (OpenAI, 2022), the lead had already been declared for five years.

The title of this paper is therefore a claim about structure. The cold war was always coming in the sense that matters: it could not have been avoided *through the mechanism by which governments ordinarily act*. It could have been avoided through the mechanism by which they do not. That is the failure the title records.

VI. Openness Is the Proliferation Path

There is a standard rejoinder to everything above, and it is usually offered as the cooperative alternative: if pooled development was unavailable, publish everything, openly, to everyone, so that no actor holds a monopoly. The rejoinder is wrong for artificial intelligence in a way that it was not wrong for genomics, and the difference is instructive.

A. The Genome Precedent

The Human Genome Project is the one large public scientific enterprise that faced a commercial race and won it, and it did so through a publication rule rather than through a treaty. At Bermuda in 1996 the sequencing centers agreed that all human genomic sequence information from large scale sequencing should be released freely and placed in the public domain, on a daily basis (Kabata and Thaldar, 2023; Collins, Morgan and Patrinos, 2003). When Celera Genomics entered in 1998 with a faster method and a plan to release data quarterly and hold proprietary databases, the public consortium's daily release rule held the line, because a private actor cannot monopolize what is already public; the race ended in a joint announcement in June 2000 (Kabata and Thaldar, 2023). The genome model, stated precisely, is that the cooperative pool publishes fast enough that racing it is pointless. That is a model with a wall in it, and it is the model the regret should have been reaching for instead of fusion.

B. Why It Does Not Transfer

The genome model transfers to artificial intelligence exactly as far as the published object is inert. Sequence data is inert; publishing it within 24 hours carried no proliferation cost. A frontier model's weights are the opposite. A downloaded set of weights can be retrained, stripped of its safety training, and redeployed by any actor with the hardware, and the actor with the least restraint gains the most. Bostrom (2017) made the strategic point before the technology made it concrete: openness about source code, science, and capability tends to tighten the competitive situation around the introduction of advanced systems, raising the probability that winning the race becomes incompatible with any safety method that incurs a delay. Open publication of the frontier is the fastest proliferation path, and it removes the one lever, control over the object, that any subsequent arms control would need.

C. Born Secret

Nuclear law reached this conclusion in 1946 and has held it since, and the doctrine is worth stating because it names a distinction that artificial intelligence policy has not yet drawn. The Atomic Energy Act defines Restricted Data as all data concerning the design, manufacture, or utilization of atomic weapons, the production of special nuclear material, or the use of special nuclear material in the production of energy, excepting only data formally declassified (42 U.S.C. § 2014(y)). The definition turns on subject matter, not on origin: the information is restricted from the moment it exists, wherever it exists, which is why the doctrine is called born secret. In *United States v. Progressive, Inc.* (1979), a federal court enjoined a magazine from publishing an article on hydrogen bomb design even though much of its content had been assembled from public sources, finding that the article provided a more comprehensive and accurate account of the weapon's construction than anything then in the public literature, and that its publication could materially reduce the time required by certain countries to achieve a thermonuclear weapon capability (*United States v. Progressive, Inc.*, 467 F. Supp. 990 (W.D. Wis. 1979)). The court weighed the disparity of risk, the prospect of proliferation against the presumption against prior restraint, and found the presumption overcome.

The point here is narrower than a proposal to classify model weights under an Atomic Energy Act for

artificial intelligence. The nuclear regime distinguishes between the material, which is countable and can be safeguarded, and the information, which once published cannot be recalled and which the law therefore treats as the proliferant in its own right. Artificial intelligence has the same two layer structure with the layers named differently: compute is the material, and weights are the information. A governance regime that publishes the weights and then proposes to verify the compute has already given away the thing the verification was for.

VII. What the Completed Path Leaves Available

Only one technology with a competitive gradient has run the full course from contest to a functioning international control regime, and it did so in a fixed order. The order is the finding, because it says what is available now and what is not.

A. Race, Parity, Verification

The atomic race began when the Baruch Plan failed and ran until the Soviet test of August 1949 produced parity (Alario and Freudenburg, 2007). The institutions came afterward and only afterward: Eisenhower's Atoms for Peace address to the United Nations on December 8, 1953, which proposed a new international agency to monitor proliferation and regulate trade in nuclear materials; the International Atomic Energy Agency, founded in 1957 (Alario and Freudenburg, 2007); and the Treaty on the Non-Proliferation of Nuclear Weapons, opened for signature in 1968 and in force in 1970 (United Nations Office for Disarmament Affairs, n.d.; International Atomic Energy Agency, 1970), under which states without nuclear weapons accept Agency safeguards on their nuclear material. Every one of those institutions was built after both sides already held the capability. None of them ran on denial. The Cold War's arms control worked because it verified parity. The denial instruments belong to a separate argument, and this paper takes as a premise that constraint imposed on a rival tends to accelerate the rival's innovation rather than arrest it. The sequencing that the Baruch Plan could not solve, inspection before or after disarmament, dissolved once there was nothing left to disarm asymmetrically, and the verification institutions were built on that dissolved problem.

