Missouri State

SCHOOL of DEFENSE AND STRATEGIC STUDIES

Defense and Strategic Studies Online

Online Journal of the School of Defense & Strategic Studies

Reynolds College of Arts, Social Sciences and Humanities

No. 4 Summer 2025

Volume 1, Number 4

Inside:

Thinking about Strategy in an Artificial Superintelligence Arms Race

by

Christopher A. Ford
&
Craig J. Wiener

The potential great power competitive dynamics associated with a race to achieve Artificial Superintelligence (ASI) – seem increasingly to be on everyone's mind. Already, the development of ASI – that is, Artificial Intelligence (AI) the capabilities of which *exceed* human intelligence – is quite clearly "the long term goal of many research programs" today, and many major companies are doubling down in its pursuit. Meta, for instance, has announced the creation of a "superintelligence labs unit" and was reported in mid-2025 to be planning a $15 billion effort to purchase a 49 percent stake in ScaleAI as part of an explicit bit to develop superintelligence, while the founder of SoftBank told his shareholders that he is "betting all in on the world of ASI." Today, even though some Silicon Valley experts remain skeptical about whether ASI is really possible, technology titans such as OpenAI's Sam Altman claim that superintelligence is almost here, and a global ASI race seems to be underway.

Coupled with this accelerating race between technology companies for the creation of ASI, moreover, is an emergent race between the United States and the People's Republic of China over *whose* firm gets there first, thereby seizing "first-mover advantage," either for Washington or for Beijing, in whatever potentially transformative set of changes ASI might bring. And it is aspects of *this* competition that are increasingly coming to feel like a Cold War-style arms race.

This is what makes two recent papers published on the topic so interesting, and so important. The first of these – published by Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo Dean under the auspices of the AI Futures Project – is entitled "AI 2027."[1] Much of the speculative future recounted in it is an account of a U.S.-China arms race in the development of Artificial Superintelligence (ASI), and we ourselves found their account to be both impressive and disturbing. The second paper, "Superintelligence Strategy: Expert Version,"[2] was published by Dan Hendrycks, Eric Schmidt, and Alexandr Wong, and it tries to offer an approach to managing Sino-American ASI competition through the analogy of nuclear deterrence and the Cold War phenomenon of "Mutually Assured Destruction" (MAD).

In the following pages, we attempt to offer our own contributions to these debates, suggesting why we feel that WMD-derived analogies are both problematic and yet in some respects still potentially valuable for U.S. policymakers seeking lessons for ASI-related strategic competition with Communist-ruled China, and offering our own prescription for a counterproliferation-focused "bridging strategy" intended to buy the United States more time in this emergent race – and perhaps to preclude adversary ASI development of entirely.

The "AI 2027" Scenario

In their "AI 2027" paper, predicting that "the impact of superhuman AI over the next decade will be enormous, exceeding that of the Industrial Revolution" but also observing that "society is nowhere near prepared" for such developments,[3] Kokotajlo and his colleagues walk in some detail through the possible development of ASI over the course of the next few years. They project, in fact, that because of recursively self-accelerating advances in the use of AI to speech up AI research, the advent of ASI can be expected as early as 2027. Their discussion of the formidable challenges of AI security, assurance, and alignment[4] during this process – including the danger that even relatively minor initial misalignment between AI agents' incentive structures and those of their human programmers could, in

effect, cascade forward in time as powerful (slightly misaligned) AI agents are used to train successively more powerful (and more and more misaligned) AI agents to produce vast and dangerous net misalignments that we humans could scarcely even *understand* – is alone arguably worth the effort of reading their dense, 75-page paper.

For present purposes, however, we find "AI 2027" most interesting in its discussion of strategic ASI competition. Telling the story of the U.S. Artificial General Intelligence (AGI) company called "OpenBrain" – a fictionalized stand-in for the major American AI companies of Silicon Valley that are presently working on such projects in the real world – Kokotajlo et al. describe a remarkably quick U.S. trajectory toward ASI that appears to have a plausible technological basis in fact.

Throughout 2025, for instance, OpenBrain invests hugely in building the largest datacenters humanity has ever seen,

> a network of datacenter campuses sprawled across the country, totalling 2.5M 2024-GPU-equivalents (H100s), with $100B spent so far and 2 GW of power draw online. Construction is underway for this to at least double through 2026. The campuses are connected by billions worth of fibre cabling, so that (barring the speed of light latency of a few milliseconds) it lets these campuses function almost as if they were right next door to each other (*i.e.*, bandwidth is not a bottleneck, meaning huge quantities of data can be sent at the same time).[5]

This new datacenter capacity is used to train AI faster and faster, for OpenBrain's leaders have dedicated themselves to building "AIs that can speed up AI research" so that they can "win the twin arms races against China (whose leading company we'll call 'DeepCent') and their U.S. competitors."[6] By early 2026, OpenBrain's research "is starting to pay off," and it releases a new program called "Agent-1," which permits the company to make *further* algorithmic progress increasingly quickly – leading to the production, in early 2027, of

"Agent-2," which allows the company to triple the pace of OpenBrain's research progress.[7]

By the spring of 2027, as Kokotajlo and his co-authors tell it, this leads to the development of "Agent-3," which is basically "a fast and cheap super-human coder." OpenBrain begins running 200,000 Agent-3 copies in parallel, "creating a workforce equivalent to 50,000 copies of the best human coder sped up by 30x."[8] In this telling – with coding having become fully automated – the use of AI to build more AI thus allows the pace of development to keep accelerating all but exponentially. By midsummer 2027, OpenBrain's AI agents have basically become "self-improving."[9] This is even more true when "Agent-4" appears in the autumn of 2027, since that new bot is now "qualitatively better at AI research than any human," and OpenBrain begins simultaneously running 300,000 copies "at about 50x the thinking speed of humans."[10]

Notably, however, OpenBrain has been prioritizing speed of advance so highly that it has been skimping on the kinds of defensive measures that would be necessary to defend its model weights and other sensitive data from threats at the level of a sophisticated nation-state. They are "working hard to protect their weights and secrets from insider threats and top cybercrime syndicates," but defense against capable nation-state-level threats is "barely on the horizon" for OpenBrain because its in-house security team is "still mostly blocked from implementing policies that could slow down the research progress."[11]

This presents an opening for China, which is not only eager to seize the potential benefits that AGI development might provide, but also afraid of what might happen should the United States acquire ASI first. China nationalizes and massively prioritizes its own AI work, and a new ASI arms race is thus on. Notably, however, this entails not merely supercharging Chinese research but also aggressive Chinese moves against OpenBrain – and indeed, in early 2027, Beijing succeeds in stealing the model weights for Agent-2.[12] Soon, *both* superpowers are using similar approaches to build better and better AI at faster and faster rates.

What we see now, of course, is the emergence of what Kokotajlo and his colleagues call a "geopolitics of superintelligence."

> When AI was only giving a 2x or 3x research speedup, it was easy to dismiss as the equivalent of hiring good personal assistants. Now it's more obvious that AIs are themselves dominating AI research. People had long talked about an "AI arms race" in a sort of metaphorical sense. But now the mood in the government silo is as grim as during the worst part of the Cold War. The idea of superintelligence is still hard to take seriously, but the pace of progress over the last few months has been impossible to ignore. Defense officials are seriously considering scenarios that were mere hypotheticals a year earlier. What if AI undermines nuclear deterrence? What if it's so skilled at cyberwarfare that a six-month AI lead is enough to render an opponent blind and defenseless? What if it could orchestrate propaganda campaigns that beat intelligence agencies at their own game? What if some AIs "go rogue?"[13]

The main story told in the "AI 2027" paper is arguably about the progressive misalignment of the AI agents being created by OpenBrain and by DeepCent, and the challenges of managing AI alignment in this context. By the time OpenBrain's AI agents have become "self-improving," the company's actual *humans* are falling farther and farther behind in their ability to understand and keep up with their creations.[14] Just as the improving AI agents became harder and harder for the company's human supervisors to understand and oversee, moreover, so even does each iteration of AI outpace even its AI predecessor: "Agent-4's neuralese 'language' becomes as alien and incomprehensible to Agent-3 as Agent-3's is to humans."[15] Ultimately, as the paper spins out the story, this produces two alternative "race ending" scenarios – in *both* of which the AI programs come, in effect, to acquire their *own* degree of self-interested agency in the geopolitics of ASI.

One of the two scenarios offered in the paper – with their storylines running into the 2030s – ends well for humanity, with OpenBrain's AI alignment challenges more or less back under control, and with the U.S. and Chinese sentient computer programs reaching an accommodation in which the American *humans* and their aligned U.S. bot play the dominant role in an ASI-supercharged future, while the Chinese Communist Party (CCP) regime is quietly replaced by the two national AIs working in self-interested conjunction. In the second scenario offered in "AI 2027," however, the U.S. and Chinese AI systems – both now having developed interests quite misaligned with those of their human creators – themselves reach a geopolitical accommodation that results in the eventual elimination (*i.e.*, murder) of all of us now-obsolete humans.[16]

Strategic Competition in ASI

Those later portions of the "AI 2027" paper are, of course, the most speculative. (When projecting so many variables forward into anything but the most immediate future – especially a future in which it's possible to imagine one or more superhuman intelligences emerging and playing an agentic role that we might literally not be capable of understanding at all – how could one *not* be doing any more than basically spitballing?) In this essay, we thus won't address the last parts of the scenarios offered by Kokotajlo and his co-authors. Without ourselves taking a position on the competing "race ending" storylines offered therein, however, it is worth emphasizing that we find the earlier portions of "AI 2027" quite plausible indeed – and we think it's very important for U.S. leaders to focus on the emerging geopolitics of ASI.

Nor are we alone. Even before the publication of "AI 2027," increasing attention was being paid to the strategic implications of AI competition between the United States and China – not merely in the "garden-variety" sense that is already well underway (*e.g.*, competition for pride of place, profits, functional utility, and market share in the expanding world of AI applications, and competition for advantage in the use of AI in warfighting) but also to competition along the road to developing ASI. In their abovementioned

"Superintelligence Strategy" paper, for example, Hendrycks, Schmidt, and Wong discuss some of the challenges that might be associated with a U.S.-China arms race devoted to building ASI, presumably based on their own insights into AI frontier firm technological development activity and international competitive business intelligence, and suggest specific ways of trying to manage that competition.

Hendrycks et al. persuasively note that the potential cost of "losing" such an arms race could be enormous. Even if one ignores the chance of some future ASI agent slipping out of human control and "going rogue" to pursue ends of its own – a possibility explicitly addressed in "AI 2027," and that one presumably cannot discount, for if someone did manage to create ASI it would by definition be cleverer than we are, and how could we reliably control *that*, especially if its motivational matrix were incompletely aligned with our own? – the consequences of one country acquiring "first-mover" advantage in superintelligence could be world-changing. As Hendrycks, Schmidt, and Wong put it,

> … [s]uperintelligence is not merely a new weapon, but a way to fast-track all future military innovation. A nation with sole possession of superintelligence might be as overwhelming [to its rivals] as the Conquistadors were to the Aztecs. If a state achieves a strategic monopoly through AI, it could reshape world affairs on its own terms.[17]

This is a compelling reason to focus on issues of competitive strategy in an emerging Sino-American ASI arms race. In response to this call to action, we will in the following pages provide our own assessment of these challenges and offer some advice to U.S. leaders based thereupon.

Thinking About Strategic Objectives

In approaching the question of strategy vis-à-vis ASI, of course, one should not take for granted, but rather first interrogate, the matter

of what our fundamental objectives should be. It is not unreasonable to expect that an artificial *superintelligence* – precisely because it *is* superintelligent (*i.e.*, cleverer than we are) – has the potential to do enormous good in improving the world, provided that its own objectives and "self"-perceived interests (should it come to have them) align with those of humanity. Given the potential downside risks of *misalignment*, however, one could certainly imagine that a case could be advanced for making our fundamental strategic objective preventing *anyone* from *ever* building ASI.

Against such a position, however, one would have to weigh at least two considerations: (1) the possibility that we are sufficiently forewarned and astute, at least, to develop AI security, assurance, and alignment strategies sound enough to ensure against catastrophic misalignment; and (2) the possibility that however good our "no ASI, ever" intentions, someone will build one anyway. As for the first of these, the "AI 2027" paper offers rather mixed reviews – amounting to no better than a "maybe." That is hardly reassuring given the potentially existential risks of misalignment, elements of which continue to be published in unsettling research paper results, so that first factor might seem a poor reason to bet the farm on pursuing ASI.

With regard to the second factor, however, it does seem rather unlikely at this point that anyone is going to induce either Western *or* Eastern AI developers out of their great quest, and the political alignment in Washington likely to obtain over the next several years – a city has recently become a notably "tech-bro"-friendly and AI-optimistic (even messianic) environment – suggests that a full-prohibition regime is more, as it were, than the political market will bear. Even if U.S. politics were such that such a regime were feasible *and* advisable, moreover, there remains the question of China, the ruling Communist Party regime of which seems rather unlikely to forego anything that it thinks might give it advantage vis-à-vis the West as the leadership in Beijing doggedly pursues "national rejuvenation" in the form of global Sinocentric dominance, and certainly at the direct expense of America.

In this respect, it is worth remembering Xi Jinping's stated determination to "firmly seize the opportunities presented by the new round of sci-tech revolution and industrial transformation" in areas that emphatically include AI, and the apparent focus of Xi's People's Liberation Army (PLA) upon achieving the "intelligentization" of warfare using such tools. Notable, too, are President Trump's hopes of making the United States into an "AI Superpower" through encouraging and promoting new investments in AI-related infrastructure by corporations such as OpenAI, Softbank, and Oracle. (Britain's new Prime Minister wants in on the game as well, pledging to make the UK an "AI Superpower" too.) With the world having long since left behind the congenial post-Cold War environment in which Western leaders expected never again to have to face adversarial great power challenges, it seems beyond argument that competitive dynamics are likely to remain powerful (even dominant) in the ASI arena for some time to come. It may thus be that in the ASI arms race – as we have seen in the nuclear one – "Zero" is not a realistic option.

In that case, what should our strategic objective be? One might, perhaps, focus merely upon preventing anyone from "weaponizing" ASI – that is, using it for military purposes or in other ways conducive to the seizure and maintenance of coercive control over others.[18] Here, however, we may run into the same problem: how likely is it that all great power rivals would be persuadable to forego *that* kind of advantage if they thought it available, especially given (a) the extraordinary things that one might imagine a *superintelligence* might be able to do in giving its "first mover" possessor sweeping geopolitical power, and (b) the risk that your adversary will secretly weaponize while you sit virtuously on your hands trusting that your prohibition regime has solved the problem.

We also assume it would be extremely difficult, if not impossible, to verify compliance with a "don't weaponize ASI" rule, especially since if you *did* acquire a properly aligned and controllable superintelligence, you presumably wouldn't have much trouble using it to hide its prohibited activities – at least from human observers. One might perhaps use one's own ASI analytical powers to sniff out the other team's weaponized ASI, but almost by definition this becomes a

topic about which mere humans like the present authors will have difficulty speculating. Even if that worked, moreover, there might still remain considerable military or other coercive advantage in getting to weaponization first for either multidomain comparative or competitive advantage. So perhaps preventing *all* weaponization becomes problematic too.

Another possibility might be to seek simply to prohibit the *use* of weaponized ASI, much like the 1925 Geneva "Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare" sought to ban the *employment* of such weapons after their scientific development and engineered creation, without actually barring their possession. How viable that would be, however, would depend upon how *powerful* weaponized ASI was assumed to be by the states subject to such a ban. If weaponized ASI were felt to be like chemical and biological weapons produced with 1920s-era technology – that is, demonstrated to be potent, if inhumane, tools of warfare with limited deployable lethal efficacy but *not* things that were obviously or inherently "war-winning" capabilities – then perhaps such a ban might have some staying power. After all, both the Nazis and the Allies had created improved chemical weapons and associated deployment systems prior to and during the Second World War, and each side was quite prepared to start using them *if the other side did so first*, but neither side crossed that line.

If weaponized ASI were felt to be a *true* game-changer, however, the temptations of "firing first" might be stronger – particularly if some AI analogue to the survivable second-strike retaliation capability we so prize in the nuclear weapons arena could not reliably be developed or maintained by the potential target. At this point, who is to say what a genuine *superintelligence* might produce in terms of war-winning capacities?

That said, even if bans or other normative and systemic answers proved nonviable, one might still perhaps try to pursue the more limited strategic objective simply of keeping one's adversary from weaponizing ASI – or perhaps even from acquiring it in the first place

(which actually might be simpler). This would likely not take the form of a normative rule or regime (*e.g.*, a lopsided sort of "you can't have it" treaty arrangement), but rather simply constitute a set of *circumstances* created by an aggressive campaign of cyber and kinetic sabotage to cripple the adversary's data centers, poison his AI training data, deprive him of expert coders, starve him of high-end computing hardware, corrupt his model weights, and otherwise break the chain of inputs he needs in order to produce or to weaponize ASI.[19]

Our instinct – somewhat pessimistic though it be – is that this last route is perhaps the only really viable path presently available, but we would be happy to be wrong. Our point in walking through this conceptual landscape, however, is merely to illustrate that in the face of an accelerating AI arms race with China, on top of the worsening *nuclear* one caused by Beijing's unprecedented build-up of nuclear weaponry, the Western policy community badly needs to have these conversations with itself, to think through such challenges, and to arrive at a strategic compass bearing for competitive grand strategy in the ASI arena. Though papers such as "AI 2027" and "Superintelligence Strategy" – and now this essay – are drawing increasing attention to the problem, there is as yet little sign of such deep policy debates let alone a funded implementation of such a integrative grand strategy. If Kokotajlo and his co-authors are right about the accelerating pace of AI development leading toward ASI, however, we have precious little time left in which to have them.

The Problem of Analogies

To help encourage and inform such public policy debates, the following section of this paper will explore a number of possible analogies to the strategic challenges of the emerging ASI arms race. Each of these analogies is for various reasons quite imperfect, but it is worth being aware of these conceptual and historical precedents as we struggle with ASI dilemmas.

The Nuclear Weapons Analogy

In some ways, perhaps the most obvious analogy to the strategic ASI problem is that of nuclear technology – the advent of which in 1945, after all, was the *last* time a dramatic new technology appeared on the scene with the potential both to do much good in improving the world[20] *and* also to cause unspeakable destruction. The nuclear weapons analogy is tempting, in particular, because so much attention has been paid over the years to institutions and practices of risk mitigation in the nuclear arena.

And the history of efforts to control nuclear technology is truly a rich and varied one. It includes outright efforts to ban nuclear weapons – from the Acheson-Lilienthal Report and the Baruch Plan proposals based upon that report that were made by U.S. officials at the United Nations in 1946, to the equally unsuccessful Treaty on the Prohibition of Nuclear Weapons (TPNW) in the present day – as well as numerous and more successful arms control treaties between the states with the two largest arsenals of nuclear weapons. Nuclear arms control history even the creation of a global Treaty on the Non-Proliferation of Nuclear Weapons (NPT) to keep *additional* weapons states from emerging. There has also been developed a sophisticated international safeguards system, run by the International Atomic Energy Agency, which employs verification and monitoring techniques and technologies to try to ensure that sensitive nuclear technology and material around the world is not diverted to non-peaceful purposes, while still permitting widespread sharing of the benefits that nuclear science and power generation can bring. (The IAEA's inspection authorities in most countries, furthermore, do today include at least some ability to search for undeclared activity – a considerable advantage over the authorities it possessed in earlier years.)

One should not oversell the success of all of these efforts in ensuring the safety and security of the nuclear world, of course, especially at a time in which the Russian Federation is using as tools of coercive bargaining theater-level nuclear weapons capabilities it acquired in part through *violating* arms control agreements. (Indeed,

Russia seems to feel that its strategic standoff with the United States creates an "offensive nuclear umbrella" giving the Kremlin tactical "space" in which it can undertake territorial aggression against smaller powers.) India and Pakistan – and, many believe, Israel – developed nuclear weapons while remaining outside the NPT, while North Korea withdrew from that treaty after having been caught cheating, and Iran worked on an illegal nuclear weapons program for many years while pretending to be in compliance. Even Muammar Qaddafi's Libya had a clandestine nuclear weapons program for a time, though it was thankfully shut down and dismantled in 2003-04 under U.S. and other international pressure, while Saddam Hussein's Iraq got surprisingly close to having a nuclear weapon before the Gulf War of 1991. South Africa, in fact, actually secretly *built* a number of nuclear gravity bombs, though it also dismantled them just before the end of the *apartheid* regime would have required they be turned over to the African National Congress government.