The nuclear precedent has been read for artificial intelligence before, and the reading agrees on the order. Maas (2019) draws three lessons from nuclear weapons for military AI arms control: that institutionalized norms can slow proliferation, that organized epistemic communities of experts can catalyze arms control, and that many applications will remain susceptible to normal accidents, so that assurances of meaningful human control are largely inadequate. All three are post-parity instruments; none of them is a pooled development regime. Read against artificial intelligence, the sequence is a warning against one instinct and an instruction for another. The instinct it warns against is the technical working group convened to prevent the race, which is an ITER instrument applied to a technology that already has a weapon and a product; it is the right institution for the wrong decade. The instruction it gives is that verification is the achievable object, that it becomes achievable at parity, and that a state which understands this should be designing the verification regime now, for deployment when parity arrives, rather than spending the interval on denial that accelerates the party it is aimed at.

B. Where the Fissile Analogy Holds and Where It Breaks

Nuclear safeguards work because fissile material is physical, countable, and rare, and the regime's own history shows what happens when the count slips: the 1953 Oppenheimer report concluded that the growth in fissile production had made disarmament agreements nearly impossible to verify because the material could no longer be fully accounted for (Alario and Freudenburg, 2007). Accountancy is the regime; when accountancy fails, so does the regime. The safeguards agreements built under the Treaty are agreements about nuclear material accountancy: what a state holds, where it is, and whether the books balance (International Atomic Energy Agency, 1972). The analogy to artificial intelligence therefore holds exactly where the object is physical and countable, and breaks exactly where it is not.

Compute is countable. Frontier training runs require chips that are manufactured in a small number of facilities, shipped, installed, and powered, and every one of those steps leaves a record that an inspector can audit. This is why export controls on advanced chips exist at all, and it is why a verification regime at the compute layer is the direct descendant of material accountancy: declare the training runs, meter the hardware, and let the books be balanced by a third party.

Evaluations are auditable. A completed model can be put through a battery of capability and safety evaluations by an independent body, and the results can be disclosed on a schedule, the way test ban regimes disclosed seismic data. This is the layer where the genome model's publication rule survives its transfer: publish the safety findings and the evaluation results on a clock, so that the cooperative pool's knowledge of what the frontier can do is always ahead of any single actor's claim about it.

Weights are neither. They are copied without leaving a residue, moved without crossing a border, and modified without a facility. There is no accountancy for weights, and any regime that assumes otherwise will be verifying the two layers it can while the third leaks. The placement of verification follows from that asymmetry. The nuclear regime did not try to inventory the equations; it inventoried the plutonium and restricted the equations. The equivalent for artificial intelligence is to inventory the compute, audit the evaluations, and keep the weights inside the regime rather than publishing them.

C. The Instrument Already Exists in Draft

The compute layer is not a hypothetical. The European Union's Artificial Intelligence Act already contains a declared training run instrument. Under Article 51(2), a general purpose model is presumed to have high impact capabilities, and so to carry systemic risk, when the cumulative computation used for its training exceeds 10^{25} floating point operations; under Article 52(1), a provider whose model meets that condition must notify the Commission within two weeks; and under Article 51(3), the Commission may amend the thresholds by delegated act to track algorithmic improvements and hardware efficiency (European Union, 2024). That is material accountancy translated into compute: a declared quantity, a reporting duty, a threshold, and a mechanism for revising the threshold as the technology moves. It is unilateral, it is a notification rather than an inspection, and it verifies nothing about what the model can do. But it is the first half of the instrument, and it is in force. The second half, an independent evaluation layer reporting on a fixed clock, is what the genome model contributes, and the two together are the regime this paper proposes.

D. The Modified Genome Rule

Put together, the available regime is the genome model with a proliferation wall. Pool the compute under declared and metered training runs. Release the evaluation results and the safety findings on a fixed clock, so that racing the pool's knowledge is pointless. Keep the weights inside the consortium. That combination gives up nothing the nuclear regime did not also give up, and it takes the one part of the fusion model that was ever transferable, the shared institution, and places it at the layer where a shared institution can hold.

VIII. Limitations

The argument has five boundaries that a careful reader will find, and they are stated here in the author's terms.

First, the dating of the symmetry window is an argument, not a measurement. The window is placed at 2014 to 2017 on the ground that the stakes were arguable in print from 2014 and a major power declared the technology a national objective in 2017. A reader who dates the corporate lead earlier, to the 2012 demonstration that began the deep learning era, can shorten the window to two years or close it before the formal race models appeared. The argument survives a shorter window, because a shorter window makes the non-overlap claim stronger, but the specific dates should be read as the author's placement.