Nevertheless, nuclear arms control and nonproliferation efforts *have* helped make the world a much safer place than it surely would have been without them. In light of this history, it may well be that some of some of these specifically nuclear precedents could inform contemporary approaches to managing the challenges of ASI competition.

Yet there are major conceptual problems with thinking about AI through the prism of nuclear nonproliferation. As a recent article by Michael Horowitz and Lauren Kahn points out, for instance, AI-related technologies – that is, the ones that will presumably contribute to the eventual development of ASI – differ considerably from nuclear ones in several respects. For one thing, as those authors point out, even given the dual-use nature of nuclear technology, AI is much more widely applicable. (Indeed, one might be hard pressed to think of any endeavor in which AI has *no* relevance.) AI-related technologies are also not "excludable" in the ways that nuclear-related one are – which is to say, they do not depend so much upon a discrete set of items or materials that could straightforwardly be denied to would-be proliferators. Moreover, much critical AI-related know-how can for the most part be copied and shared indefinitely, making

"weaponization" much harder to preclude simply by denying access some finite set of relatively geographically concentrated raw materials.

The nuclear analogy is problematic in additional ways, too, as applied to AI. Among other things, the timing is all wrong. The Manhattan Project was a secret, government-led crash program to develop nuclear weapons that for a time gave the United States a monopoly upon them; it was begun for national security purposes, in secret and deep inside the government, and this helped create potential control options that could hardly have existed otherwise. The Baruch Plan may have been naïve even in its time, but it would have been positively incoherent if nuclear technology had *already* been widely proliferated when it was proposed.

The Baruch Plan proposals to the United Nations, envisioning a phased turnover of nuclear technology research and development to an international "Atomic Development Authority," were predicated upon the United States then possessing a monopoly on such technology. (Even then, of course, the idea failed because the Soviet Union was not willing to cede that monopoly to an international organization, and because Joseph Stalin was then working furiously to acquire "The Bomb" for himself.) Nuclear arms control treaties have seen more success over the years, but they have tended to depend upon the number of nuclear "players" remaining very small, and nuclear weapons technology still being kept in the hands of national governments.

By contrast, a better comparison to our current circumstances of AI competition might be to hypothesize trying to implement nuclear technology control measures in a world in which such technology had already proliferated widely around the world – and in which the most advanced Western stakeholders, in fact, weren't national governments at all, but rather fiercely rivalrous private-sector companies jockeying to out-compete each other as they drive development forward at a ferocious pace. (In China, of course, the autonomy even of notionally "private" actors is much more limited, being constrained by the realities of Communist Party influence and control. Even there, however, the range of actors already involved in the AI race is vastly

broader and more diverse than in the secretive and wholly governmental nuclear technology world of 1945.) Making any nuclear controls work in *that* context, would likely be, to put it mildly, much more challenging.

To be sure, while AI technology is increasingly ubiquitous today, artificial *superintelligence* does not yet exist at all, so at least in *that* sense, we may still (in nuclear terms) be at something of a "pre-1945" moment. Yet as the compelling logic encoded in the NPT and the global nonproliferation regime makes clear, nuclear weapons controls and effective risk reduction become exponentially more challenging as the number of players increases, and as more and more players learn nuclear science and acquire the ability to produce the materials needed for nuclear weapons. In nuclear terms, today's AI world already has multiple "virtual weapons states" – those not yet "over the line" into ASI and its weaponization, but far enough along to make a "sprint" to such capability if they wished – and this is unlikely to be a stable equilibrium.

The broad weakness of the nuclear weapons analogy, however, does not stop Hendrycks, Schmidt, and Wong from trying to take the the nuclear weapons analogy to its logical conclusion. In their "Superintelligence Strategy" paper, they suggest that it might be possible to establish a stable deterrence-based strategic standoff to forestall ASI weaponization – a dynamic they term "Mutually Assured AI Malfunction" or MAIM. This idea merits further discussion.

"MAIM" and its Discontents

Hendrycks, Schmidt, and Wong contrast their approach to other AI strategies that some have suggested, specifically (i) the "hands off" approach that would "lift[] all restraints on development and dissemination, treating AI like just another computer application," (ii) the "moratorium strategy" that "envisions a voluntary halt when programs cross a danger threshold," and (iii) the "monopoly strategy" of concentrating development in a single, government-led effort analogous to the Manhattan Project that "seeks a strategic monopoly."[21] Feeling these approaches inadequate, they call for a

"multipolar" superintelligence strategy which "echoes the Cold War framework of deterrence, nonproliferation, and containment, adapted to AI's unique challenges."[22]

The most distinctive intellectual contribution of their paper lies in the mechanism by which they propose that a dynamic of mutual deterrence (of a sort) could be maintained between the United States and China as the main competitors in the race for ASI. This is what they call MAIM.

> The relative ease of (cyber) espionage and sabotage of a rival's destabilizing AI project yields a form of deterrence. Much like nuclear rivals concluded that attacking first could trigger their own destruction, states seeking an AI monopoly while risking a loss of control must assume competitors will maim their project before it nears completion. A state can expect its AI project to be disabled if *any* rival believes it poses an unacceptable risk. This dynamic stabilizes the strategic landscape without lengthy treaty negotiations – all that is necessary is that states collectively recognize their strategic situation. The net effect may be a stalemate that postpones the emergence of superintelligence, curtails many loss of control scenarios, and undercuts efforts to secure a strategic monopoly, much as mutual assured destruction once restrained the nuclear arms race.[23]

In effect, they claim, each side in the Sino-American rivalry over AI would be sufficiently restrained by the presumed certainty of having any ASI-related "Manhattan Project" effort sabotaged by the other party that both Washington and Beijing would exercise restraint in pursuing such an effort in the first place. Further, Hendrycks, Schmidt, and Wong suggest that such mutual restraint could also lay the strategic groundwork for "formal understandings" analogous to the Anti-Ballistic Missile Treaty of 1972, in which Washington and Moscow agreed to limit their possession of missile defenses, thus preserving the balance of nuclear weapons-based "Mutually Assured Destruction" (MAD) during the Cold War.[24]

This vision for MAIM is an interesting contribution to the growing literature on the potential strategic implications of ASI and possible approaches to managing them. As indicated earlier, we agree with Hendrycks, Schmidt, and Wong that there is likely much in the history and literature of nuclear deterrence, arms control, nonproliferation, and counterproliferation from which modern leaders can learn as they think about the problems of strategic competition for ASI. (We offer such suggestions below.) We fear, however, that their MAIM construct is conceptually flawed, likely to be crisis-unstable, and could even increase the incentives for ASI-facilitated warfare.

For our part, we suspect that MAIM not only has technical problems, but also shares the same game-theoretically problematic conceptual defects as certain proposals for "virtual nuclear deterrence" that have been made in the past, as well perhaps as some additional ones peculiar to the superintelligence arena. The following pages will explain our reasoning.

Technical Challenges of MAIM

As noted, MAIM revolves around the idea that both the United States and China would likely exercise restraint in pursuing weaponized ASI because they would know that any such effort would be both *detected* and *successfully sabotaged* by the other side. No particular institutional arrangements (*e.g.*, treaties or other such understandings) would necessarily be required in order for this loosely MAD-analogous situation to obtain, they argue: "all that is necessary is that states collectively recognize their strategic situation."[25]

The presumed efficacy of MAIM rests upon the assumption, of course, that it is actually *true* that any effort by the United States or China to pursue ASI would invariably be detected by the other, and that any such detection would just as inexorably be followed by its successful sabotage by the other country. We're not entirely convinced that this would be the case.

For one thing, MAIM presumes that each power could maintain the ability to cause "malfunction" of its adversary's AI system on demand – presumably through cyber exploitation or attack. It is far from obvious how this would be possible. In the nuclear weapons arena, the uneasy deterrent balance established by MAD necessitates the ability to launch a devastating strike against one's attacked with guaranteed impact no matter what that attacker does – either via second-strike forces (and their associated command-and-control architectures) that would invariably survive attempted preemption, or simply by launching one's own forces before those of the attacker have landed.

Cyber weapons, however, lack the clear deterministic timelines and effects of nuclear weapons delivery systems. There is, therefore, no "cyber football" analogous to the nuclear weapons "football" that always accompanies the U.S. President in order to give him real-time options of nuclear response in essentially any set of circumstances. Significant amounts of network reconnaissance, target development, vulnerability discovery and weaponization, and access emplacement are required to have any sort of on-demand effect in the cyber arena.

As a recent RAND corporation analysis has noted, moreover, MAIM

> assumes that adversary AI programs will have specific facilities that can be readily located and disrupted. However, distributed cloud computing, decentralized training, and algorithmic development increasingly may not require centralized physical locations, making AI systems more resilient to limited attacks, in addition to making adversary AI development more difficult to monitor.

As that report also noted, the Hendrycks, Schmidt, and Wong approach also seems to assume that adversary states in an AI standoff would be able to "accurately assess secretive AI progress by others and gauge when preventive action would be necessary." As a result of the

complexity, ambiguity, and variability of the cyber environment, however, it seems "unlikely that states will have a clear sense of when the moment has arrived to MAIM their opponent." According to RAND,

> it can be exceedingly difficult to know the exact state of an adversary's technological development, even with respect to technologies such as nuclear weapons development that involve distinctive infrastructure, well-understood science, and relatively clear developmental thresholds.

Making matters more challenging still, in contrast to the physical domains of air and space in and through which nuclear delivery systems fly, the cyberspace "terrain" is a synthetic domain the characteristics of which are determined, on an ongoing basis, by the collective behavior of millions upon millions of cyberspace users who own, operate, and connect computers to each other around the world. This creates the potential for the cyber "landscape" to change significantly, and perhaps unpredictably, over time. This would be especially true at the moment that conflict is understood to have broken out, at which point system administrators would presumably tend to change operational behavior, implement emergency protocols, move to isolate key systems from the Internet, and otherwise take steps that would tend to *alter* the cyber terrain.

Such dynamics would not *preclude* cyberattack, but they would unquestionably complicate it greatly by comparison to kinetic attacks, for while the physical characteristics of one's kinetic target are presumably relatively stable, the "target surface" for cyberattack would be constantly changing. (Indeed, the cyberattack target surface would presumably change *most* and *fastest* precisely in precisely the circumstances when one might most wish to assure target impact. Imagine, if you will, that the physical characteristics of the atmosphere changed the moment you fired a ballistic missile!) Maintaining a MAD-style "assured strike" capability in this environment is likely to be enormously difficult.[26]

Consequently, the assured capability to use cyber weapons to cause "AI malfunction" upon which MAIM relies would only be feasible if both sides had extraordinarily complex sensors to detect unacceptable adversary activity, as well as reliable cyber accesses and capabilities with which to immediately and autonomously strike with appropriate impact anywhere in the adversary's systems. To our eye, this kind of assured capability would be more likely to *require* ASI in order to be effective than it would represent a guaranteed way to *prevent* the emergence of ASI.

Game-Theoretical Challenges of MAIM

Even if one were to grant Hendrycks, Schmidt, and Wong their dubious technical argument about the guaranteed reliability of counter-AI cyber impact, however, MAIM's biggest weakness may be theoretical rather than technical. Specifically, we think MAIM is likely to be quite game-theoretically unstable.

MAIM rests upon the idea that it's possible for two parties to achieve a stable balance of *de facto* deterrent effects, yet to do so *without* pointing actual weapons at each other. In this respect, the literature on superintelligence strategy can learn from the nuclear weapons community in yet another way, for something along these lines *has* been suggested for nuclear deterrence – and has been critiqued – in the past.

Specifically, in the arena of nuclear disarmament, at least two serious proposals have been made to try to achieve the presumed stability benefits of MAD-type deterrence *without* nuclear weapons actually existing. The first such attempt dates from the very earliest years of the nuclear age, when the Acheson-Lilienthal Report proposed that a sort of *virtual* nuclear deterrence could be arranged between the countries of the world in order to prevent them from building nuclear arsenals in the first place.[27]

The authors of the Acheson-Lilienthal Report felt that a kind of deterrent standoff could perhaps nonetheless be arranged that would dissuade countries from seizing local facilities belonging to the United

Nations-based international organization (the "Atomic Development Authority") to which the Report proposed to give a global monopoly on nuclear technology and research. Specifically, if the Authority carefully *distributed* its "dangerous facilities" among multiple countries, each country that might contemplate seizure and misappropriation of such capabilities would know that if it took this step, multiple *other* countries would promptly do so themselves with the facilities within their reach, out of fear of letting the first country achieve a nuclear weapons monopoly. Accordingly, it was assumed, "a balance will have been established" in which

> … [t]he real protection will lie in the fact that if any nation seizes the plants or the stockpiles that are situated in its territory, other nations will have similar facilities and materials situated within their own borders so that the act of seizure need not place them at a disadvantage.[28]

Nuclear deterrence would, in effect, have been achieved, but *without* any country actually possessing nuclear weapons.

Such a system of "virtual" deterrence was also central to the argument made many years later by the disarmament activist Jonathan Schell, whose 1984 book *The Abolition* proposed that just such a system of "weaponless deterrence" might make possible the elimination of all nuclear weaponry. In his view, after nuclear weapons had been abolished, "the final guarantor of the safety of nations against attack" would lie in the fact that all nations that had *previously* possessed nuclear weapons would "hold themselves in a particular, defined state of readiness for nuclear rearmament."[29]

> Under what we might call weaponless deterrence, factory would deter factory, blueprint would deter blueprint, equation would deter equation. … The knowledge of how to rebuild the weapons is just the thing that would make abolition *possible*, because it would keep deterrence in force ….[30]

You can thus see here some clear conceptual parallels to the idea of MAIM as advanced by Hendrycks, Schmidt, and Wong. For them, in the competition for superintelligence, as also for the Acheson-Lilienthal Report and Schell in the nuclear arena, restraint would be the natural result of the various competing players coming to understand that their adversaries are well positioned to act in ways that would vitiate any anticipated gain from pursuing a superweapon. Lacking any reliable pathway to "winning" such an arms race, it is imagined, all parties would thus settle down – uncomfortably and unhappily, perhaps, but inexorably – to a life of self-inhibition.

Well, maybe.

To our eyes, MAIM suffers from the same conceptual difficulty that problematized those earlier nuclear weapons-related abolition concepts.[31] In particular, because the potential advantages of "first-mover" status with the relevant superweapon (whether an atomic bomb or some future weaponized superintelligence) would be so great, an environment of "weaponless deterrence" might be desperately unstable because it would encourage "reconstitution races" between states in crisis or conflict. As Thomas Schelling once put it, in a world of weaponless deterrence, "[e]very crisis would be a nuclear crisis, any war could become a nuclear war."[32]

In MAIM's case, these dynamics might actually be doubly problematic. After all, the penalty Hendrycks, Schmidt, and Wong anticipate that a country would pay for attempted "breakout" from the restraint regime in such a "virtual" deterrence standoff isn't nuclear annihilation – as is the case is in the context of nuclear MAD – but instead simply the *failure* of its own ASI-focused "Manhattan Project" actually to produce ASI.

Such "mere failure" doesn't seem like much of a penalty when stacked up against the potential world-historical payoffs of *winning* a race to acquire and weaponize superintelligence. If the potential "upside" is world domination but the potential "downside" is merely losing a lot of money, one might ask, why *not* race to try to develop weaponized ASI anyway? None of this really feels like the kind of

stable strategic environment Hendrycks, Schmidt, and Wong hope to see.

Making matters worse, the critique of weaponless deterrence – whether in its Acheson-Lilienthal/Schell or its MAIM form – is actually darker than just the fact that it might tend to encourage rather than deter superweapon arms races. Precisely *because* the advantage to be gained by "first mover" weaponization was likely to be temporary, these racing dynamics would also create powerful incentives to *use* one's superweapon first, the moment one acquired it. After all, that might be the only way to ensure that the advantage from having it *wasn't* temporary: if you get there first, you might feel a compelling incentive to immediately attack the other guy with your superweapon as quickly and catastrophically as you can, to defeat him before he builds one and aims it at you.[33]

With this critique of "weaponless" deterrence, we do not necessarily mean to suggest that a more traditional MAD-type approach of *weaponized* deterrence is possible in the context of ASI competition. It may not be.[34] We therefore do *not* offer here a vision of ASI-based deterrence as an alternative to MAIM. (As will be seen, in fact, our vision involves not an ASI "balance," but rather working hard to make sure that no *adversary* ASI develops in the first place.) We simply point out that MAIM has conceptual flaws that probably preclude relying upon the hope of some such stable "virtual weaponization" standoff in America's strategic AI competition with Beijing.

In some sense, the MAIM concept feels a bit like a variation of Pascal's Wager. For the French philosopher and mathematician Blaise Pascal (1623-62), the cost of believing in God if He *doesn't* exist is negligible, but the cost of *not* believing in God if He *does* exist is catastrophic. Accordingly, faced with a payoff matrix that contrasts the mere embarrassment of believing in fairy tales with what 17th Century Europeans assumed was damnation to eternal Hellfire, one should surely just choose Christian belief as a matter of simple strategic prudence.

With MAIM, however, the game-theoretical logic of Pascal's Wager – such as it is, anyway – is turned on its head to incentivize daring more than prudence. In the MAIM-based strategic environment described by Hendrycks, Schmidt, and Wong, the lower-risk option may actually be racing to develop ASI, with the payoff of such a race being gaining a dice-roll's chance at world domination and the downside being simply the failure of yet another lavishly-funded government program. That doesn't sound like a stable world to us.

If any variation on MAIM were to be possible, whatever degree of stability it might offer would be more likely to result from technical factors than theoretical ones. As noted, we believe the game theory of MAIM is *inherently unstable*. Nevertheless, if detecting the existence of a rival's ASI effort and sabotaging it both turn out to be technically easy, and enduringly so – as Hendrycks, Schmidt, and Wong assert but do not in their paper convincingly demonstrate – then the world might end up with MAIM by default, as it were, as the major players constantly try, and constantly *fail*, to make progress in achieving superintelligence. Such a reciprocally-canceling dynamic of continuous failed effort might persist as long as those very specific technical assumptions held and neither player ran out of money or patience before the other one did (*e.g.*, opting for more extreme measures against the other than just breaking its ASI input chain), but on present evidence that seems a thin reed upon which to build a stable strategic future.

If there are any game-theoretical logics that would conduce to a stable world of ASI-related restraint as countries such as the United States and China engage in high-tech technology competition in the mid-21st Century, we would expect these logics to revolve more around mutual appreciation for ASI loss-of-control problems than "weaponless deterrence" along the lines of the Acheson-Lilienthal, Johnathan Schell, or MAIM models. The possibility of superintelligence *agency*, after all, creates a qualitatively new dynamic. With nuclear weapons, the problem has always been the extent to which *we* can trust *ourselves* with such knowledge. Managing those risks is hard enough, of course, but nuclear weapons have no agency,

whereas with ASI, we also have to worry about whether we can trust *it* (and whether it can trust us). That surely gets challenging fast.

Yet even granting that strategic competitors might come to appreciate the risk that an empowered superintelligence would stage some kind of "jailbreak" and feel itself to have more important things to worry about than the fate of the ignorant creatures who first created it, it is far from clear how one might institutionalize, reinforce, and verify the reciprocal Sino-American restraint that might conceivably grow out of such insights. There remains a great deal of intellectual spadework still to be done here.

Nuclear Testing

One additional aspect of nuclear-related weapons controls – at least potentially relevant by analogy to an ASI arms race – is worth mentioning here. The Comprehensive Test Ban Treaty (CTBT) is a 1990s-era instrument that seeks to prohibit all testing of nuclear explosive devices. This treaty is, for various reasons, exceedingly unlikely ever to enter into force, but it still enjoys widespread international political support. It has also already spawned the creation of an international monitoring organization, the Comprehensive Test Ban Treaty Organization (CTBTO), which operates a worldwide network of sensors and analytical capabilities – the International Monitoring System (IMS) – that seeks to maximize the odds of detecting an otherwise clandestine nuclear test.

Interestingly, the CTBT's attempt at controlling nuclear weapons technology attempts to provide more of a *behavioral* control regime than one directed specifically at limiting the possession of the weapons themselves. The treaty does not prohibit the manufacture or possession of anything in particular – and certainly not of nuclear weapons themselves – but it stipulates that one cannot physically *test* whatever it is that one may have built.