Second, the mechanism predicts the outcome for technologies with a competitive gradient, and it is confirmed by the two cases that have one, fission and artificial intelligence, and by the one case that does not, fusion. Three cases is a pattern, not a law. Other dual gradient technologies, recombinant biology among them, are not examined here, and a reader could reasonably ask whether the mechanism holds for them or whether the biology community's early self-regulation is a counterexample. The paper does not settle that.

Third, the claim that verification at the compute layer is the direct descendant of material accountancy assumes that frontier capability continues to require frontier compute. If algorithmic efficiency decouples capability from hardware faster than the paper assumes, the countable layer stops being the load bearing one, and the argument's constructive section loses its ground while its diagnostic sections do not. The Artificial Intelligence Act's delegated act mechanism for revising its compute threshold (European Union, 2024) is an acknowledgment of the same risk by the one legislature that has written the instrument down, and it is a mechanism for chasing the decoupling, not for preventing it.

Fourth, two published positions cut against the paper's framing and deserve naming. Katzke and Futerman (2024) argue that a race to artificial superintelligence is a trust dilemma rather than a prisoner's dilemma, so that cooperation to control development is both preferable and achievable; Schmid, Lambach and Diehl (2025) argue from government documents that the arms race metaphor misdescribes AI competition, which is better read as a geopolitical innovation race in which actors are open to positive sum approaches and prioritize economics and status alongside security. This paper's answer to both is the same: a trust dilemma still has to be solved inside a window in which trust costs nothing, and an innovation race still has a leader. Neither position identifies a mechanism by which the symmetry and salience windows are made to overlap, and until one is identified, they describe the preferable outcome rather than the available one. That is a reply, not a refutation, and the reader should weigh it as such.

Fifth, the paper treats governments as the relevant agents for the salience window and firms as the relevant agents for the symmetry window, which is a simplification. The firms that signed the race avoidance principle in 2017 were the parties whose behavior the principle was meant to govern, and the paper reads their later conduct as the gradient operating on a structure with no wall. A reader who attributes more agency to those firms will read the same conduct as a choice, and the paper's inevitability claim weakens for the corporate case even where it holds for the state case.

IX. Future Work

Two items follow directly from the limitations above.

The second limitation is addressable by extending the three technology taxonomy to the technologies it omits. Recombinant DNA is the obvious candidate: it had a research community that self-regulated early, and whether that self-regulation held once a commercial gradient appeared is a question that the framework makes precise and that the historical record can answer. If the biology case fits the mechanism, the pattern becomes four cases; if it does not, the framework has found its boundary condition.

The third limitation is addressable by specifying the compute layer verification regime in enough detail to say what it would measure. The Artificial Intelligence Act supplies the declaration half of the instrument; the audit half and the evaluation clock remain unspecified. A design that stated what a declared training run consists of, what a metering audit would inspect, and what disclosure clock the evaluation layer would run on would convert the paper's constructive section from a placement into a proposal, and would expose whether the fissile analogy holds under the efficiency trends that the third limitation names.

X. Conclusion

The regret that artificial intelligence should have been a worldwide pursuit from the start is correct about the destination and wrong about the road. The fusion road requires a technology that never becomes salient, because only such a technology keeps its symmetry window open long enough for governments to walk through it on their own schedule. Fission had a weapon and closed the window three weeks before the salience that would have prompted an agreement; the agreement was then proposed anyway, by the leader, and could not be accepted by the follower, and the sequencing problem that killed it was complete by 1947. Artificial intelligence had a weapon and a product, closed its window in about three years, and did so while a formal literature that had already modeled the race sat in the journals and a signed race avoidance principle sat in a drawer. The failure was foreseeable, it was foreseen, and it was not acted on, because the mechanism by which governments act requires an event and the event arrived after the lead.

What remains is the road that one technology has completed. It runs through the race rather than around it, reaches verification at parity rather than before it, and verifies the countable layer while restricting the uncountable one. For artificial intelligence that means metered compute, audited evaluations released on a clock, and weights kept inside the regime; one legislature has already written the declaration half of the first of those. It is a narrower destination than the shared institution the regret imagines, and it is the one the

record supports.

ACKNOWLEDGMENTS

The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All intellectual contributions, research questions, theoretical arguments, methodological decisions, and conclusions are solely the author's own.

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org.