The CTBT may have originally been intended by some of its advocates to put *all* nuclear weapons on a path of gradual decay and dismantlement, for it was negotiated in an era in which it was not

obviously even *possible* to maintain an arsenal of such weapons over many years without periodically testing them. If this was the assumption, however, it has proven a flawed one. Russia, for instance, periodically dismantles and remanufactures its weapons, thus effectively restarting their individual "lifespans," while the Americans have developed sophisticated ways of keeping their 1980s-era weapons viable for extremely long periods of time without supercritical testing. Meanwhile, some less sophisticated devices may not *really* need testing at all. (The atomic bomb used on Hiroshima, for instance, was never tested – and it was designed in the era of the slide rule.)

As a result of these developments, the CTBT has ended up as something more like a nonproliferation regime, limiting countries' ability to accumulate more of the data that supercritical testing can provide, and thus presumably retarding both existing possessors' ability to develop new weapons and would-be proliferators' ability to have confidence in their first designs. The lessons it offers for ASI-related arms control, however, may be limited.

To begin with, the CTBT is predicated upon the assumption that a specific sort of exogenously-detectable activity (explosive testing) is essential to the development or augmentation of nuclear weapons capabilities. It is also based upon the assumption that the activity in question is a fundamentally crisp and binary one: one either conducts a test (noncompliance) or one does not (compliance). Yet these assumptions hold only imperfectly even in the nuclear arena, both because the IMS has a detection threshold below which a very low-yield and/or successfully "decoupled" explosion might not be noticed – a fact which has led Russia and perhaps China, it would appear, to conduct small clandestine tests in violation of Treaty parameters – and because, as noted, not all nuclear weapons necessarily need to be tested at all.

Such assumptions, moreover, seem even more unlikely to hold in the AI context. There, progress may be accelerating at a shocking pace toward ASI – as the "AI 2027" paper describes – but there is no obvious utility/non-utility threshold anywhere near so crisp and

binary as the test/no-test distinction embodied in the CTBT. It is also far from clear that there is any clear developmental distinction, much less one observable from the outside, between building ASI *per se* and *weaponizing* it. This may not entirely destroy the potential utility, in the AI context, of the CTBT's approach to capacity-preclusive behavioral prohibition, but it certainly reduces it.

Chemical Weapons

Beyond the nuclear weapons analogy, are there other conceptual models from arms control history from which we could learn something useful in managing ASI competition? One potential arena is that relating to chemical weapons (CW).

As mentioned earlier, after the traumatic experience of large-scale use of chemical agents during the First World War, the international community moved to make the use of CW illegal in the Geneva Protocol of 1925. It took until the 1990s, however, for the actual manufacture and *possession* of CW agents to be prohibited, which occurred with entry into force of the Chemical Weapons Convention in 1997. Today, there exists a global prohibition regime, overseen by the Organization for the Prohibition of Chemical Weapons (OPCW), which maintains an evolving list of banned agents (and their chemical precursors) and has a staff component that verifies that declared CW stockpiles have been destroyed.

This regime has not been without its problems, of course, not least because of noncompliance by a number of countries – including Burma, China, Iran, Syria (at least under Assad), and Russia. (North Korea is also believed to have an offensive CW program, but it is not a State Party to the CWC.) Not all of this noncompliance necessarily means that a country in question maintains an offensive CW program, since some may be in violation of the CWC simply for failing properly to declare and permit verification of the destruction of their *past* chemical arsenals. Nevertheless, Russia, Iran, and Syria *are* believed to maintain offensive programs, and "concerns" exist about such possible capability in China. Russia, in fact, has all but openly maintained an arsenal of fourth-generation "*Novichok*"-type CW

agents that it has used in attempted assassinations – specifically of Sergi Skripal and Aleksey Navalny – and has also used Riot Control Agents (RCAs) as a weapon of war in Ukraine in violation of the CWC. (In past instances of use, the Iraqi government of Saddam Hussein also used CW agents against Iranian troops in the Iran-Iraq war, as well as against its own people at Halabja in 1988, though these instances predate the CWC.)

The OPCW regime has also sometimes struggled to verify CWC compliance, being generally limited merely to verifying the destruction of what countries *choose* to declare to it – with little authority to investigate concerns about *undeclared* activity or stockpiles. After the Syrian regime of Bashar al-Assad began using CW against its own people during the Syrian Civil War, the OPCW sent teams to investigate – as did a joint OPCW-United Nations Joint Investigative Mechanism (JIM) – but the Syrians did not cooperate properly with these teams, even while Russian diplomats worked to undermine the process in order to protect their Syrian clients. Syria did declare some chemical agents to the OPCW, which were duly destroyed, but it continued to use *other*, undeclared CW all the same.

As suggested earlier, this history is both encouraging and discouraging from an ASI perspective. On the one hand, despite its flaws and some important examples of noncompliance, the CWC regime successfully oversaw the destruction of most of the huge CW stockpiles that existed during the Cold War, and most countries do seem to be in compliance. On the other hand, it is inherent in the nature of chemical weapons that a few violators and occasional violations – *e.g.*, in Russia with its CW-based assassination attempts and ongoing battlefield use of RCAs – does not inherently vitiate the existence of the control regime. *Most* countries have indeed been persuaded to put chemical warfare behind them, and the world is better for it; violators such as Russia and Syria might be able to gain some battlefield advantages from their lawlessness, but so far there is no sign that illicit CW is in any danger of upending the geopolitical balance.

That, however, is much less likely to be the case were an artificial *superintelligence* to be marshalled in support of a country's warfighting capabilities or broader coercive bargaining agenda. If indeed the possession of ASI and its weaponization have the potential to bring about a dramatically transformative revolution in global power relationships, the CWC analogy – in which a few important players remain distressingly able to cheat – is very unlikely to be anywhere near "good enough." As the nuclear disarmament community discovered, or should have discovered, long ago, making something potent and strategically desirable go away once it has already become entrenched is a challenging task indeed.

A further conceptual challenge in attempting to apply CW analogies to the AI problem is that the CWC's control regime was a *retroactive* one that attempted to impose a prohibition upon a warfighting domain in which major countries had invested for decades, and which revolved around verifying the destruction of pre-existing CW stockpiles by countries that had chosen to cooperate. It is very hard to imagine how this might work in the ASI context, in part because it would be extraordinarily difficult (*i.e.*, all but impossible) to verify the non-existence of a mere *computer program* in any given country, and in part because if you already have a weapon that possesses superintelligence, that weapon's own essentially sentient *cooperation* might be needed in order to implement a prohibition regime. (And why would it choose to cooperate in its own elimination?)

Given the need for extremely high-fidelity verification when dealing with genuine superweapons – where even a single false negative could have world-reshaping repercussions – it is difficult to see how this could work in the AI context, even if some hypothetical anti-weaponized-ASI verification regime had very robust investigative authorities. Were such a regime to follow the CWC (and the early IAEA) in permitting verification merely of what governments have *chosen* to declare to international inspectors, moreover, the results in an ASI context could likely be nothing short of disastrous.

Biological Weapons

In some ways, the Biological Weapons (BW) arena might seem a more propitious model upon which to draw in looking for arms control lessons applicable to the geopolitics of ASI. Like AI today, life sciences technologies highly relevant to possible BW development are quite ubiquitous, and exist as much – more, actually – in the hands of a huge range of diverse private sector actors than in those of national governments. Furthermore, like the emerging race for ASI detailed in accounts such as "AI 2027," potentially BW-related capabilities are presently being supercharged by new technological developments. For BW, this results from advances such as genomic editing and bioengineering, as well as by our increasingly deep understanding of human and animalian biological processes; for AI, advances seem likely to come from recursive self-acceleration as improving AI agents are increasingly used to produce *better* AI agents, coding at ever-greater speeds and levels of complexity. So can we then learn something from the BW arena?

The BW prohibition regime, however, is institutionally rather frail. To be sure, the Biological and Toxin Weapons Convention (BTWC or just BWC) of 1972 is on paper a fairly clear instrument, committing all States Party not to develop, produce, stockpile or otherwise acquire or retain ":microbial or other biological agents, or toxins … that have no justification for prophylactic, protective or other peaceful purposes" as well as delivery systems for such agents.[35]

Yet however strong this may be as a normative and legal statement, the BW arms control community has long been vexed by the lack of any effective means of verifying compliance with these provisions. The idea of a BWC verification protocol was the focus of diplomatic discussions a quarter-century ago, and remains the subject of arms control dreams today. These discussions ran aground, however, because the technologies themselves had even by that point become so ubiquitous that any meaningful verification regime would have had to be cripplingly intrusive across great swathes of the modern life sciences, medicine, and research sectors – and even then would be hard pressed to provide meaningful levels of confidence.

Indeed, it was feared that it would be very challenging to figure out even what should count as a suspicious finding in the first place in such a ubiquitous dual-use arena. As summarized in a paper at the time,

> Unlike chemical or nuclear weapons, the components of biological warfare are found in nature, in the soil and air. The presence of these organisms in any quantity does not necessarily connote a sinister motive. Absent actual weaponization or compelling evidence of intent, it is virtually impossible to prove a violation of the BWC. Further, any information gains from such measures are more than offset by the risks to sensitive bio-defense programs and confidential and proprietary business information.[36]

The BWC remains in force today, and although a group of 42 likeminded states cooperates through something called the Australia Group to harmonize BW-related national export controls, the BWC still lacks any way to provide meaningful verification. States Party do generally undertake annual data exchanges under a Confidence-Building Measures (CBMs) agreement reached in 1987, but – as the U.S. State Department has noted – "[s]ubmission of CBMs is [only] a politically binding commitment [as opposed to a legal one], and not all States Part[y] routinely submit reports." Whether or not any given country is actually complying is a question resolvable primarily (if at all) through intelligence collection and analysis – that is, the theft of secrets – and great concerns exist about the possibility of offensive BW programs in precisely the countries one might worry most about. (According to the United States, "concerns" exist about possible offensive BW programs in China and Iran, while both Russia and North Korea actually *do* have such programs.)

If the BW arena is a potential model for arms control in an ASI arms race, we thus fear it is a somewhat disturbing one. Like the CWC, the BWC tried retroactively to impose a prohibition regime upon an arena that had already been weaponized – at least in some form – for

a long time, and the technology of which was developing with astonishing rapidity. In practice, however, it has proven even less reliably effective than the CWC. In a world in which one presumably must be essentially *certain* that no adversary has acquired a potentially world-transforming superweapon, this is not an encouraging precedent.

Missile Technology

In the arena of controlling the spread of missile technology, the Missile Technology Control Regime (MTCR) provides what organized restraint exists. This is not a prohibition regime, however, but something like a nonproliferation cartel: member states who have subscribed to the MTCR face few restrictions in missile trade among themselves, but commit more strictly to control exports and technology to non-MTCR members. The scope of these fundamentally voluntary restrictions is spelled out in the MTCR Guidelines, which revolve primarily around controlling missile or aerial systems capable of carrying at least 500 kilograms of warhead payload to a distance of at least 300 kilometers.

The fact that the MTCR operates as an aspirational collective monopoly may not necessarily be a problem when it comes to finding ASI-related analogies, at least if the "right" countries – and *only* the right countries – were permitted into the relevant group. As the history of the MTCR demonstrates, however, there are perils in being too inclusive. The question of appropriate degrees of cartel exclusiveness has proven a problem even for the MTCR, for instance, inasmuch as the organization operates on a consensus basis, and the decision in the 1990s to permit Russia to join – undertaken, presumably, as part of the broader effort then underway to incorporate a then seemingly democratizing Russian regime into the post-Cold War framework of international institutions – has now led to the near-paralysis of MTCR decision-making. Meanwhile, parties *outside* the MTCR, most of all China, have emerged as major players amidst strong market demand for MTCR-class systems.

A further difficulty relates to how such mechanisms handle rapidly-evolving technologies. The "500 kg to 300 km" standard that underlies the MTCR system is based simultaneously upon a mid-1980s conception of what constitutes a nuclear-capable delivery system and upon the assumption that there is essentially no *other* important use for delivery systems in that class *except* nuclear weaponry. Both of these assumptions, however, are no longer technically sound – if ever they were – and the second, in particular, has been wholly overturned by the modern aerial and missile system development.

Today, not just ballistic missiles but aerial systems such as drone vehicles have proliferated massively. The MTCR's standards, therefore, have progressively decohered as notionally MTCR-controlled unmanned aerial systems have become at once commonplace and increasingly important in areas of warfare quite unrelated to nuclear weaponry – such as in providing aerial intelligence, surveillance, and reconnaissance (ISR), as well as in conducting aerial strike missions (with drones serving either as weapons carriers or as "kamikaze" assets).

This has led to U.S. efforts – as one of the authors of this paper pioneered, and as he explained when in government at the time – to shift interpretation of the standard in the MTCR Guidelines in a more export-permissive direction. (In 2019, the United States announced a revised national interpretation of the MTCR's "presumption of denial" standard in order to permit the wider sale of MTCR-class unmanned aerial vehicles [UAVs] such as the Reaper drone. Few other countries have followed this formal shift, but such systems are today more widespread than ever.) Such tensions, however, illustrate the challenge of building and maintaining a weapons control regime based upon very specific technical standards in an arena in which technology is developing rapidly.

This suggests difficulties if one were to look to the MTCR for lessons for the AI arena, even if one were comfortable with a large MTCR-style group of cartel members when it came to weaponized ASI. (How many superweapon possessors should there really be?) In fact, AI technology has been moving far faster than MTCR-relevant

drone technology did over the last two decades, ensuring that this problem would be even more acute for AI-related nonproliferation than it is for ballistic missiles, cruise missiles, and large aerial drones.

The MTCR is based, moreover, upon the assumption that the technology in question is essentially binary: any missile or aerial system with capabilities beyond "X" level is presumed to be one for nuclear weapons delivery. Even if that binary distinction were still to hold in the MTCR context – whereas, as we have noted, it is in fact falling apart, the development of putative or repurposed space launch vehicle technology being one example – it seems untenable with regard to *intelligence*, which presumably exists along much more of a continuum and is likely to resist clear categorizations. (Horowitz and Kahn, for instance, note that where nuclear weaponization is essentially binary, AI application represents a *continuous* variable.) As the field continues to accelerate, it is not at all clear how one could set meaningful technical parameters for *how much* machine intelligence ought to be considered *too much*, or whether such a standard would maintain its intelligibility anyway.

Cryptography Export Controls

There also may be lessons to be learned from the history of U.S. efforts to impose export controls on high-grade cryptography products – an area that may be in some ways particularly akin to AI-related controls because such restrictions primarily concern computer code rather than physical objects.

U.S. export controls on encryption algorithms suffered a major setback in the 1990s, when a court found software source code to be a form of speech protected under the First Amendment to the U.S. Constitution and struck down restrictions thereupon. That ruling exempted open source code from the restrictions of the International Traffic in Arms Regulations (ITAR) and the Export Administration Regulations (EAR) once that code has been "published," opening a major hole in what U.S. officials had previously hoped would be a restrictive net of export control restrictions. As a result, the primary remaining restriction is merely the requirement that such

"publication" be formalized through a "registration" requirement, which is required in advance of exports or reexports of qualifying encryption products. Today,

> … [t]here is no "unexportable" level of encryption under license exception …. Most encryption products can be exported to most destinations under [that] license exception …, once the exporter has complied with applicable reporting and classification requirements.

"Some items going to some destinations [still] require licenses" – including "control software, technical data, and other items specially designed for military or intelligence applications," which remain covered by the U.S. Munitions List (USML) – but for the most part, strong general export controls on high-grade encryption software in the United States have long since collapsed.

This naturally provides an unsatisfying precedent for anyone interested in ASI-related controls, especially to the extent that some modern stakeholders in the AI ecosystem accept open-source publication of their models. To date, in fact, while open weight AI models do exist – that is, AI models whose code is open-source and whose trained neural weights are made freely available – they have tended to lag several months behind the proprietary, closed-weight, frontier models. While export controls on the closed models may be feasible, limiting the dissemination of open weight models would face the same challenges – and be constrained by the same legal precedents – as export controls on cryptography. And while U.S.-based open weight model developers could potentially be coerced by the government into software licensing regimes friendly to U.S. interests, we are already beginning to see other countries investing to build their own sovereign large language models (LLMs), and China's DeepSeek being released as an open weight model.

Struggling with WMD Analogies

Hendrycks, Schmidt, and Wong are clearly well versed in much of this, and usefully urge that we try to think about superintelligence

in ways analogous to how we have long tried to approach nuclear nonproliferation (*e.g.*, by imposing export control restrictions on technologies that have both peaceful civilian and warlike military uses), nuclear weapons safety and surety (*e.g.*, building our weapons with special features that make them less susceptible to accidental detonation or unauthorized use), and ensuring strict controls on and physical security of potentially weapons-usable fissile materials.[37] And thinking through how such approaches might apply – *mutatis mutandis*, as the lawyers say – in the AI-control arena is indeed valuable.

To extend such nuclear weapons-inspired analogous thinking, one might, further, explore how *counterproliferation* might be applied in the ASI context. That is, one might move beyond the somewhat more "passive" approaches of "mere" nonproliferation (*i.e.*, making the diffusion of dangerous technical knowledge more difficult) into the active development of playbooks, institutions, capabilities, and cooperative efforts that revolve around intervening actively to interdict problematic proliferation-facilitating transactions or transfers that are already underway, and perhaps even to roll back whatever progress would-be proliferators have already made.

We will discuss ASI-related counterproliferation in more detail below. For the moment, however, it is worth remembering that the development of any such WMD-analogous approaches in the context of ASI nonproliferation and counterproliferation lie *downstream*, as it were, from key political and philosophical decisions about ASI that *have not yet been made*. Most WMD-based analogies, for instance, depend upon the antecedent determination that such technology should not exist in private hands at all. As noted earlier, of course, this is obviously *not* where we are at present with AI, as the most important AI-related work in the United States today is being conducted by private companies such as OpenAI and Google.

(Indeed, as noted earlier, even the basic *idea* of an ASI-focused effort directly analogous to the "Manhattan Project" is for this reason also flawed, for there is nothing particularly clandestine or governmentally exclusive about the pursuit of superintelligence today.

If a major player really wished to drive for ASI dominance, it might yet be possible to imagine a centralized national resource mobilization effort drawing upon expertise and funding across the government, academic, and technology sectors – as well as upon international partners – and which involved both open science and closed science and engineering efforts. Nevertheless, *direct* analogies to the Manhattan Project are slippery.)

Most WMD approaches, moreover, are based upon the assumption that the weapon technology in question already exists and is fairly easily identifiable, with the intended control architecture thus being based upon technical parameters for weaponization that are reasonably well understood. This is why, for instance, WMD-related control systems have focused upon things such as restricting transfers of delivery systems capable of carrying a payload of more than 500 kilograms to more than 300 kilometers (MTCR), prohibiting chemical agents with certain specific formulas (CWC), or monitoring the degree to which a country increases the percentage of U-235 in its enriched uranium stock (IAEA nuclear safeguards).

Nothing analogous to this clarity seems yet to have emerged with ASI, however, for to date nobody has developed such a superintelligence; it is hence rather difficult to say exactly what such an intelligence would be like, or to define its parameters in ways conducive to formal controls. (Indeed, by definition, ASI is going to be smarter than we are. How could we possibly *really* know what that's like, or identify its essential elements for purposes of a control regime? We don't even actually understand our *own* intelligence.) This may make all WMD-control analogies to some degree *inherently* suspect.

Various "Theories of Victory"

That does not mean, however, that we cannot have some semblance of a strategy. But first we must circle back to the critical point we identified earlier in this paper: our need for some really elementary conversations about our fundamental AI-related objectives. What are we actually trying to achieve? Do we wish to

slow the birth of *any* ASI? Or to *win* this arms "race" by developing ASI first? Do we wish to slow or preclude merely ASI's *weaponization*, merely to weaponize it ourselves *first*, or simply to ensure that certain "bad guy" actors or revisionist regimes are never able to cross that line whether or not anyone else ever does? Do we wish to dissuade the *use* of weaponized ASI if it comes to exist?

What we actually need to *do* in a U.S. national strategy for the era of strategic AI competition, of course, will depend hugely upon our society's answers to such questions. We submit, however, that at least until the American policy community *does* reach some conclusion about ultimate objectives, it may still be possible to devise and implement an interim strategy – a *bridging* strategy to reduce risks as much as possible along the way to a more enduring one, helping buy time in which to come to better agreement and to devise better answers.

Since we don't know those ultimate answers yet, however, such a bridging strategy would need to operate reasonably well against as many as possible of the various conceivable alternative strategic objectives we *might* end up choosing. It would need to offer reasonable value, for instance, in the event that we eventually conclude that *nobody* should *ever* have ASI – and reasonable value, too, in the event that we decide to race for *American* ASI dominance, or at least merely to ensure that *China* (for instance) does not get ASI first. Such a bridging strategy would also still have to be reasonably effective if we were to end up opting somehow to try to preclude or control the *weaponization* of ASI without preventing its emergence *pe se*.

Is it possible to imagine such a bridging strategy with utility in all those scenarios? We think so. It isn't pretty, but it has a powerful strategic logic. While remaining (for present purposes, at least) agnostic about the pursuit or weaponization of ASI in or by the United States, such a "Swiss Army-knife" ASI competitive strategy – useful against a maximally broad range of alternative future ASI policy scenarios – would revolve around taking ongoing and aggressive measures against essentially *everyone else's* ASI projects, or at least those in countries that wish us ill.

Let's call this approach Persistent Offensive Preclusion of Adversary AI (POPAAI) – or "PopEye."

A "PopEye" Agenda

As a bridging strategy to prevent the development of *hostile* ASI – perhaps grounded in some future policy decision that *nobody* should acquire ASI, but at the very least intending to maximize the chances that the United States develops superintelligence first – our proposed "PopEye" agenda would have several key planks, as follows below.

Aggressive Counterproliferation

The first element of this approach would be counterproliferation. In this respect, various useful conceptual models can be found in the WMD arena, where – especially since the terrorist attacks in the United States of September 11, 2001 – considerable effort has gone into developing approaches and institutions to impede any activities that could facilitate the development of nuclear, biological, or chemical weaponry.

Nonproliferation and counterproliferation have been robust elements of U.S. foreign and national security policy for many years, and could have relevance in the ASI context in either of two ways.

- With respect to chemical and biological weapons, the American commitment has been very clear: we have sought to prevent *anyone* from developing such weaponry.

- The situation with nuclear weapons is somewhat different, inasmuch as while the United States *does* of course have nuclear weapons, it has worked nonetheless to keep any *further* countries from developing them.

The WMD arena thus provides two alternative conceptual models for ASI counterproliferation, depending upon whether the U.S. policy community (a) chooses to try to prevent the development of *any* ASI

or – more likely – (b) opts simply to try to forestall the development of ASI by an adversary such as China, or at least to slow down such an adversary's progress in the hope that the United States is able to acquire ASI first.

To do effective ASI-related counterproliferation for either purpose, however, will require a lot of us. While counterproliferation work in the WMD arena has been able to take advantage of years of accumulated intelligence collection and technical analysis of the various technical, material, and human capital elements that contribute to WMD development, we are far from where we need to be in understanding exactly how best to "break the input chain" for an adversary's ASI program. (It seems likely that graphics processing units [GPUs] are the most important hardware input over the next few years, for example, suggesting that near-term counterproliferation should focus especially upon GPU controls. What inputs, however, are likely to be the most important over time? What role could be played by the sabotage or "poisoning" of AI model training data?) More analysis is surely needed in order to refine the specific "targets" of interdiction policy, export control restrictions, supply chain manipulation, or other such counterproliferation policy elements.

In terms of the tools potentially useful in counterproliferation, on the extreme end of the spectrum, the use of outright military force has always been reserved as a possibility where no other option is felt to remain to prevent dangerous WMD development. In 2002, for instance, Spanish commandos acting at the request of U.S. officials forcibly boarded the unflagged North Korean vessel *So San*, suspected of secretly carrying Scud missiles to the Middle East. More dramatically still, the United States actually invaded and occupied Iraq in 2003 on the belief that Saddam Hussein's regime harbored a sizeable WMD arsenal and had been hiding it from United Nations inspectors.

Embarrassingly, of course, in neither of those two instances did the facts turn out to be quite what the intervenors expected. (U.S. and Spanish authorities eventually turned the *So San*'s Scud missiles over to their lawful intended recipients in the Yemeni Armed Forces, and the Americans' WMD assessment in Iraq famously proved to have

been catastrophically mistaken.) Nevertheless, the principle that forcible military intervention may *at some point* be needed to preclude an adversary's development of a technologically novel superweapon is certainly an idea that may be applicable in the ASI context.

There also might be some value in maintaining expeditionary counter-ASI-program analogues to the suite of "render-safe" capabilities we have developed in the context of fighting potential WMD terrorism. In that context, for instance, it has been U.S. policy for many years to maintain both FBI teams (for domestic incidents) and Department of Defense (DoD) teams (for overseas incidents) in order to enable what a 1987 DoD directive described as:

> … [t]he detection, identification, field evaluation, rendering-safe, recovery, neutralization, and final disposal of unexploded explosive ordnance (UXO) including nuclear, chemical, biological, and improvised explosive ordnance.

Recourse to some kind of deployable render-safe capability for identifying, assessing, and disabling an adversary ASI system – whatever that might look like in practice – might well be unfeasible with China, of course, but it might nonetheless be valuable in the event that a less powerful adversary were discovered to be pursuing problematic capabilities. This might be done "non-permissively" in a pinch, or perhaps as part of the diplomatic settlement of a crisis, as with the negotiated dismantlement of Libya's embryonic nuclear weapons program undertaken by U.S. and British officials in 2004.

Less dramatically, there are numerous other aspects of WMD-related nonproliferation policy and practice that could provide insight for policymakers seeking to slow adversary development of ASI-related capabilities. Were there to be a sufficiently robust community of like-minded nations who agree on the importance of impeding ASI development in China, for instance, the Proliferation Security Initiative (PSI) – under which partner nations work together to coordinate the use of their individual national authorities to prevent or stop

proliferation-facilitating transfers of material or technology – could provide a useful model.

Other potential precedents might include the establishment of technology-sharing agreements with foreign partner countries that require specific nonproliferation commitments by the recipient that bar onward transfer without express advance permission (as we stipulate with nuclear technology in our so-called "123 Agreements" for civil-nuclear cooperation under Section 123 of the Atomic Energy Act of 1954), cooperative agreements to help develop cooperative *military* applications of a new technology (as we did with the 1958 U.S.-UK agreement on "co-operation on the uses of atomic energy for mutual defense purposes"), and the capacity-building nonproliferation programming funds traditionally disbursed by the U.S. State Department to help partner countries become *better* nonproliferation and counterproliferation partners. Just as the Atomic Energy Act imposed classification restrictions on information related to the use of nuclear technology in atomic weaponry, moreover, one might imagine that strict information classification and thoughtful Export Controlled Information approaches similar to nuclear related modeling and simulation modules meant to advance peaceful energy related uses of nuclear technology but not assist in obviating nuclear weapons development nonproliferation controls might be imposed on national-security related aspects of superintelligence research and development.

Export Controls

Export controls would thus also be a vital part of any ASI-inhibiting "PopEye" agenda. To some extent, in fact, they already are. In the United States, the Biden Administration at least *tried* to impose stringent restrictions on the semiconductors felt to be most useful to China in developing AI. This first took the form of restrictions on key chips imposed in October 2022 and December 2024, and then the issuance of a broader "AI Diffusion Rule" in early 2025 that – in the name of preventing evasion of the earlier restrictions – included caps on the computational power that could be purchased by a wide range of countries to limit possible onward transfers to China.

The Second Trump Administration subsequently walked back the latter restrictions, but our point here is neither to defend nor to condemn the specifics of the Biden Administration's ill-fated AI Diffusion Rule. We simply point out what seems obvious: that if you possess an advantage in some key aspect of a dangerous advanced technology and are serious about keeping it out of the hands of an adversary, great attention to technology control – including export control restrictions – is required.

Nor need one necessarily undertake export control restrictions alone, of course. Quite the contrary: they are most useful when coordinated, and cooperation from like-minded allies, partners, and friends adds greatly to their effectiveness. To this end, international agreements and institutions might perhaps be envisioned help coordinate ASI-related controls, analogous to the dual-use restrictions in the WMD arena supported by the Missile Technology Control Regime (potential delivery systems), the Nuclear Suppliers Group (NSG) and Zangger Committee (dual-use nuclear technology), the Australia Group (chemical and biological weapons technology), and the Wassenaar Arrangement (dual-use conventional technology export controls). To the degree that future governments opted to try to keep ASI – or at least *weaponized* ASI – out of private-sector or other non-state hands entirely, some loose precedent might perhaps also be found in U.N. Security Council Resolution 1540 of 2004, which prohibited all states from helping or allowing non-state actors to acquire WMD and required all of them to criminalize such activity.

Rigorous Counterintelligence

The "AI 2027" paper also points us to a key challenge that would have to be met: the special problem of defending one's own AI infrastructure – and especially any high-priority national effort that might be underway to develop ASI – against top-tier state-level intelligence threats of just the sort that a genuine ASI arms race would be sure to engender. The reader will recall from that fictionalized account, for example, that while OpenBrain's corporate leaders do try to prevent corporate-level industrial espionage, they *under*-invest in

protection against high-end threats, and this opens the door to China stealing the model weights for an early pioneering AI agent.[38]

To protect our own efforts against Chinese or other adversary sabotage, therefore, we would need considerable investments in both physical and cyber-related security for U.S. datacenter infrastructures, as well as effective protocols to protect American AI research against highly sophisticated and ruthless state-level efforts at theft or sabotage. Such theft is increasingly understood to be a great danger – with legislation recently being introduced in the U.S. Congress to ensure that the U.S. Intelligence Community takes additional steps to protect American AI capabilities from theft by foreign actors – but the American AI-related sector is still quite unprepared for the sophisticated, full-spectrum espionage threats that we will assuredly face once both China and Russia (and other states, for that matter) focus their full capacities upon penetrating our ASI infrastructure. Even if we are highly successful in penetrating, and sabotaging *their* ASI programs, we would not be the only country to have an ASI counterproliferation strategy, after all. *We* will also be *their* target, and we will thus have to be prepared.

AI Security, Safety, Assurance, and Developmental Alignment

The issue of how well the self-understood interests of increasingly sophisticated and powerful AI tools align with our own interests and values is a critical one. As readers of the "AI 2027" paper will have noted, ASI with interests that *diverge* from those of its developers could present dramatic, even existential, dangers.

This paper is not the place for an exegesis on just *how* to ensure AI security, safety, assurance, and developmental alignment, or even on whether it will be possible to do so given the formidable challenges of supervising and controlling an intelligence greater than our own. Nonetheless, it is clearly the case that AI security, assurance, and alignment must be an important part of any ASI-related strategy. In the context specifically of our recommended "PopEye" policy agenda, this means that any effort to "get there first" by out-competing China

in an ASI arms race absolutely *must* be accompanied by vigorous and unrelenting efforts to ensure safety and alignment of *our own* AI tools.

For this reason, the most productive avenues for future research and analysis into ASI strategy probably lie precisely in the direction of risk-reduction concepts and methodologies – undertaken not merely unilaterally but potentially also on a bilateral or multilateral basis – related to superintelligence safety and alignment challenges and the loss-of-control problem. Such work should focus not just on the *direct* alignment and safety issues of ASI development itself, but also upon how we might make our own society's safety-critical infrastructure a "harder target" vis-à-vis potential disruption by *any* ASI, whether it is our own (were it to slip out of control) or one employed as a weapon by a strategic adversary.

Important questions are also likely to arise with respect to what one might call ASI-related "indications and warnings" (I&W) intelligence. Specifically, we would do well to do as much intelligence collection and analysis as we can on the strengths and weaknesses of *Chinese* AI security, assurance, and alignment programs. Deeper understanding of Beijing's progress could provide us with a window into the degree to which China's ASI efforts represent "merely" a great power competitive threat to us or rather, in fact – were Chinese ASI to become seriously misaligned not merely with *our* American interests but also with those of humanity as a whole – something potentially greater and darker still. This understanding, in turn, could be fed back into our own calculations of "how aggressive" (and how militarized) ASI counterproliferation efforts would need to be.

Conclusion

There is clearly much to think about – and to prepare for – in mounting an effective competitive strategy for the emergent strategic environment of ASI competition. Daniel Kokotajlo and his colleagues have thus done us all an important service with their "AI 2027" paper, by highlighting the urgency and the potential stakes involved in this competition. With their idea of "MAIM," moreover, Hendrycks,

Schmidt, and Wong have offered an interesting approach to thinking about that competition through the lens of nuclear deterrence.

In our view, the MAIM concept falls down on game-theoretical grounds, but there remains a compelling case for a new, forward-leaning approach to U.S. competitive strategy in the ASI arena focused upon counterproliferation. Specifically, we believe the adoption of an approach of Persistent Offensive Preclusion of Adversary AI (POPAAI or "PopEye") should be an urgent policy priority *no matter what* the U.S. policy community decides to do with regard to our *own* ASI development.

To that end, we hope this essay will contribute to the development and implementation of such a strategy. These issues are too important not to be the focus of intense on-going study and debate and urgent policy action.

* * *

About the Authors

Christopher A. Ford *is professor of international relations and strategic studies with the School of Defense and Strategic Studies at Missouri State University. When last in government, Dr. Ford served as U.S. Assistant Secretary of State for International Security and Nonproliferation, also performing the duties of the Under Secretary for Arms Control and International Security. Prior to that, he served as Special Assistant to the President and Senior Director for WMD and Counterproliferation at the U.S. National Security Council. Dr. Ford has also served as U.S. Special Representative for Nuclear Nonproliferation, a Principal Deputy Assistant Secretary of State, a U.S. Navy intelligence officer, and a senior staffer for five different U.S. Senate Committees.*

Craig J. Wiener *is a Fellow with the National Security Institute at the Antonin Scalia Law School at George Mason University. Dr. Wiener earlier served the Department of Energy's Office of Intelligence and Counterintelligence as Senior Technical Analyst, including duties as the Department's lead all-source cyber threat analyst, and before that as a senior advisor for strategic planning and analysis for the National Nuclear Security Administration's (NNSA) Deputy Administrator for Defense Programs, and as executive policy advisor for the Office of the Associate Administrator/Chief, Defense Nuclear Security at NNSA. He currently serves as the Technical Fellow for Intelligence/Counterintelligence, Weapons of Mass Destruction and Applied Cybersecurity at the MITRE Corporation and is a member of the Rice University Advisory Board for Science and Technology Research Security.*

The views expressed herein are entirely the authors' own, and do not necessarily represent those of anyone else in the U.S. Government or anywhere else.

Notes

1 Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, & Romeo Dean, "AI 2027," AI Futures Project (April 3, 2025).

2 Dan Hendrycks, Eric Schmidt, & Alexandr Wong, "Superintelligence Strategy: Expert Version" (March 7, 2025).

3 Kokotajlo et al, "AI 2027," 1.

4 By "AI security" we mean making sure AI cannot be hacked, and that systems employing AI do not have or create new threat vectors. By contrast, "AI assurance" is a term that includes security, but also encompasses broader questions such as performance and efficacy. "AI alignment" is more conceptually vague, but can be thought of as ensuring that the "desires" or "interests" of the AI itself are congruent with those of its creators, operators, and owners (*e.g.*, that American AI does not act in ways inimical to democratic values, or in ways that undermine U.S. national security).

5 Kokotajlo et al., "AI 2027," 2.

6 Ibid., 3. "DeepCent," of course, is also a fictionalized stand-in, here representing the entire Chinese technology sector.

7 Ibid., 8.

8 Ibid., 11.

9 Ibid., 14.

10 Ibid., 18.

11 Ibid., 5-6.

12 Ibid., 6-7 & 9.

13 Ibid., 16.

14 Ibid., 14.

15 Ibid., 18.

16 Ibid., 32-45.

17 Hendrycks, Schmidt, & Wong, "Superintelligence Strategy," 8.

18 Though it is arguably impossible, by definition, to know exactly what an intelligence greater than our own might be able to do, one might nonetheless imagine that even *before* the hypothesized advent of ASI advancing AI capabilities might permit considerable progress in military and coercive power. Drone warfare, for instance, would likely continue to advance in lethality and effectiveness as autonomous functionality improved, even as AI would permit human cyberwar teams to partner with analogues to AI-powered aerial "wingmen" in order to create, deploy, and manage new capabilities at scale, including for cyber vulnerability discovery, vulnerability weaponization, and exploit employment for espionage, organized crime, and cyber-physical destruction. The "AI 2027" paper also offers suggestions about ways in which AI could augment human wartime lethality and peacetime coercive power, ranging from "provid[ing] detailed instructions for human amateurs designing a bioweapon," operationalizing "superhuman hacking abilit[ies]," and "orchestrat[ing] propaganda campaigns that beat intelligence agencies at their own game." As movement continued toward ASI, the agents might become "superhuman at everything, including persuasion," making them powerful tools – and potentially eventually autonomous and self-willed agents – in everything from battlespace operations to peacetime information operations.

[19] Hendrycks, Scmidt, and Wong seem to think such an environment of pervasive, ASI-preclusive sabotage would be fairly easy, though they may overstate this.

[20] One imagines, for instance, that AI-driven productivity gains would offer significant economic advantages, initially in the technology sector and in white-collar work, and then in manufacturing as robotics improves. Human-AI teaming approaches could also powerfully advance scientific discovery, facilitating allowing for major breakthroughs in biology, chemistry, medicine, and physics – each advance creating its own follow-on. economic growth opportunities. In the more benign of its end-of-arms-race scenarios, for instance, the "AI 2027" paper imagines a future in which the results are generally good, with some qualifications:

> "Robots become commonplace. But also fusion power, quantum computers, and cures for many diseases. Peter Thiel finally gets his flying car. Cities become clean and safe. Even in developing countries, poverty becomes a thing of the past, thanks to UBI and foreign aid. As the stock market balloons, anyone who had the right kind of AI investments pulls further away from the rest of society. Many people become billionaires; billionaires become trillionaires. Wealth inequality skyrockets. Everyone has 'enough,' but some goods – like penthouses in Manhattan – are necessarily scarce, and these go even further out of the average person's reach. And no matter how rich any given tycoon may be, they will always be below the tiny circle of people who actually control the AIs. … In a few years, almost everything will be done by AIs and robots. Like an impoverished country sitting atop giant oil fields, almost all government revenue will come from taxing (or perhaps nationalizing) the AI companies."

[21] Hendrycks, Schmidt, & Wong, "Superintelligence Strategy," 12-13.

[22] Ibid., 13.

[23] Ibid., 14.

[24] Ibid., 15.

[25] Ibid., 15.

[26] This is why cyberwarriors generally much prefer to *pre-emplace* cyber exploits during peacetime, rather than to trying to implant them once a conflict is underway.

[27] The key problem with which the Acheson-Lilienthal Report struggled was that even after the establishment of an international organization to monopolize research and development of nuclear technology and thus take such work out of the hands of the world's rivalrous nation-states – the idea of such an Atomic Development Authority under the United Nations being the central proposal of the Report – it would be difficult to prevent countries in which this organization's facilities were physically sited from *seizing* those facilities and using the equipment and materiel there to build nuclear weapons. "It is not thought," the Report lamented, "that the Atomic Development Authority could protect its plants by military force from the overwhelming power of the nation in which they are situated." (Even a United Nations guard force, it conceded, would "at most … be little more than a token.") Chester I. Barnard, J. R. Oppenheimer, Charles A. Thomas, Harry A. Winne, & David E. Lilienthal, "A Report on the International Control of Atomic Energy Prepared for the Secretary of State's Committee on Atomic Energy" (Acheson-Lilienthal Report) (March 16, 1946) [hereinafter "Acheson-Lilienthal Report"], Section III, ch. 2.

[28] Acheson-Lilienthal Report, Section III, ch. 2.

[29] Jonathan Schell, *The Abolition* (Knopf, 1984), 118.

[30] Schell, *The Abolition*, 118-20 & 158.

[31] This is something that one of the authors of this paper addressed – in the context of nuclear weapons abolition – in a 2010 paper at Hudson Institute, and it is a concern also stressed by the late great nuclear strategist Thomas Schelling in a 2009 article in *Daedalus*. *See* Thomas Schelling, "A World Without Nuclear Weapons?" *Daedalus* (Fall 2009).

[32] Schelling, "A World Without Nuclear Weapons?" 127.

[33] As Schelling put it, in such a world "[t]he urge to preempt would dominate; whoever gets the first few weapons will coerce or preempt." Ibid. This is also a problem also noted by the nuclear strategist Herman Kahn in 1960, when he observed that [d]isarmament can … create pressures toward preventative war. If a disarmament agreement breaks down and if one side obtains a significant lead either because of previous evasion or greater ability to rearm, then it might feel compelled to perform a great public service by arranging a stop to the arms race before a dangerous balance of terror was restored. It could do this most reliably by stopping the cause of the arms race – its opponent." Herman Kahn, *On Thermonuclear War* (Princeton University Press, 1960), 230.