REFERENCES

Alario, M. V., & Freudenburg, W. R. (2007). Atoms for peace, atoms for war: Probing the paradoxes of modernity. *Sociological Inquiry*, 77(2), 219–240. https://doi.org/10.1111/j.1475-682X.2007.00188.x

Alexandrova, P. (2015). Upsetting the agenda: The clout of external focusing events in the European Council. *Journal of Public Policy*, 35(3), 505–530. https://doi.org/10.1017/S0143814X15000197

Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: A model of artificial intelligence development. *AI & Society*, 31(2), 201–206. https://doi.org/10.1007/s00146-015-0590-y

Birkland, T. A., & Warnement, M. K. (2017). Focusing events, risk, and regulation. In E. J. Balleisen, L. S. Bennear, K. D. Krawiec, & J. B. Wiener (Eds.), *Policy shock: Recalibrating risk and regulation after oil spills, nuclear accidents and financial crises*. Cambridge University Press. https://doi.org/10.1017/9781316492635.005

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

Bostrom, N. (2017). Strategic implications of openness in AI development. *Global Policy*, 8(2), 135–148. https://doi.org/10.1111/1758-5899.12403

Cave, S., & ÓhÉigeartaigh, S. S. (2018). An AI race for strategic advantage: Rhetoric and risks. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3278721.3278780

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. *Science*, 300(5617), 286–290. https://doi.org/10.1126/science.1084564

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, consolidated text of 27 July 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02024R1689-20260727

Future of Life Institute. (2017). *Asilomar AI principles*. https://futureoflife.org/open-letter/ai-principles/

Gerber, L. G. (1982). The Baruch Plan and the origins of the Cold War. *Diplomatic History*, 6(1), 69–96. https://doi.org/10.1111/j.1467-7709.1982.tb00364.x

Ikeda, K. (2010). ITER on the road to fusion energy. *Nuclear Fusion*, 50(1), 014002. https://doi.org/10.1088/0029-5515/50/1/014002

International Atomic Energy Agency. (1970). *Treaty on the Non-Proliferation of Nuclear Weapons: Notification of the entry into force* (INFCIRC/140). https://www.iaea.org/sites/default/files/publications/documents/infcircs/1970/infcirc140.pdf

International Atomic Energy Agency. (1972). *The structure and content of agreements between the Agency and States required in connection with the Treaty on the Non-Proliferation of Nuclear Weapons* (INFCIRC/153 (Corrected)). https://www.iaea.org/sites/default/files/publications/documents/infcircs/1972/infcirc153.pdf

ITER Organization. (n.d.). *The ITER story*. https://www.iter.org/proj/iterhistory

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics*, 30(2), 167–214. https://doi.org/10.2307/2009958

Kabata, F., & Thaldar, D. (2023). The human genome as the common heritage of humanity. *Frontiers in Genetics*, 14, 1282515. https://doi.org/10.3389/fgene.2023.1282515

Katzke, C., & Futerman, G. (2024). The Manhattan trap: Why a race to artificial superintelligence is self-defeating. arXiv. https://arxiv.org/abs/2501.14749

Maas, M. M. (2019). How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. *Contemporary Security Policy*, 40(3), 285–311. https://doi.org/10.1080/13523260.2019.1576464

Mueller, B. S. (2015). Waging peace in a disarmed world: Arthur Waskow's vision of a nonlethal Cold War. *Peace & Change*, 40(3), 339–367. https://doi.org/10.1111/pech.12134

Office of Scientific and Technical Information. (n.d.-a). *The Trinity test, July 16, 1945*. Manhattan Project: An Interactive History, U.S. Department of Energy. https://www.osti.gov/opennet/manhattan-project-history/Events/1945/trinity.htm

Office of Scientific and Technical Information. (n.d.-b). *The atomic bombing of Hiroshima, August 6, 1945*. Manhattan Project: An Interactive History, U.S. Department of Energy. https://www.osti.gov/opennet/manhattan-project-history/Events/1945/hiroshima.htm

OpenAI. (2015, December 11). *Introducing OpenAI*. https://openai.com/index/introducing-openai/

OpenAI. (2019, March 11). *OpenAI LP*. https://openai.com/index/openai-lp/

OpenAI. (2022, November 30). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

Schmid, S., Lambach, D., & Diehl, C. (2025). Arms race or innovation race? Geopolitical AI development. *Geopolitics*, 30(4), 1907–1936. https://doi.org/10.1080/14650045.2025.2456019

State Council of the People's Republic of China. (2017). *New Generation Artificial Intelligence Development Plan* (English translation). DigiChina, Stanford University. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

United Kingdom Government. (2023). *The Bletchley Declaration by countries attending the AI Safety Summit, 1–2 November 2023*. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

United Nations Office for Disarmament Affairs. (n.d.). *Treaty on the Non-Proliferation of Nuclear Weapons (NPT)*. https://disarmament.unoda.org/wmd/nuclear/npt/

United States v. Progressive, Inc., 467 F. Supp. 990 (W.D. Wis. 1979).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. arXiv. https://arxiv.org/abs/1706.03762

42 U.S.C. § 2014(y).