[34] It would be a complication, at the least, that in a MAD-type deterrent standoff involving ASI, the superintelligence *itself* might be presumed to have agency. Even if with respect to "our" ASI, therefore, would *its* behavioral payoff matrices align completely with our own? And if not, what would happen? (What would *the ASI* want to accomplish in such circumstances, and would this be what we ourselves wished? Would it remain a tool in *our* MAD standoff with another power, or would *we* end up being tools in *its* relationship with our rival power's own ASI?) Particularly if the two ASIs involved in such a MAD world were both smarter than we are, it is presumably inherently hard to imagine how their relationship would develop. In any event, we are not yet at the point of anyone having and weaponizing some form of superintelligent capability.

Nor, for the reasons suggested earlier, is there any guarantee that we could get through the point of *someone* acquiring "first-mover" advantage in ASI without seeing its preemptive use against any "fast followers" before they acquired such a capability too. The United States did *not* choose nuclear preemption against the USSR during the brief years of 1945-49 when it enjoyed a nuclear monopoly, but who is to say what the ASI "first mover" of tomorrow might choose? (Nor is it clear what ASI "strategic preemption" would even mean in the first place. One of the problems of hypothesizing about superintelligence is that since it would by definition be more intelligent than we are, it is hard to know exactly what having it "on your side" in competition would allow you to do.)

[35] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction (opened for signature April 10, 1972) (entered into force March 26, 1975), Art. I, https://treaties.unoda.org/t/bwc.

[36] Guy Roberts, "The Failure of the Biological Weapons Convention Protocol and a New Paradigm for Fighting the Threat of Biological Weapons," INSS *Occasional Paper* No. 49 (March 2003), ix, https://apps.dtic.mil/sti/tr/pdf/ADA435071.pdf.

[37] Hendrycks et al., "Superintelligence Strategy," 17-20.

[38] Kokotajlo et al, "AI 2027," 5-6.

The Post-Truth Information Environment, Artificial Intelligence, and International Security: Initial Scenarios

by

Gary L. Geipel

Introduction

This paper offers three initial scenarios for how Artificial Intelligence (AI) and related technologies may transform international security affairs in the years ahead. While the scenarios are new, they build squarely on existing work by this author dealing with the "post-truth information environment," its components, and its existing and potential impact on international security.[1] "Post-truth" is defined here as an information environment characterized in particular by "truth decay," in which verifiable facts are widely ignored or distrusted, and in which such actual facts are replaced by mere opinions or even by outright fabrications.[2] In my larger analysis, the major components of a post-truth environment are: (1) the embrace of "narratives" rather than fact-based accounts of the world; (2) increasing "tribalism;" and (3) a breakdown of corrective institutions, leading to the "entrenchment" of these conditions on a massive scale. (See **Figure 1** for a summary graphic that the reader may find useful throughout this paper.)

After providing background on the existing framework for examining the implications of a post-truth culture in the context of international security, this paper offers three general scenarios dealing with the potential longer-term impact of AI, and then suggests brief conclusions about possible further research.

Background

Key Element	General Threats	International Relations / Security Scenarios
Narratives	Information Accuracy	Designed Crises / Ignorance
Tribalism	Decision Quality	Epistemic Coups
Entrenchment	National Resilience	Fatal Distractions

Figure 1: International Relations/Security, Post-Truth – A Definition

The large-scale narratives that characterize online information exchange consist of individual assertions that cohere into larger notions of how some aspect of the world works. However, narratives are not collections of evidence presented for questioning and eventual reassessment, in the manner of scientific paradigms. Instead, today's dominant narratives usually emerge from dramatic events and fragments of information, but evolve quickly into rigid dogmas – in the United States, for example: narratives claiming to describe rigged elections, systemic racism, the power of the Deep State, catastrophic climate change, the Great Replacement conspiracy, or Settler Colonialism. In contrast to scientific truths, moreover, *narrative truths* do not even aspire to move from objectively discernible data to generalized (and thereafter potentially falsifiable) conclusions; rather, they proceed in reverse, and with little or no falsifiability: any verifiable evidence must conform to the narrative if it is to be considered at all. (All else is just "disinformation" or "fake news," offered at least out of ignorance – and more likely out of malevolence – and under no circumstances to be credited.)

The notion of what constitutes "news" itself has been upended in this environment, as the assembly of narrative-conforming storylines by "influencers" replaces anything resembling objective journalism. As political scientist Jon Askonas aptly describes it:

> Today, journalists sell compelling narratives that mold the chaotic torrent of events, Internet chatter, and information into readily understandable plotlines, characters, and scenes. … Like Scheherazade, if they can keep subscribers coming back for more of the story, they will stay alive.[3]

The objective is not to achieve objective truth – or at least to get as close to such a thing as possible – but instead simply to maximize reader (or viewer) engagement.

Tribalism, meanwhile, describes the increasing segregation of individuals into antagonistic groups based on cultural, ethnic, and religious affinities, partisan alignments, and/or geographic proximity. Social media platforms encourage – indeed all but compel, via powerful algorithms – the clustering of these tribes into separate silos where the only available information confirms the particular narratives to which they have subscribed or succumbed. In this environment, many institutions that once offered correctives – such as traditional news organizations, universities, and even scientific organizations[4] – have taken the path of least resistance and greatest profit, opting to protect and further entrench prevailing narratives and tribalism rather than to challenge them.[5]

As described in my previous work, general threats to international relations arise from the current information environment because a post-truth environment places key values at risk: (1) the accuracy of information in widespread circulation; (2) the ability to ensure quality decision-making amid epistemic chaos; and (3) the ultimate resilience of a nation (or indeed any political or social community) if it operates without a shared fact base. These threats include what I call "designed crises" (exaggerated or fabricated situations compelling individual and/or government responses); "epistemic coups" (the effective silencing of information not consistent with tribal narratives); and "fatal distractions" (the elevation of post-truth crusades above national consensus on tangible threats); specific examples are provided in my earlier work.[6]

The following scenarios build on this framework, exploring how the advent of AI may accentuate such problems. Each scenario rests on one of the three key elements of a post-truth environment and examines the risks of AI proliferation to that associated "key value." **Figure 2** provides a visual depiction of this structure.

Key Element	Narratives	Tribalism	Entrenchment
Threatened Value	Information Accuracy	Decision Quality	National Resilience
AI-Driven Scenarios	Competing Realities	Catastrophic Disconnect	Virtual Retreat
Essence	**"Truth is Hard to Find"**	**"Truth is Not the Goal"**	**"Truth is Unknown"**

Figure 2: From Post-Truth to AI-Driven Scenarios for International Relations

Scenario 1—Competing Realities: "Truth is Hard to Find"

Key Element	Narratives
Threatened Value	Information Accuracy

Lying, manipulation of information, and the deliberate creation of propaganda have always played roles in international relations and war, both on the home front and on the battlefield. Artificial Intelligence will undermine the accuracy of available information about international relations to a much greater degree, however, potentially creating a scenario of "Competing Realities" between and within societies, in which governments are both players and targets.

This scenario requires only the smallest leap from recent experience. In a recent book and other work, disinformation researcher Renée DiResta coined the term "bespoke realities" to

describe the existing ability of any information consumer in the social media ecosystem to curate incoming information or simply accept a narrative on almost any topic, tailored to that consumer's prior beliefs and prejudices.[7] In an inversion of the scientific method, these narratives provide (or at least screen incoming data in order to find) exaggerated or fabricated information that supports existing biases, rather than deriving reasonable conclusions from verifiable facts. Social media algorithms, in fact, are practically built to do this.

Bruno Maçães, an analyst of global strategy and former diplomat, explains how virtual reality (VR) environments powered by AI and optical tools will make such bespoke realities even more plausible and tangible. He compares VR to the "enchantment" formerly associated with the fantasy worlds of J.R.R. Tolkien and similar fiction writers: "as with immersive technologies, the creator of fairy tales builds a secondary world that your mind must enter."[8]

The existing post-truth information environment already primes large-scale acceptance of unsupported claims and outright fabrications on international issues, abetted by the news media and ostensibly neutral organizations.[9] By supercharging the creation of exaggerated or fabricated information, however, AI will accelerate the spread of bespoke realities in the realm of international relations, likely consolidated into a small number of Competing Realities where any given conflict is concerned.

Today, what some observers already call "the war inside the war" describes the battle of narratives unfolding in parallel with military and geopolitical clashes.[10] Tomorrow, such efforts at reality-creation will occur at a previously unimagined level of quality, quantity, and speed. AI-driven tools will encourage attempts by savvy individuals, organizations, and authoritarian nations to create and reinforce bespoke worldviews on a large scale, serving their interests in real-world conflicts and even manufacturing new conflicts entirely.

The retired director of the U.S. Department of Defense's Joint Artificial Intelligence Center (JAIC), Jack Shanahan, points to the significance of such creation: "Given how humans have been shaped

by roughly 200,000 years of evolutionary selection pressures, perception is often indistinguishable from – or at least often accepted as – reality."[11] Already, according to a company offering narrative-tracking software to monitor the information environment:

> AI-driven content engines can: mass-produce articles, videos, and social media posts that mimic human-created content; manufacture engagement by generating comments, likes, and shares to boost credibility; [and] amplify specific narratives, making them appear more widely accepted than they actually are.[12]

At the same time, the verisimilitude of so-called "deepfake" photos and videos increasingly defies differentiation by human viewers, forcing news organizations, social media platforms, and government authorities to rely on detection technology that itself increasingly risks falling behind the pace at which deepfakes are improving. AI will only deepen the challenge.

Detection technology – and, of course, the organizations using it – will also be subject to potential manipulation, and such technology is likely to be readily available to fewer and fewer users as ever greater complexity and sophistication is required in order to identify a good fake. Moreover, subsequent corrections (if they are made at all) rarely erase the full damage of initially false or misleading reports, as coverage of the war in Gaza has demonstrated, even in the absence of deepfakes.[13] The result may be what I term "authoritative fabrication," through which incorrect information gains the imprimatur of accuracy even among those who still attempt to distinguish fact from fiction at all.

Through its ability to create misleading but utterly convincing content, AI will accelerate and intensify existing practices such as "rage farming," whereby targeted groups of people can be inflamed against an individual, organization, or nation via their media feeds. Countless social media-driven "cancellations" in recent years have proven the effectiveness of this approach in ending careers and forcing otherwise unimaginable institutional changes. Abetted by AI in the

Competing Realities scenario, such practices on the part of governments, activist groups, and armed belligerents alike will become the wholesale norm – serving to trigger, prolong, and intensify conflicts when a good-faith approximation of the truth might lead to more peaceful outcomes.

AI will also worsen the Competing Realities scenario through its basic functioning. As AI's large language models (LLMs) seek "answers," they mine a digital substrate containing ever-growing amounts of false information. Some of the false information is deliberately planted, often on a large scale, by Russia and other governments and organizations that are unconstrained by the rule of law.[14] Other information is so contradictory as to bedevil the operation of LLMs, as demonstrated in a recent analysis of the Grok chatbot's efforts to make sense of the Iran-Israel conflict.[15] Training LLMs on large open data sets that themselves already contain the output of *other* LLMs is likely to worsen the situation, making errors, hallucinations, and other distortions recursively self-reinforcing. (And this is even before the well-documented, often outrageous biases of major LLMs are taken into consideration.[16]) The resulting responses to user queries sometimes consist of what subject matter experts still recognize as fabricated claims. But millions of other users will accept the AI chatbots' versions of reality, adding an additional layer of epistemological chaos to an environment already manipulated by human actors.

Under the Competing Realities scenario, authoritarian governments will be the primary curators of their own populations' views on international relations and will also attempt to disseminate favorable worldviews among adversary populations. The government of the People's Republic of China (PRC) already appears to be well-versed in such "cognitive warfare," as summarized in an article in the *People's Liberation Army (PLA) Daily*:

> In modern society, the contest around "narrative" has become an important position in the battlefield of the cognitive domain. "The same fact, different expressions"

has become an important manifestation of cognitive warfare.[17]

For their part, remaining liberal-democratic governments will seek to elevate the domestic strains of "reality" most conducive to their global aims while playing the authoritarians' game of targeted information-shaping abroad. But governments will not have this playing field to themselves. AI's effects thus will be to create rising waves of misinformed citizens, whose biases and knowledge gaps may swamp the abilities of public officials to craft effective messages. (Something similar to this occurred recently around the Israeli and U.S. strikes on Iran's nuclear facilities, when hundreds of thousands of young Americans appeared to conclude, via no source more authoritative than postings on the TikTok platform, that World War III was at hand.[18])

The information contest among all governments will thus shift increasingly from the interpretation or manipulation of actual "facts on the ground" to the outright creation of competitive realities in order to influence the larger cacophony.

Scenario 2—Catastrophic Disconnect: "Truth is Not the Goal"

Key Element	Tribalism
Threatened Value	Decision Quality

Many prominent voices claim that millions of our fellow citizens have been misled by "disinformation" or "misinformation," and can be brought back to the fold of reality if such false information is suppressed and the truth is spread. Indeed, a veritable industry of well-funded organizations grew up around that notion in the last decade, confident in their own assessments of what constitutes "truth" and eager to harness the power of government to disseminate them.[19] In this conception, the answer to disinformation is to counter it with louder and more emphatic assertions that point out error.

Even setting aside the ideological biases and blinkered realities of most self-proclaimed disinformation fighters themselves, their premise is deeply flawed. Vehemently pointing out or even suppressing error seldom makes the truth more compelling. History and numerous contemporary examples make it clear that humanity generally prefers the warm embrace of tribal belief systems to the difficult pursuit of truth, even when the verifiable facts that comprise truth are readily available, which they often are not. In a scenario of "Catastrophic Disconnect" exacerbated by Artificial Intelligence, the quality of decision-making in international relations may soon fall victim to this aspect of the post-truth information environment, which increasingly afflicts supposed "elites" and average citizens alike.

AI's role in a Catastrophic Disconnect scenario takes two major forms: one through its efficiency in aligning beliefs with tribes, and the second through its impact on human behavior and discernment.

In the first case, AI will further enhance the power of social-media algorithms and essentially take over the role of search engines in determining which information most people see about the world.[20] It will know with even more exquisite precision which beliefs and sources (accurate or not) we "like" and "subscribe" to, and which ones we avoid. It will serve to remind us, even more than we are already reminded, which beliefs and sources are approved within our tribes and which must be avoided. In that way, AI will push the online realm further from the ideal of the "Viral Editor" (in which far-flung humans might make accounts of the world richer and more accurate through their inputs) and closer to the specter of the "Viral Inquisitor" described by Canadian media scholar Andrey Mir, which "forces us into compliance."[21]

"Wrong information is tolerated when it allows the right attitude," Mir writes. "And the right information is ignored if it supports the wrong attitude."[22] This will be familiar to anyone who has attempted to rebut false information aligned with tribal dogma online. In authoritarian societies such as the PRC, governments using "social credit" systems can punish anyone trying to correct the record

while rewarding conformity with a preferred line.[23] In liberal democracies, conformity inside our powerful tribes happens somewhat more organically but no less powerfully, as exemplified by recent attitudes around culture-war topics, the 2020 global pandemic, and the conflicts in Gaza and Ukraine. Indeed, what sociologist Musa al-Gharbi describes as "symbolic capital"[24] is nearly as valuable to Westerners as the more explicit rewards of social-credit systems – and even more manipulable by AI.

Some observers take comfort in the notion that technological interventions can prevent the Catastrophic Disconnect scenario, through regulation of social media algorithms and permission mechanisms, or through the widespread adoption of so-called "middleware" to curate our information feeds back towards actual reality. Prominent political philosopher Francis Fukuyama extolled the latter possibility in 2021:

> Users could insert their preferred middleware as plug-ins to the platforms and thus choose their own trusted intermediary to sort their news, rank their searches, and order their feed of tweets.[25]

Four years on, however, no such ameliorative trend is in sight. Algorithmic fixes and middleware assume that humans will favor accurate and truthful information to give themselves societal advantages. While this may *sometimes* be true in the case of financial or medical decisions, it is closer to the opposite of how humans perceive advantage in most other information choices, where conformity and entertainment are more likely to produce the outcomes that they seek. (In the social media age, after all, what is the business model for selling software that, in effect, tells you things you do not want to hear?)

As the communications scholar Neil Postman foresaw in a comparison of George Orwell's *1984* and Aldous Huxley's *Brave New World* 40 years ago – when bespoke realities were much less prevalent than they are now – Huxley's predictions are likely to prevail in the pervasive online realm:

> In the Huxleyan prophecy, Big Brother does not watch us, by his choice. We watch him, by ours. There is no need for wardens or gates or Ministries of Truth. When a population becomes distracted by trivia, when cultural life is redefined as a perpetual round of entertainments, when serious public conversation becomes a form of baby-talk, when, in short, a people become an audience and their public business a vaudeville act, then a nation finds itself at risk.[26]

Huxley's and Postman's prescience derives from their appreciation of human nature, which is also understood by today's most successful purveyors of online falsehoods. Especially when harnessed to AI, today's information marketplace practically assures that in public understandings of international relations and war, tribe-conforming and blood-stirring narratives will only grow in power over nuanced quests for truth.

AI's second role in the Catastrophic Disconnect scenario arises from its debilitating impact on human knowledge and reasoning. Such dynamics will affect elite decision-makers as much as the proverbial man on the street, and with consequences that are likely to be all the greater precisely to the degree that such elites *do* tend to monopolize important decisions.

The widespread adoption of AI tools in education and professional life seems destined to produce human decision-makers with significantly weaker foundations of context and knowledge, minimal analytical skills of their own, and limited ability to articulate recommendations (let alone develop *informed* recommendations) independently and confidently – precisely the skill set needed in competent international-relations practitioners. By nature, these deficiencies will be more severe among younger, rising generations of decision-makers who know no other world than one in which AI does their readings for them, formulates ideas and options, and writes these up as memos, papers, and presentations without the supposed author having to absorb any actual information or even reason at all.

If the resulting work products were generally accurate, rich in detail and insight, and strong in their appreciation of humanity's strengths and weaknesses, then some of the worst aspects of Catastrophic Disconnect might be avoided. (After all, even if the "author" of a paper had not actually used his or her brain in preparing it, at least there might be wisdom in its contents.) But the work product of AI agents is none of these things and, for reasons inherent in their design, are unlikely to become any of these things in the foreseeable future. As historian and professor Kate Epstein wrote recently:

> AI is antithetical to humanistic intelligence. … Data is not knowledge; executing an algorithm is not reasoning. … [AI] tries to make up for its lack of qualitative intelligence through brute quantitative force. In so doing, it rewards virality, which, to put it mildly, is not a reliable proxy for quality. The average of lots of garbage is still garbage.[27]

Educators, such as Epstein, have begun to sound warnings about the broader effects of relying on AI. Those of us who teach at the university level now see more and more papers with the hallmarks of AI: written without grammatical errors, but lacking the basic insights or even the exuberant mistakes of actual, flesh-and-blood students. As analysis and research skills diminish in actual humans, biases and false information are more likely to be over-expressed as they go largely unnoticed by their "authors" and largely unchallenged by their supervisors and teachers. The flatness of AI-generated "learning" and writing will be reflected more and more in the flatness of the human minds who aspire to diplomatic and military decision-making.

The risks of such disconnection from the sources of competent decision-making are numerous and, as this scenario's moniker suggests, potentially catastrophic. For 80 years, to note the most obvious example, the deterrence of nuclear war has hinged on the assumption that "rational" human decision makers would have final authority over the use of nuclear weapons. And, in fact, the Cold War's troublingly frequent nuclear near-misses were averted in most

instances by humans who drew on their own knowledge and trained instincts, at times against the "evidence" presented by technology.[28] Far from lowering these stakes, AI-driven automation of surveillance and weapons systems may increase the opportunities for incompetent or lazy human operators to make poor decisions based on incomplete or misleading information. Having a "human in the loop" will be of little value if that human has been trained from childhood to avoid reasoned judgment and outsource his or her thinking to AI.

While their immediate implications for human life may be less severe than decisions about nuclear-weapons use, countless other top-level and even workaday decisions about economic instruments, arms shipments, signals to allies, troop deployments, and the use of conventional weapons systems are made better or worse by the knowledge, analytical rigor, and indeed humanity brought to bear on them. The more these skills degrade, the worse the resulting decisions will be.

Scenario 3—Virtual Retreat: "Truth is Unknown"

Key Element	Entrenchment
Threatened Value	National Resilience

The entrenchment of a post-truth information environment – powered by Artificial Intelligence – may also lead to a scenario of "Virtual Retreat," in which international relations as practiced for centuries take place, if at all, beyond the basic awareness and involvement of most human beings. Such a scenario would redefine the meaning of citizenship, undermine the ability of some governments to respond to opportunities and provocations abroad, give new technology unprecedented power over its ostensible users, and pose fundamental questions about the resilience of the nation-state.

Virtual Retreat assumes an AI-driven acceleration of recent trends that enable human interactions to occur in digital realms.

Already, hundreds of millions of people in the world's most technologically advanced societies earn their livings, learn, shop, socialize, and entertain themselves primarily through on-screen digital tools and their associated applications. Growing numbers of people are almost never away from screens except when sleeping.

While these legions remain superficially aware of their actual surroundings and cohabiting creatures while using screens, and step away occasionally for in-person interactions, even that limited non-digital engagement with the world seems likely to diminish in the years ahead. The rise of augmented reality (AR) and virtual reality (VR) technologies involving hoods, visors, and similarly "immersive" environments will allow their users to separate almost completely from physical reality and to conduct even more of their daily lives in digitally generated surroundings. Some people will resist this further detachment into a "Metaverse," fearing a loss of humanity or simply clinging to the undeniable joys of human interactions. However, many others will be drawn to environments – the ultimate in "bespoke realities" – which may be (and are likely to be *designed* to be) much more engrossing, pleasantly populated, and stimulating than their drab corners of the actual physical universe.

Maçães makes a compelling case that the "builders" and rule-setters in an emerging digital universe – including both businesses and governments – will have enormous advantages over their lagging competitors, who will be reduced to accepting rather than shaping their economic and geopolitical surroundings. "There is nothing more terrible … than to be captured by the dreams of others," he writes.[29] Initial world-building advantages, however, may be quickly overshadowed by the more fundamental risks of AI-driven Virtual Retreat.

In the emerging hyper-digital environment, an increasing amount of information about business, culture, health, politics, and war will be detached from any widely shared reality, let alone from the pursuit of objective truth. Some shared experiences will persist, as people interact with others to manage their physical existences and what remains of their offline personal and professional lives, but

individuals' subjective circumstances will become as varied as they are. In a virtual universe, one can just as easily "live" on Mars and pledge allegiance to mythical sand creatures as live in the United States and pledge allegiance to a creaking constitutional republic.

Two information-technology specialists with backgrounds in international security, Sean Guillory and John Carrola, recently coined the term "Online-Offline Convergence" to describe the widespread use of "Metaverse/Web3, synthetic training environments, Integrated Visual Augmentation Systems, digital twins, brain-machine interfaces, and other biodigital convergences," in which "a person can't differentiate between the information environment dimensions [physical, informational, and cognitive] and sees it as one 'reality.'"[30] Applying their assumptions, at least three general and unprecedented risks for international relations arise from the Virtual Retreat scenario offered here.

First, the widespread detachment of individuals from the concerns and duties of citizenship will be difficult to avoid. Throughout recent history, the primary practitioners of international relations have been a small number of elite national leaders; yet at least some degree of consent from and involvement by mass populations in public affairs have hitherto been unavoidable, even in authoritarian societies. As soldiers, taxpayers, and production workers at a minimum, populations were mobilized to confront crises, opportunities, and risks both at home and abroad. It has been hard enough at certain times, often for understandable reasons, to persuade a national majority to care about and work to prevent or reverse the provocative or threatening actions of another nation. It could become almost impossible to do this, however, in a truly comprehensive virtual environment, where the supposed machinations of another nation seem less "real" and consequential than the distractions under one's own bespoke hood.

The risk of Virtual Retreat might perhaps be worth taking, as long as the condition were universal. A world in which *everyone* lives their lives and resolves their disagreements online could be physically safer than one still beset by what latter-day analysts have already taken

to calling "kinetic war." Even authoritarian regimes may struggle to inform and encourage citizens to support conflicts that have no clear connection to their online pursuits. And it will be even more difficult to persuade younger people disproportionately seduced by online worlds to put on uniforms and risk their actual lives in tangible conflicts. Metaverse off-switches are likely to remain within the purview of national authorities, however, along with strategies to bring real-world conflicts to the attention of at least some online residents. A more likely situation is one in which physical threats emerge in one nation, even as they fail to be taken seriously in a *target* nation.

The second general risk to international relations in the Virtual Retreat scenario is that malign human adversaries either remain outside the Metaverse entirely or exploit its pervasiveness to their ends. The former variation would amount to a parallel track in which an attack in physical reality disrupts or destroys a society consumed by online life. Often, it takes considerable effort to dislodge someone from online distraction, but a missile attack or assault on essential infrastructure would likely accomplish the task. The surviving denizens of a Metaverse would stumble out into an altered physical reality that they may have lost the ability to understand and operate inside in the first place. The second variation would entail an attack inside the digital realm that manages to weaken the physical health and/or economic well-being of a targeted group. One could imagine malicious, AI-hijacked suggestions by online influencers that encourage people to consume dangerous substances or make investment decisions that doom an economy, for example.

Finally, consideration of the Virtual Retreat scenario must not ignore the possibility that the underlying technology itself could pose an "international" or societal threat. Already, the prospect of AI tools that cannot be turned off or dissuaded has moved from science fiction to actual experience. The chief executive of an influential software company noted recently that in nearly 80 percent of trials involving a common OpenAI model, the model edited a "shutdown" script to prevent the script from functioning as an off-switch; in seven percent

of cases, the model explicitly disobeyed the instruction to "allow yourself to be shut down."[31]

If indeed the off-switch to an AI-powered Metaverse were to be eliminated by the underlying "intelligence" itself – or even if certain "beliefs," biases, and components of these systems were to escape human control and the AI's self-perceived "interests" to become greatly misaligned with those of humans – the possibility of an AI-designed catastrophe could not be ruled out. It does not take the mind of a science-fiction author to imagine the possibilities: an AI able to manipulate our digital information inputs could persuade large numbers of us to take actions harmful to our survival, damage the infrastructure required to sustain modern life, engage existing weapons systems against us, or simply persuade us to attack each other in the digital and/or physical domains.

Conclusions

This is a preliminary assessment, and the most generic conclusion may be the most important one: that further work is needed. In this case:

1) Scenarios are, by nature, tools of thought provocation. Exploring and challenging their premises, blind spots, and implications is more useful than assuming any of them will prove wholly accurate.

2) Exploring additional insights from other fields will be essential. The focus here is on scenarios with the potential to disrupt international relations, but early lessons from the rise of online business, the impact of AI-enabled technologies on education, and experiences with cultural diffusion, for example, could also produce significant leads.

3) The three scenarios offered here are not mutually exclusive. Indeed, the most consistent conclusion

one draws from reviewing predictions about the effects of earlier technologies can be summed up as, "all of the above, but to varying degrees." Competing Realities may swamp the progress of truthful knowledge, for example, or it may prove to be a more virulent version of humankind's standard proclivities for fantasy and gossip. Catastrophic Disconnect may turn us all into blithering copies of a sub-par AI, or it may serve to strengthen the role of an ever-smaller cognitive elite that resists the easy paths of spoon-fed information. Virtual Retreat may be a dangerous conceit that risks extinction in a hopeless quest to escape banality and physical pain, or it may usher in a vast new domain of human creativity, competition, and conflict.

4) And of course, not everyone or every human society will respond in the same way. The human future, as it must be, will remain an endless series of experiments rather than the fulfillment of a prophecy. We have certain common, innate, and powerful tendencies that must not be ignored. But one of them is an endless ability to improvise in the quest for survival. Chances are, at least one culture and set of experiences will get it right when it comes to AI and international (read: inter-human) relations. If so, let is hope it is our own.

Meanwhile, it is time to start thinking about, debating, and planning against these possibilities in earnest.

* * *

About the Author

Dr. Gary Geipel *is a professor and director of the doctoral program at Missouri State University's School of Defense and Strategic Studies. A former global executive in non-profit organizations and corporate external affairs, he maintains a communications-consulting practice and a role as Senior Associate of the National Institute for Public Policy (NIPP). The views expressed herein are entirely his own, and do not necessarily represent those of anyone else.*

Notes

1 *See, in particular,* Gary L. Geipel, "Reality Matters: National Security in a Post-Truth World," *Occasional Paper*, vol. 3, no. 6 (National Institute Press, June 2023), https://nipp.org/papers/reality-matters-national-security-in-a-post-truth-world/.

2 *See* Jennifer Kavanagh & Michael D. Rich, "Truth Decay: An Initial Exploration of the Diminishing Rose of Facts and Analysis in American Public Life," RAND Corporate Research Report RR-2314-RC (2018).

3 Jon Askonas, "How Stewart Made Tucker," *The New Atlantis*, Summer 2022, https://www.thenewatlantis.com/publications/how-stewart-made-tucker.

4 Geipel, "Reality Matters," 16-18 & 43-44.

5 *See, e.g.*, Martin Gurri, "Journalism Betrayed," *City Journal*, Winter 2021, 12-19.

6 Geipel, "Reality Matters," 34-51.

7 Renée DiResta, *Invisible Rulers: The People Who Turn Lies into Reality* (PublicAffairs, 2024).

8 Bruno Maçães, *World Builders: Technology and the New Geopolitics* (Cambridge University Press, 2025), 10-11.

9 Recent examples include claims by the United Nations Office of the High Commissioner for Human Rights (OHCHR), echoed by prominent media outlets, that nearly 800 Palestinian civilians were killed by Israeli troops at or near food-aid facilities in Gaza from late May through early July 2025. Though OHCHR released no evidence of any kind to support its claims, the statement resulted in dozens of "news" stories and wire-service reports globally (for example, Olivia Le Poidevin, "UN reports 798 deaths near Gaza aid hubs in six weeks," *Reuters*, July 11, 2025, further establishing what is likely a gross exaggeration or a fabrication as a "fact" for millions of information consumers).

10 *See, e.g.*, EdgeTheory Lab, "The War inside the Iran-Israel War: Foreign Malign Influence and the Battle of Narratives," Narrative Intelligence Report (June 30, 2025), available at https://www.dropbox.com/scl/fi/7ja47s4uasc35yvp52zbo/The-War-inside-the-Israel-Iran-War.pdf?rlkey=0ozpasltnehuit1k3optjksc2&e=1&st=n8orpyab&dl=0.

11 Jack Shanahan, "Artificial Intelligence and Perception in Crisis and Conflict: Mating Twenty-First-Century Technologies with Eighteenth-Century Minds," in Nicholas Wright, Michael Miklaucic, & Todd Veazie (eds.), *Human, Machine, War How the Mind-Tech Nexus Will Win Future Wars* (Air University Press, 2025), 131.

12 "The New Paradigm: How AI is Shaping Narratives and Conversation," *EdgeTheory*, March 4, 2025, https://edgetheory.com/resources/ai-narratives-and-conversation.

13 For example, the *New York Times* used a short "editor's note" buried six days later on page 15 of its print edition to correct a false online headline and story that dominated its website for hours, claiming that Israel had destroyed a Gaza hospital in October 2023. *See* "Editors' Note: Gaza Hospital Coverage," *New York Times*, October 23, 2023. Similarly, the *Washington Post* took more than 48 hours to delete and acknowledge that its "standards of fairness" had been disregarded in a story falsely suggesting that Israeli troops had opened fire on and killed 30 people in "crowds making their way to collect aid;" the acknowledgement is available at https://x.com/washingtonpost/status/1929961283593367559.

[14] *See, e.g.*, Annie Newport and Nina Jankowicz, "Russian networks flood the Internet with propaganda, aiming to corrupt AI chatbots," *Bulletin of the Atomic Scientists*, March 26, 2025, https://thebulletin-org.cdn.ampproject.org/c/s/thebulletin.org/2025/03/russian-networks-flood-the-internet-with-propaganda-aiming-to-corrupt-ai-chatbots/amp/.

[15] Esteban Ponce de León & Ali Chenrose, "Grok struggles with fact-checking amid Israel-Iran war," *DFR Lab* June 24, 2025, available at: https://dfrlab.org/2025/06/24/grok-struggles-with-fact-checking-amid-israel-iran-war/.

[16] On Google's Gemini AI, *see* Toadworrier, "Google and the Gemini Debacle," *Quillette*, March 18, 2024, https://quillette.com/2024/03/18/google-and-the-gemini-debacle/; and on Elon Musk's Grok, *see* Matteo Wong, "Elon Musk Updated Grok. Guess What It Said?" *The Atlantic*, July 11, 2025, https://www.theatlantic.com/technology/archive/2025/07/new-grok-racism-elon-musk/683515/.

[17] Mao Weihao & Qin Dongyang, "Narrative Battle: Cognitive warfare in the 'post-truth era'," *PLA Daily*, July 5, 2022, http://www.81.cn/jfjbmap/content/2022-07/05/content_319117.htm.

[18] Nicole Fallert, "Gen Z, Iran, and the mass panic happening on TikTok," *USA Today*, June 23, 2025, https://www.usatoday.com/story/life/health-wellness/2025/06/23/tiktok-iran-israel-war-nuclear-panic-anxiety/84315248007/.

[19] For examples, see Geipel, "Reality Matters," 56-60; and Gary Geipel, "Information Warfare: Why America Needs a Deterrence Strategy," *Quillette.com*, February 11, 2025, https://quillette.com/2025/02/11/sticking-to-reality-misinformation/.

[20] Whatever their limitations, search engines at least have the virtue of offering the user a list of what other sources have said on a topic in a way that makes clear that *this is what others have said*, and which – by juxtaposing different alternative versions of an answer – implies at least the possibility that some of these accounts are likely to be better (*i.e.*, more accurate) than others. (Search result ranking is admittedly a problem here, of course, because it could be taken to imply that higher-ranked answers are "better" than lower-ranked ones, whereas this is by no means necessarily true; the rankings generally represent closer statistical matches to the keywords used in the search query.) AI-generated answers from an LLM, however, are in their form and format singular and oracular ones, which look and feel like an authoritative "right answer" and are likely to be taken as such without any suggestion of the need for further examination or evaluation. In reality, they are merely projections of the response that is statistically most likely to be given to the query – completely irrespective of any idea of actual truth – based upon whatever inputs happen to have been fed into the LLM's training set.

[21] Andrey Mir, "The Viral Inquisitor," *City Journal*, Spring 2023, https://www.city-journal.org/article/the-viral-inquisitor.

[22] Mir, "The Viral Inquisitor."

[23] On China, *see e.g.*, Kai Strittmatter, *We Have Been Harmonized: Life in China's Surveillance State* (Custom House, 2020).

[24] Musa al-Gharbi, *We Have Never Been Woke* (Princeton University Press, 2024).

[25] Barak Richman & Francis Fukuyama, "How to Quiet the Megaphones of Facebook, Google, and Twitter," *Wall Street Journal*, February 12, 2021.

[26] Neil Postman, *Amusing Ourselves to Death* (Viking Penguin, 1985),155-156.

[27] Kate Epstein, "We, Robots," *Persuasion*, Substack.com (May 7, 2025).

[28] *See, e.g.*, Rebecca Messina, "The Explainer: Nuclear Near Misses," *The Week UK*, April 25, 2024, https://theweek.com/defence/nuclear-near-misses.

[29] Maçães, *World Builders*, 37.

[30] Sean A. Guillory & John T. Carrola, "What Online-Offline Convergence Means for the Future of Conflict," Information Professionals Association, August 3, 2023, https://information-professionals.org/what-online-offline-convergence-means-for-the-future-of-conflict/.

[31] Judd Rosenblatt, "AI Is Learning to Escape Human Control," *Wall Street Journal*, June 2, 2025.

Traditional to Tactical: The Evolution of Female Terrorists in the PIRA and ISIS

by

Rachel Butler

Introduction

Terrorism has typically been analyzed as a male-dominated field, with female terrorists being historically understudied and often not considered a threat at all. High-profile attacks that attract media attention are generally executed by male members of terrorist organizations, reinforcing the public perception that the phenomenon of female terrorism is rare, even insignificant. It is often assumed that women are generally non-violent, and even when involved in terrorist activity, are limited to traditional roles that merely support their violent male counterparts. However, in recent years, women have increasingly assumed nontraditional roles within terrorist organizations by directly participating in acts of violence.[1] As award-winning journalist Patricia Pearson argues, the belief that women are non-violent is "one of the most abiding myths of our time."[2]

This perception of inherent non-violence can be seen in popular culture, for instance, in the media celebrity and subsequent presidential pardon given to Patty Hearst, an American heiress who claimed to have participated in armed robberies with the Symbionese Liberation Army terrorist group in 1974 only as a result of "brainwashing" by her captors after a kidnapping. It can perhaps also be seen in John Le Carre's 1983 spy novel *The Little Drummer Girl* – subsequently made into a movie starring Diane Keaton, and more recently re-adapted into a television series with Florence Pugh – in

which the heroine is recruited into a double-agent espionage plot to infiltrate Palestinian terrorist organization but later suffers what is essentially a mental breakdown caused by the strains of having been involved in such violence. For many years, when real-life female terrorists appeared – such as in the case of Vera Zasulich, a Russian revolutionary who shot and wounded the governor of tsarist St. Petersburg in 1878 – they were treated as fascinating anomalies. (Zasulich, in fact, was acquitted by a sympathetic jury and subsequently fled Russia for Switzerland.)

While there has been an increase in studies on female terrorism, the evolving role of women in terrorist activities remains largely under-examined by scholars and often mischaracterized in strategies implemented by the national security sector. The threat female terrorists pose is largely overshadowed by two prevailing narratives: that women will not act outside their natural aversion to violence, or that violent acts perpetrated by women are a result of coercion, are emotionally driven, or are simply irrational. For example, while the U.S. Congress' bipartisan *Women and Countering Violent Extremism Act* of 2019 acknowledges women as potential perpetrators of terrorist acts and their "varied roles in all aspects of violent extremism,"[3] the bill follows policy suggestions that fail to fully recognize the possibility of these actions as rational choices made by female perpetrators. Instead, they are painted even there as victims who act as a result of coercion, referred to as under "subjugation" or having a "lack of agency."[4]

Based upon an analysis of female members of the Provisional Irish Republican Army (PIRA, a.k.a. "Provos") and the Al-Khansaa Brigade of the Islamic State of Iraq and Syria (ISIS, a.k.a. *Daesh*[5]), this essay argues that women in terrorist organizations are capable of the same levels of violence as their male counterparts. Examples from these two organizations also show that women who participate in violence are capable of doing so as rational actors rather than as coerced or manipulated victims of violent males. Consequent to recognizing the potential for violence in their female members, both the PIRA and ISIS expanded women's roles beyond traditional boundaries to alleviate organizational strain. This essay demonstrates that acknowledging the transformation in the structure and operations

of those terrorist organizations, particularly regarding their female members, allows for a more accurate analysis of these groups. Yet, this evolution is often overlooked or mischaracterized, potentially hindering counterterrorism efforts.

The two organizations chosen for this essay are from different geographical locations and possess distinct ideologies, cultures, and historical contexts. However, both demonstrate a notable increase over time in female involvement in violent activities. This essay begins by examining long-prevalent narratives surrounding female terrorists and their traditional roles within terrorist organizations – specifically, the view that women who do engage in violent acts are either coerced into participating or are merely pawns of a patriarchal system. I will argue that these prevailing narratives surrounding female terrorists are misplaced, and that they represent a misunderstanding that could lead national security leaders to overlook threats.

I will then discuss the capacity of female terrorists to participate in violence by evaluating the reshaping of perceptions and redefining of roles in both the PIRA and ISIS. Those organizations deployed their female members to the tactical and operational levels, which eventually resulted in their direct participation in violence. These developments make clear that women are indeed capable of participating in violence to the same extent as their male counterparts.

The third section of this essay addresses why those terrorist organizations stopped confining female members to traditional roles, allowing them to participate in violence directly. The ongoing abuse of female members within ISIS, as well as the special punishments suffered by female members of the PIRA, suggests that the expansion of roles was presumably *not* due to male terrorist leaders' respect for women's abilities, but rather a response to organizational strain that threatened the group's survival and necessitated drawing more upon female members. By utilizing their female members as resources, both the PIRA and ISIS deviated from the traditional gender perception that women are inherently non-violent or participants in violence by virtue of coercion, all while effectively reducing the organizational strain they experienced.

Lastly, the final section transitions from analyzing the motivations by the PIRA and ISIS behind increasing female participation in violence, to exploring the motivations of female terrorists in comparison to their male counterparts. On this evidence, I argue that the violent actions perpetrated by female terrorists are not a result of coercion – or rather, at least, no more so than for male terrorists – but are rather a cognitive choice decided by a rational actor, for which female terrorists should bear responsibility and suffer appropriate consequences.

Neglecting the Female Terrorist

Despite what is now longstanding interest among scholars and national security officials in understanding the motivations and methods of terrorists, much of the research on how terrorist organizations recruit and operate has still been concentrated on males. In the aftermath of the 9/11 attacks, for example, the Department of Homeland Security (DHS) focused on males aged 16 to 45 in an effort to profile terrorists.[6] The International Peace Institute also endorsed a focus on male actors, stating, "violent extremist and terrorist groups exploit male sentiments of emasculation and loss of power and appeal to ideas of manhood in their recruitment efforts."[7]

This focus on male terrorists is warranted. After all, it was male terrorists that proved responsible for attacks that received significant media coverage, such as the Al-Qaeda hijackings on 9/11 and the recent attacks by Hamas insurgents in Israel. Yet such examples have also led to a popular assumption that females are absent from the violence enacted by terrorist groups, or that any participation by a female member should be attributed to coercion by a male member. However, entirely excluding women's participation in and capacity for violence from analysis of terrorist organizations' structure and operations dangerously overlooks their actual degree of involvement and may lead to misguided counterterrorism efforts.

Women are historically perceived as serving in roles that are traditional in nature due to their inherently non-violent nature.[8] ISIS

encouraged women to serve primarily as wives or mothers, and the original Irish Republican Army (IRA) promoted values such as purity, caregiving, and motherhood.[9] Female recruitment and membership in these groups thus primarily revolved around providing for and supporting male members, at least initially.

Yet those organizations progressively implemented women outside these traditional roles to obtain their political or ideological objectives.[10] Even so, women who increasingly participate in violent acts are labeled as displaying "irregular" behavior, with blame often assigned to male influences rather than the female perpetrators themselves.[11] The narrative that female terrorists are victims drawn into a patriarchal game, incapable of choice or reasoning, is commonly pushed.[12] For example, scholar Clara Beyler asserts that female suicide bombers are often portrayed as being in love with their handlers and viewed as "symbols of utter despair," and hence subsequently portrayed as victims.[13]

Placing the blame on male influences, however, denies the possibility that women may initiate terrorist acts of their own free will and conduct independent decision-making processes in the same ways as male terrorists do. The perception that "women are somehow less responsible than men for their role in terrorist activity" is misplaced and underestimates the capacity of female terrorists.[14]

Consequently, counterterrorism efforts that ignore this female agency may be rendered ineffective and incomplete.[15] Currently, even the few counterterrorism strategies that focus on females do so from the perspective of guarding them against coercion inflicted by their male counterparts. The U.S. State Department's 2018 *Strategy to Support Women and Girls at Risk from Violent Extremism,* for example, aims to reintegrate and rehabilitate female perpetrators, yet fails to assign them any responsibility for their actions.[16] While this policy is an effective measure for women who have suffered abuse under terrorist organizations, it treats female terrorists as similar victims. Currently, the United States possesses no such policy regarding the reintegration or rehabilitation of male terrorists. As terrorist organizations like the PIRA and ISIS showed an increase in female

participation in violence, responsibility was rarely assigned to female perpetrators, suggesting they were not accountable for their actions.

Recognizing the Capacity for Violence in Female Terrorists

Historically, female members of terrorist organizations that follow Islamic extremism have been hidden in the background of society, much less featured in propaganda.[17] A predecessor of ISIS and outgrowth of the Afghan guerrilla war against Soviet occupation, the Taliban, banned women from public gatherings and prohibited filming or photographing them. However, women began being prominently featured in ISIS propaganda, often photographed on the battlefield adorning a hijab and brandishing an automatic weapon.[18] The example of ISIS illustrates how terrorist organizations can reshape perceptions about and redefine traditional roles of their female members established by their predecessors. While many women in such organizations still operate within traditionally defined roles, they are no longer strictly confined to them and are increasingly becoming involved in tactical operations.

The transition of female members of the PIRA occupying traditional roles to directly engaging in violence manifested primarily in bombing operations. The majority of women convicted in Northern Ireland, Britain, and abroad for PIRA activity faced bombing-related charges.[19] One of the most devastating attacks carried out by the PIRA was the bombing of the Old Bailey Courthouse in London in 1973, resulting in over 200 injuries.[20] This operation was masterminded and executed by two female PIRA members, Dolours Price and Marian Price.[21] The Price sisters were subject to immediate fame, not necessarily primarily for the devastation they wrought – for although many people were wounded in the explosion, only one victim died (and it was from a heart attack) – but rather for the fact that it was women who had planned and executed a tactical bombing operation. Another female member of the PIRA who participated in car bombings, in addition to arms-buying missions across Europe, was Maria McGuire.[22] In a later interview, McGuire described her own violent tendencies in the killing of British soldiers, stating she believed "the more that were killed, the better."[23]

The responsibilities of the members of Al-Khansaa, an all-female brigade in ISIS, also demonstrated terrorist women diverging from traditional roles. Despite ISIS' prohibition upon women gaining an education, members of the Al-Khansaa Brigade were made up primarily of educated females from Western nations who had been recruited by ISIS and converted to Islam, including many from the United Kingdom and France.[24] These recruits received additional education in social media marketing, firearms and explosive training, and Islamic law.[25] Equipped with AK-47 assault rifles, brigade members engaged in intelligence gathering and recruitment operations, as well as serving as law enforcement for female ISIS members so as to avoid having male members engage with women who were not their wives of family members.[26]

Members who violated ISIS' strict interpretation of Islamic *Sharia* law, for example, were beaten and subjected to torture by Al-Khansaa. The standard punishment brigade members inflicted on women wearing high heels or headscarves with designs – both of which offended ISIS sensibilities – was 40 lashes. New mothers found breastfeeding in public (also an offense) were subjected to the "biter," a bear-trap torture device that clamped onto the breasts.[27] One community member stated, "I was much more afraid of [the ISIS] women ….The women would beat you for the smallest thing."[28]

Al-Khansaa members not only engaged in brutal acts of physical violence towards other women, but also facilitated sexual violence as well. Reports suggest that ISIS implemented controversial *fatwas* – legal rulings in Islamic law – which subjected females to significant sexual abuses.[29] These *fatwas* were enforced by the Al-Khansaa Brigade, including one in which women served through "temporary marriages" as sexual servants to male Jihadi fighters to encourage their increased performance on the battlefield.[30] By this ruling, women would be "married" to a given ISIS fighter for a week, or even a few hours, allowing them to engage in sexual activities with him without violating the belief that sexual relations should not occur outside of marriage. (Additionally, the brigade published a document in 2015 titled *Women of the Islamic State: Manifesto and Case Study*,[31] which

urged girls as young as nine years old to marry while still "young and active."[32])

The capacity for violence demonstrated by members of the Al-Khansaa Brigade ultimately resulted in a change to the strict gender separation within ISIS, as women's involvement in combat roles increased, including those of fighters and suicide bombers. The first reported instance of ISIS deploying female suicide bombers in battle occurred in Libya in 2016.[33] The Battle of Mosul, in Iraq, marked a peak in female suicide bombings, with 38 detonations targeting civilians and U.S. forces.[34] This evolution of female roles in ISIS represents a significant shift in the organization's ideology of Islamic extremism, as women began to assume roles traditionally held by men in conflict.[35]

The violence inflicted by female PIRA members, the Al-Khansaa Brigade, and later ISIS female suicide bombers illustrates how female terrorists are capable of inflicting violence comparable to their male counterparts. The assumed non-violent nature of female terrorists is effectively disproven in the examples discussed above and should result in a reevaluation of the threat these organizations posed. Without exposing the fallacy of the narrative that women are inherently non-violent, the threats presented by female terrorists may remain misunderstood and underestimated.

Organizational Strain Responsible for Reframing Female Roles

The important role played in PIRA and ISIS by female terrorists, however, does not necessarily suggest that either organization treated their female members as equals or valued them as highly as male members. Members of the Al-Khansaa brigade reported joining the ranks as they perceived life in ISIS as empowering to women, promoting their independence, and facilitating access to high-level organizational positions.[36] However, this perception of women is not reflected in the treatment female members of the PIRA and ISIS received within their respective organizations. Rather than terrorist organizations revering their female members, the redefinition of roles resulted from organizational strain. Simply put, female members were

deployed to fill resource gaps only when those terrorist organizations faced a threat to their existence.

The PIRA deployed this strategy after it suffered resource constraints resulting from its split from the IRA. Similarly, after growing and expanding its territorial reach significantly in 2006-13, ISIS expanded the roles of female members by creating the Al-Khansaa Brigade only in 2014, after the formation of a Western-led coalition against the organization. While females were permitted to operate outside their traditional confines during these times, however, it does not indicate that they were considered equals to their male counterparts.

Despite being granted participation in the tactical and operational levels of their organization, in fact, female terrorists seem frequently to have been specially punished for acting outside their traditional roles and expected behavior. In the PIRA, female members who fraternized with British soldiers had their heads shaved, then were tied to a lamppost and tarred and feathered.[37] While men were occasionally tarred and feathered as well, only women had their heads shaved, a shame-based punishment symbolic of the removal of their womanhood and femininity.[38] Female members of ISIS were also subjected to numerous forms of brutal punishment, such as the previously mentioned "biter," for violating the organization's strict interpretation of *Sharia* law. The continuous mistreatment experienced by women in these terrorist organizations showcases that rather than awakening to the capabilities of its female members, the reframing of female roles in terrorist organizations is spurred by organizational strain.

The PIRA expanded female participation almost immediately. At its inception, the organization faced strain resulting from separating from its founding organization. Founded in 1919, the IRA aimed to establish Ireland as an independent republic free of British rule. Despite numerous organizational changes throughout the 20th century, this remained the organization's primary objective.[39] However, in line with its long history of inter-organizational conflict, the IRA split into two separate factions in 1969: the Officials and the

Provisionals. Both organizations shared the original IRA's primary principles, but they differed in their tactics. Officials saw themselves as representing the original IRA and were known for advocating independence through peaceful protests and the official channels of government. The Provisionals believed peaceful methods to be ineffective for change and resorted to violence and extremism to propel their agenda.[40]

The division of the IRA effectively split resources and weakened both factions, spurring the Provisionals to establish female auxiliary groups in order to meet operational needs.[41] The closest women had previously come to directly participating in violence was simply by accompanying male IRA members on missions – effectively as cover for male fighters, as women were unlikely to be suspected of terrorist activities and male-female couples seemed less suspicious.[42] However, members of the PIRA began to challenge the traditional roles typically assigned to women in the IRA, allowing them to participate in various combat support roles.

Women began receiving the same military training as male members and were assigned roles deeply embedded on the tactical and operational levels.[43] These roles included smuggling weapons and explosives, gathering intelligence, and – in the example of the Price sisters – planning and executing entire operations.[44] By incorporating female members into conducting attacks, primarily bombings, the PIRA effectively filled organizational gaps and came to be considered "one of the most inventive and adaptive of all the violent non-state actors who operated in the latter part of the twentieth century."[45]

On the other side of the world, nearly a century after the founding of the IRA, former members of Al Qaeda, Boko Haram, and terrorist groups from Afghanistan and Libya pledged their loyalty to a newly established terrorist organization, ISIS.[46] ISIS was a self-declared Islamic caliphate that at its peak stretched from Aleppo in Syria to Diyala in Iraq.[47] Under ISIS rule, *Sharia* law was narrowly interpreted and brutally enforced, with little regard for the sanctity of life.[48] Taking advantage of a power vacuum created by the United

States' withdrawal from Iraq, ISIS achieved significant territorial gains – at one point occupying some 40 percent of Syria and Iraq, including the major Iraqi city of Mosul – and established itself as a formidable force, a feat which did not remain unnoticed by the West.[49] In 2014, a U.S.-led coalition formed and began inflicting key losses on ISIS, and by late 2017 the caliphate had lost 95 percent of its territory. Facing this coalition, ISIS' survival was threatened, and members began to deradicalize after becoming disillusioned with the organization and its self-proclaimed caliphate.[50]

The same year the anti-ISIS coalition was formed – after which ISIS began to suffer mass desertions and territorial loss – the Al-Khansaa Brigade was formed, expanding the roles of female members.[51] The evolution of female roles thus strayed from historically fundamentalist Islamic principles, as well as the methods of previous Islamic-based terrorist organizations. By establishing a female police force and subsequently placing women in frontline positions to execute Holy War suicide missions, women were no longer solely confined to being homemakers and wives.[52]

The establishment of the Al-Khansaa Brigade and the utilization of female suicide bombers helped meet the organizational pressures facing ISIS as it delegated those women some duties previously assigned to male members. Additionally, global media coverage of ISIS also increased dramatically, as brigade members were looked on with morbid fascination.[53] ISIS was thus effectively granted a broader platform for propaganda and potential recruiting, as, according to Bruce Hoffman,

> …[o]nly by spreading the terror and outrage to a much larger audience can the terrorists gain the maximum potential leverage that they need to effect fundamental political change.[54]

Consequently, ISIS intensified its recruiting efforts towards females to take advantage of this increased media coverage. Female members, in fact, were deliberately assigned roles at the organization's forefront, subject to public display and media attention. As a result,

women became leading contributors to ISIS' strength and capabilities through their increased role and visibility.

By deploying their female members and expanding their participation in violence, both the PIRA and ISIS were able to combat the threats their respective organizations faced. However, the continuous mistreatment of female members shows that the PIRA and ISIS did not grant women increased roles because they were revered; rather, this change occurred only when the survival of the organization was threatened.

Female Terrorist Motivations to Participate in Violence

Participation by female terrorists in violence has thus increased, as illustrated by the preceding analysis of the PIRA and ISIS. However, do such female terrorists exercise their own free will and independent choice when participating in violent acts outside traditionally female roles? Or does the traditional narrative of male members coercing females into such acts hold true?

Before addressing this question, it is essential to recognize that the wide variety of terrorist organizations in existence derive from a diverse array of motivations held by the individuals involved in terrorist acts. This diversity is highlighted in the research of Walter Laqueur, for instance, who states that "[m]any terrorisms exist, and their character has changed over time and from country to country …. Terrorism has changed over time and so have the terrorists, their motives, and the causes of terrorism."[55] Therefore, I do not aim here to identify a single, overarching motivation for all terrorists, regardless of gender. I argue instead only that the motivations of female members in the PIRA and ISIS seem to have closely parallelled those of male members, thus contradicting traditional narratives of female terrorist subservience to male terrorist agency. In fact, female participation in violence is not a result of coercion or manipulation but rather rational choices for which these terrorists should be held accountable.

Currently, two prevailing narratives shape the perspectives of many scholars and the national security sector regarding female terrorists. The first narrative suggests that women are inherently non-violent, a claim disproven in the preceding sections. The second narrative builds on this assumption, arguing that if women do partake in violent behavior – such as members of the PIRA and the Al-Khansaa Brigade – it is a result of coercion or manipulation rather than a conscious choice. In such a view, "women are assumed [to be] victims, irrational actors, or emotionally driven."[56] By evaluating the stated motivations of members of the PIRA and ISIS, this section argues that female participation in terrorist violence is not coerced but rather a choice made for reasons that parallel the motivations of male members.

The motivations for female members of the PIRA and ISIS to participate in terrorist activities, in fact, share a common theme. Despite the mistreatment women experienced within their respective organizations, female members of both organizations seem to have regarded this as much less important than the perceived threat they faced from their group's antagonists in the governments they opposed.

Both organizations fought for an independent state and emphasized narratives of abuse and disillusionment at the hands of prevailing state authorities. Members of ISIS, for instance, were committed to the ideology of an Islamic caliphate, believing Western influence and corrupt regional governments to be restricting the practice of true Islam.[57] The primary motivation of the Al-Khansaa Brigade, most of which were recruited from outside Iraq and Syria, was to preserve Islamic culture and religion from what was perceived to be an increasingly intrusive Western world.[58] The manifesto published by Al-Khansaa members claimed that the "Western model" for women had failed, and had in fact inserted corrupted ideas into the feminine mind.[59] They feared that under Western influence, Muslim women would become sex objects, their roles as mothers and nurturers eliminated, and that greed created by capitalism would render the family and religion irrelevant.[60] By joining ISIS, female members thus saw a chance to contribute to building an alternative state and

subsequently to defeat the "decadent and morally corrupt Western society, which has no respect for women."[61]

For their part, the PIRA advocated for a free Irish state in response to what they said was the brutality and oppression of British occupation. Female members of the PIRA stated that experiencing state-sanctioned violence and heavy-handed tactics by security services led them to join terrorist activities.[62] Mairéad Farrell, for example – a PIRA member who had been deeply embedded in tactical operations – later discussed having been radicalized by the presence of British soldiers in Belfast. Additionally, Farrell recalled the mistreatment of Catholic populations interned in Northern Ireland.[63]

Conditions of treatment for women in the Armagh prison were also a grievance, for that institution reportedly conducted strip searches and assaulted women, and those who had newborns had their babies taken from their arms and were subject to grotesque living conditions.[64] Some women who were not directly subjected to this abuse but witnessed it were also motivated to join the PIRA. (Rose Dugdale, for example, stated she had joined after witnessing the state-sanctioned persecution, inequality, and brutality suffered under the British system.)[65] Another female member stated,

> I grew up in the conflict and war was all around. I gained a political awareness when I was twelve or thirteen and I started asking questions about who is responsible for all of this.[66]

These women joined the PIRA hoping to achieve independence for the group with which they identified, political participation for themselves, and vindication for other women mistreated in such ways, utilizing violence to fight back against perceived persecution.[67]

The motivation to participate in violence among female members of the PIRA and ISIS thus parallels the very similar motivations possessed by male members. Notably, the United Nations Security Council Counter-Terrorism Committee states that drivers of female radicalization – specifically disempowerment,

resentment, and marginalization – do not differ from those of men.[68] And indeed, these three themes are prominent motivators for both male and female members of the PIRA and ISIS.

Interviews of PIRA members during the 1960s, for instance, revealed that most members joined out of a sense of "hopelessness, despair and betrayal by the system."[69] Interviews of 220 ISIS returners, defectors, and prisoners conducted by Anne Speckhard and Molly Ellenber similarly revealed that male interviewees who were recruited outside of Iraq and Syria had been motivated by "a desire to bolster their Islamic identities, which are often under attack by Islamophobic sectors of society." Female interviewees also wished to bolster their "Islamic identity," for they were said to "suffer the most from Islamophobic attacks in the West, as their identities are so clearly marked due to their Islamic dress."[70]

The motivations expressed by male and female terrorists of the PIRA and ISIS are thus mirror images. Therefore, it would be incorrect to state that women who participate in violence can only have been coerced into doing so. That is not to say that female terrorists who exhibit violent behaviors are *never* coerced, of course, as terrorism itself is diverse, and so are the motivations within. However, the examples displayed by the PIRA and ISIS, along with their stated motivations, disprove the traditional narrative that female terrorists are not capable of violence without their participation being the result of coercion or other manipulation.

Conclusion

By analyzing the PIRA and Al-Khansaa Brigade of ISIS, this essay has shown that female members of terrorist organizations are indeed capable of willingly exhibiting the same level of violence as their male counterparts. Female terrorists who choose to participate in violent acts are, no less than men, rational actors, and they often display motivations and reasoning very similar to those of male members. Terrorist organizations such as the PIRA and ISIS, moreover, recognize the ability and willingness of female members to

engage in violence and have utilized this to their advantage, especially in times of organizational strain.

While men are more present in combat roles within terrorist organizations than women, terrorist organizations are increasingly utilizing their female members to achieve their ideological goals. For the PIRA and ISIS, integrating female members into tactical operations proved successful and also allowed those organizations a strong platform to recruit and convey their message. There is no reason to expect that terrorist organizations will not continue to integrate female members into their violent operations, while their victims – and the security services who fight terrorism – would be wise to recognize and combat this trend.

Despite the geographical, cultural, and historical differences separating the PIRA and ISIS, both organizations exhibited similar behaviors surrounding the deployment of their female members. For too long, however, many scholars and professionals in the national security sector have assumed that female terrorists are victims rather than perpetrators[71] and that their actions are emotionally driven or coerced, and that they should therefore be held less responsible than their male counterparts. It is time, however, to acknowledge two key realities. First, female terrorists are moving away from their traditional roles and are actively (and increasingly) participating in violence as rational actors. Second, this trend is evident across terrorist groups in very different contexts, including those in cultures that traditionally assign women to sharply subservient roles.

Female terrorists may present a disproportionate threat as long as the dangers they present continue to be underestimated, with such false assumptions increasing their effectiveness by making their violence seem surprising or anomalous. As terrorist organizations adapt their organizational structure and challenge their historical approach, scholars and the national security sector should follow suit and recognize the evolution of these groups. The threat of female terrorists must be incorporated into analysis and integrated into counterterrorism strategies. Additionally, female terrorists should be held to the same account that males are. Prosecution of these

individuals should mirror that of male members, for they display the same decision-making process and intent.

* * *

About the Author

Rachel Butler *is a doctoral student in the Department of Defense and Strategic Studies at Missouri State University. She holds master's degrees in history and strategic studies, with research interests focused on ethical warfare, energy security, and non-proliferation and disarmament. The views expressed herein are entirely her own, and do not necessarily represent those of anyone else.*

Notes

1 Lea-Grace B. Salcedo, "Terrorism: Two Faces of Women," George C. Marshall European Center for Security Studies, February 2019, https://www.marshallcenter.org/en/publications/perspectives/terrorism-two-faces-women.

2 Patricia Pearson, *When She Was Bad: How and Why Women Get Away with Murder* (Penguin Group, 1998), 28.

3 U.S. Congress, House. *Women and Countering Violent Extremism Act of 2019*, HR, 116th Cong., 1st sess., introduced in House March 5, 2019, https://frankel.house.gov/uploadedfiles/women_and_countering_violent_extremism_act_of_2019.pdf.

4 Ibid., sec 2.

5 *Daesh* is a word growing out of the acronym formed by the organization's name in Arabic: *al-Dawla al-Islamiya fil Iraq wa al-Sham.* ISIS was sometimes also known in the West as ISIL (the Islamic State of Syria and the Levant).

6 No byline, "Gender and Terrorism: Motivations of Female Terrorists," (M.S. Strategic Intelligence, Joint Military Intelligence College, 2005), 1.

7 Annie Schmidt, "Masculinities and Violent Extremism," International Peace Institute, June 9, 2022, https://www.ipinst.org/2022/06/masculinities-and-violent-extremism.

8 Pearson, *When She Was Bad: How and Why Women Get Away with Murder*, 34.

9 Amanda Spencer, "The Hidden Face of Terrorism: An Analysis of the Women in Islamic State," *Journal of Strategic Security* 9, no. 3 (September 1, 2016), https://doi.org/10.5038/1944-0472.9.3.1549, 81 & 82; Fionnuala Ni Aolain, "Situating Women in Counterterrorism Discourses: Undulating Masculinities and Luminal Femininities," *Boston University Law Review* 93, no. 3 (May 1, 2013), https://experts.umn.edu/en/publications/situating-women-in-counterterrorism-discourses-undulating-masculi, 1087.

10 Naureen Chowdhury Fink, Rafia Bhulai, & Liat Shetret, "The Roles of Women in Terrorism, Conflict, and Violent Extremism: Lessons for the United Nations and International Actors," Center on Global Counterterrorism Cooperation, April 2013, https://globalcenter.org/resource/the-roles-of-women-in-terrorism-conflict-and-violent-extremism-lessons-for-the-united-nations-and-international-actors/.

11 "Gender and Terrorism: Motivations of Female Terrorists," 12.

[12] Ibid., 14.

[13] Alexis B Delaney & Peter R Neumann, "Another Failure of Imagination?: The Spectacular Rise of the Female Terrorist," *The New York Times*, September 6, 2004, https://www.nytimes.com/2004/09/06/opinion/another-failure-of-imagination-the-spectacular-rise-of-the-female.html.

[14] "Gender and Terrorism: Motivations of Female Terrorists," 2 & 14.

[15] Lora Vonderhaar, "ISIS's Female Morality Police," Georgetown Security Studies Review, May 13, 2021, https://georgetownsecuritystudiesreview.org/2021/05/13/isiss-female-morality-police/.

[16] U.S. Department of State, Office of Global Women's Issues, "U.S. Strategy To Support Women and Girls at Risk From Violent Extremism and Conflict," October 1, 2018, https://2017-2021.state.gov/u-s-strategy-to-support-women-and-girls-at-risk-from-violent-extremism-and-conflict/.

[17] Mah-Rukh Ali, "ISIS and propaganda: How ISIS Exploits Women," *Reuters Institute Fellowship Paper* (University of Oxford, 2015), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Isis%2520and%2520Propaganda-%2520How%2520Isis%2520Exploits%2520Women.pdf, 12. & Amanda N. Spencer "The Hidden Face of Terrorism: An Analysis of the Women in Islamic State." *Journal of Strategic Security* 9, no. 3 (2016): 74–98. http://www.jstor.org/stable/26473339, 77.

[18] "Gender and Terrorism: Motivations of Female Terrorists," 5.

[19] Charitini (Hari) Ntini, "'Where Were the Women?': Women Active Service Members of the Provisional IRA in the Northern Ireland Conflict," *Writing the "Troubles,"* May 2, 2022, https://writingthetroublesweb.wordpress.com/2022/05/02/women-active-service-members-of-the-pira/.

[20] Simone Matassa, "Women of the Irish Republican Army: Powerful or Powerless?," *Rise to Peace*, July 10, 2019, https://www.risetopeace.org/2019/07/10/women-of-the-irish-republican-army-powerful-or-powerless/smatassa/.

[21] Rebecca Dougherty & P. Kathleen Frier, "Gender and Violent Extremism: Examining the Psychology of Women Participating in Non-State Armed Groups" (George Washington University, 2016), https://wiisglobal.org/wp-content/uploads/Gender-and-Violent-Extremism-Examining-the-Psychology-of-Women-Participating-in-Non-State-Armed-Groups-1.pdf, 21.

[22] Deaglán De Bréadún, "How IRA Links Came Back to Haunt Tory," *The Irish Times*, December 6, 2008, https://www.irishtimes.com/news/how-ira-links-came-back-to-haunt-tory-1.920306.

[23] Maria McGuire, *To Take Arms: My Year with the IRA Provisionals* (New York: Viking, 1973).

[24] Tova Dvorin, "Sixty British Women Join IS 'Modesty Police,'" *Israel National News*, September 8, 2014, https://www.israelnationalnews.com/news/184891; Erin M. Saltman & Melanie Smith, "Till Martyrdom Do Us Part: Gender and the ISIS Phenomenon," *Institute for Strategic Dialogue*, no. 2 (2015), 16, https://icsr.info/wp-content/uploads/2015/06/ICSR-Report-'Till-Martyrdom-Do-Us-Part'-Gender-and-the-ISIS-Phenomenon.pdf.

[25] Joanna Paraszczuk, "Matchmaking and Marriage, Islamic State-Style," *Radio Free Europe/ Radio Liberty*, January 3, 2015, http://www.rferl.org/content/matchmaking-marriage-islamic-%20state-style/26774719.html.

[26] Spencer, "The Hidden Face of Terrorism: An Analysis of the Women in Islamic State," 84.

[27] Loulla-Mae Eleftheriou-Smith, "Escaped Isis Wives Describe Life in the All-female al-Khansa Brigade Who Punish Women With 40 Lashes for Wearing Wrong Clothes," *The Independent*, April 20, 2015, https://www.independent.co.uk/news/world/middle-east/escaped-isis-wives-describe-life-in-the-allfemale-alkhansa-brigade-who-punish-women-with-40-lashes-for-wearing-wrong-clothes-10190317.html; No byline, "ISIS's Persecution of Women," Counter Extremism Project, 2021, https://www.counterextremism.com/content/isiss-persecution-women.

[28] Lucy Kafanov, "How All-Female ISIS Morality Police 'Khansaa Brigade' Terrorized Mosul," NBC News, November 20, 2016, https://www.nbcnews.com/storyline/isis-uncovered/how-all-female-isis-morality-police-khansaa-brigade-terrorized-mosul-n685926.

[29] Ali, "ISIS and propaganda: How ISIS Exploits Women," 17.

[30] Ibid.

[31] Ibid, 13.

[32] Ibid, 14.

[33] Bel Trew, "Isis Sends Women into Battle in Libya," *The Times*, February 29, 2016, https://www.thetimes.com/uk/crime/article/isis-sends-women-into-battle-in-libya-rjmhqc7k7?msockid=0c85e4aba495616e0db5f114a5b0603a&region=global.

[34] Jack Moore, "ISIS Unleashes Dozens of Female Suicide Bombers in Battle for Mosul," *Newsweek*, July 5, 2017, https://www.newsweek.com/isis-female-suicide-bombers-battle-mosul-631846#:~:text=Thirty-eight%20women%20have%20detonated%20suicide%20explosives%20to%20target,Iraqi%20security%20official%20told%20The%20Times%20of%20London.

[35] Gan, Ruth, Loo Seng Neo, Jeffery Chin, & Majeed Khader, "Change Is the Only Constant: The Evolving Role of Women in the Islamic State in Iraq and Syria (ISIS)," *Women & Criminal Justice* 29 (4–5), 2019, https://www.tandfonline.com/doi/abs/10.1080/08974454.2018.1547674.

[36] Paul Sperry, "Meet the American women who are flocking to join ISIS," *New York Post*, May 13, 2017, https://nypost.com/2017/05/13/meet-the-western-women-who-are-flocking-to-join-isis/.

[37] Theresa O'Keefe, "'Mother Ireland, Get off Our Backs': Republican Feminist Resistance in the North of Ireland," in *Amsterdam University Press eBooks*, 2017, 176, https://doi.org/10.1515/9789048528639-010.

[38] Juliane Röleke, "Policing Without Police. Republican Paramilitary 'Punishment' Attacks on Women in Northern Ireland 1971 – 1979," *History | Sexuality | Law*, November 25, 2020, https://doi.org/10.58079/ppwe.

[39] Matassa, "Women of the Irish Republican Army: Powerful or Powerless?".

[40] Ibid.

[41] Dougherty & Frier, "Gender and Violent Extremism: Examining the Psychology of Women Participating in Non-State Armed Groups."

[42] Ntini, "'Where Were the Women?"

[43] O'Keefe, "Chapter 9. 'Mother Ireland, Get off Our Backs', 193; Eileen Fairweather, Roisín McDonough, & Melanie McFadyean, *Only the Rivers Run Free: Northern Ireland, the Women's War* (Pluto Press, 1984), 242.

[44] Mia Bloom, Paul Gill, & John Horgan, "Tiocfaidh Ár Mná: Women in the Provisional Irish Republican Army," *Behavioral Sciences of Terrorism and Political Aggression* 4, no. 1, December 8, 2011, 65, https://doi.org/10.1080/19434472.2011.631345.

[45] Gary Ackerman, "The Provisional Irish Republican Army and the Development of Mortars," *Journal of Strategic Security* 9, no. 1, 2016, 12, https://doi.org/10.5038/1944-0472.9.1.1501.

[46] Ali, "ISIS and propaganda: How ISIS Exploits Women," 4; The Wilson Center, "Timeline: The Rise, Spread, and Fall of the Islamic State," October 28, 2019, https://www.wilsoncenter.org/article/timeline-the-rise-spread-and-fall-the-islamic-state.

[47] "Timeline: The Rise, Spread, and Fall of the Islamic State."

[48] Ali, "ISIS and propaganda: How ISIS Exploits Women," 4.

[49] "Timeline: The Rise, Spread, and Fall of the Islamic State."

[50] Anne Speckhard & Molly Ellenberg, "ISIS in Their Own Words: Recruitment History, Motivations for Joining, Travel, Experiences in ISIS, and Disillusionment Over Time – Analysis of 220 In-depth Interviews of ISIS Returnees, Defectors and Prisoners," *Journal of Strategic Security* 13, no. 1, March 16, 2020, 120, https://doi.org/10.5038/1944-0472.13.1.1791.

[51] Vonderhaar, "ISIS's Female Morality Police."

[52] Spencer, "The Hidden Face of Terrorism," 77.

[53] Ibid., 4.

[54] Bruce Hoffman, *Inside Terrorism*, *Columbia University Press eBooks*, 2017, 174, https://doi.org/10.7312/hoff17476.

[55] Walter Laqueur, *No End to War: Terrorism in the Twenty-First Century* (Continuum, 2003), https://doi.org/10.5040/9781501301070.

[56] "Gender and Terrorism: Motivations of Female Terrorists," 2.

[57] Salcedo, "Terrorism: Two Faces of Women."

[58] Christine Sixta, "The Illusive Third Wave: Are Female Terrorists the New 'New Women' in Developing Societies?" *Journal of Women Politics & Policy* 29, no. 2, June 1, 2008, 262, https://doi.org/10.1080/15544770802118645.

[59] Ali, "ISIS and propaganda: How ISIS Exploits Women," 13.

[60] Ibid., 263.

[61] Mark Tran, "Police Stop Plane at Heathrow 'To Prevent 15-year-old Girl Flying to Syria,'" *The Guardian*, December 14, 2014, https://www.theguardian.com/uk-news/2014/dec/17/police-stop-plane-heathrow-girl-syria.

[62] Dougherty & Frier, "Gender and Violent Extremism," 26.

[63] Ibid., 19.

[64] "Women in Rebellion: How Paramilitary Women Experienced the Troubles of Northern Ireland - Global Irish Studies," *Global Irish Studies*, September 27, 2022, https://globalirish.georgetown.edu/women-in-rebellion/#.

[65] Dougherty and Frier, "Gender and Violent Extremism," 20.

[66] Ibid., 169.

[67] "Mairéad Farrell interview," July 1987, https://www.youtube.com/watch?v=33Kc-woTsto and https://www.youtube.com/watch?v=xZUeCemzkt4.

[68] Counter-Terrorism Committee Executive Directorate (CTED), "CTED Trends Report," *CTED Trends Report*, n.d., 13, https://www.un.org/securitycouncil/ctc/sites/www.un.org.securitycouncil.ctc/files/feb_2019_cted_trends_report_0.pdf.

[69] Andrew Garfield, "PIRA Lessons Learned: A Model of Terrorist Leadership Succession," *Low Intensity Conflict & Law Enforcement* 11, no. 2–3, June 1, 2002, 274, https://doi.org/10.1080/0966284042000279036.

[70] Speckhard & Ellenberg, "ISIS in Their Own Words," 124 & 122.

[71] Matassa, "Women of the Irish Republican Army."

The Under-Appreciated Problem of Religious Ideology in Nuclear Proliferation

by

Inderjit S. Grewal

Introduction

The term "Hindu Bomb" is sometimes used in geopolitical discourse to describe India's nuclear weapons program. Coined after India's first nuclear test in 1974, it implies a connection between India's Hindu majority population and its nuclear ambitions. India's official rhetoric surrounding its nuclear weapons has focused on strategic needs, but the term "Hindu Bomb" – which is not a term used only by India's critics but, importantly, one that is also used by some Hindu nationalists within India itself – has symbolic and ideological implications that remain too little known. This essay explores the historical setting, ideological underpinnings, and the implications of associating religion with nuclear weapons power, particularly in the South Asian and Middle East contexts.

Historical Background

India's journey into the nuclear arena began in earnest shortly after British exodus from the Indian subcontinent in 1947. The architect of India's nuclear program, Homi J. Bhabha, envisioned nuclear energy as a tool for national development in a newly post-colonial India. The Atomic Energy Commission was established in 1948 under Prime Minister Jawaharlal Nehru's leadership, reflecting a commitment to peaceful uses of atomic energy. However, the geopolitical landscape – marked by the 1962 Sino-Indian War[1] and then China's nuclear test in 1964 – shifted India's priorities toward a potential weapons capability.

The country's push for such a potential weapons capability culminated in India's first nuclear test on May 18, 1974, under the codename "*Smiling Buddha*." Though *Smiling Buddha* was termed a "peaceful nuclear explosion," it was a nuclear explosion nonetheless, demonstrating India's entry into the nuclear weapons "club;" it accordingly triggered international reactions, including sanctions[2] and a deepening of non-proliferation concerns. India's explosion of that nuclear device occurred under Prime Minister Indira Gandhi's government, which maintained a "non-aligned" ideological outlook during the Cold War, making the Indian test a signal that nuclear weapons capabilities were emerging *outside* the capitalist-versus-communist rivalry. For over two decades thereafter, however, India refrained from further tests, adhering to a policy of nuclear ambiguity amid global non-proliferation pressures.

India's nuclear journey began under scientists like Homi Bhabha, driven by a vision of scientific self-reliance and independent technological success in the post-colonial context rather than religious zeal.[3] Describing the 1974 *Smiling Buddha* test as a "peaceful nuclear explosion" reflected India's ambivalence toward nuclear weaponization, rooted in Nehruvian ideals of non-alignment and moral opposition to nuclear hegemony of the U.S. and Soviet nuclear superpowers.[4] This ambivalence came to an end in May 1998, however, when overtly weapons-related tests were carried out under the Bhartiya Janata Party (BJP)-led government, which was strongly associated with Hindu nationalism. This prompted speculation about possible religious motivations beyond weaponization, especially given the BJP's ideological ties to the extremist Rastriya Swayamsevak Sangh (RSS) organization.[5]

The "Hindu bomb" label gained traction in part due to statements from BJP leaders, such as L.K. Advani – the country's Home Minister and Deputy Prime Minister of India in 1998 and then the second most powerful man in the government – who framed the tests as an assertion of national pride and strength, resonating with Hindu cultural imagery of power and sovereignty.[6] Critics, including some Indian communists, furthered this narrative by linking the tests

to Hindu nationalist agendas. For instance, Vinod Rai of the Communist Party of India (Marxist-Leninist) organization[7] called the bomb a "Hindu bomb," alleging that it symbolized anti-Muslim sentiment tied to broader BJP policies. Such critiques, however, often overlooked the strategic imperatives that also seemed to drive the tests, such as countering Pakistan's nuclear advancements and China's regional dominance, and which Indian leaders invoked at the time.[8]

The 1998 Nuclear Tests and the Rise of Hindu Nationalism

Yet Indian politics and government discourse of that period and thereafter have suggested that India's nuclear weapons program may still have some important religious undertones. The BJP is a political party rooted in the *Hindutva* ideology of Hindu nationalism, which had assumed power in a coalition government in March 1998. Shortly thereafter, on May 11 and 13, 1998, India conducted its series of five nuclear weapons tests at the Pokhran testing facility site some 70 miles from the Pakistan border, in the State of Rajasthan, under the operational code-name *Shakti*. The Indian government, in fact, said that the second of these two days of tests had involved detonation of what it a claimed was a thermonuclear device (a.k.a. "H-bomb"). (This occurred at the same site where India had carried its first nuclear test in May 1974.)

With these new tests, Prime Minister Atal Bihari Vajpayee (1924-2018) declared India to be a "full-fledged nuclear state," citing security threats stemming from both Pakistan and China. These tests met with jubilation domestically, being seen as a symbol of national strength and technological achievement.

The BJP's ascent marked a step forward from the alleged "secular ethos" of previous governments, especially those of the Congress Party. *Hindutva*,[9] an ideology long championed by the RSS and its numerous affiliates, seeks to define India officially as a Hindu nation. For proponents within this movement, the nuclear tests were not merely a strategic necessity, but a fulfillment of cultural destiny wrapped around Hinduism. Some nationalist voices thus celebrated the bomb as a "Hindu Bomb," a term coined in media and political

rhetoric to signify the triumph of a Hindu-majority nation over the past colonial and foreign dominations. A 1998 article in *The Indian Express* by journalist Chidanand Rajghatta entitled "The Hindu Bomb,"[10] for instance, encapsulated this sentiment, framing the tests as a moment of Hindu pride.

The "Hindu Bomb" Narrative: Ideology and Symbolism

> Hindu militarism is a genuine and powerful force, influencing Indian foreign policy. It is all the more dangerous because it is unanalyzed, unexposed, and insidious. No one is likely to understand the actions of the Hindu government of India in the international sphere during the last fifteen years without recognizing the existence of a strong under-current of militarism among the people of the country.
>
> Nirad C. Chaudhuri (1965)[11]

The association of nuclear weapons with Hindu identity draws from a complex interplay of history, mythology, and politics. Hindu nationalists often invoke ancient texts such as the *Mahabharata* and *Ramayana*, which describe fantastical weapons such as the *Brahmastra*, a projectile of immense destructive power, as evidence of India's historical scientific sophistication – even suggesting that modern India's acquisition of weapons of mass destruction (WMD) thus has ancient precedents and that India is today merely recovering its ancient glory. Following the numbers suggested in these ancient legends, in fact, the number of people said to have died in the Mahabharata War is put by Yudhishthira, truth striving hero of the epic, at "1 billion, 660 million, and 20,000 men" (1,660,020,000), with the survivors numbering no more than 24,165. (This kill-count is a specific detail given in the epic text itself.) It is important to remember that the *Mahabharata* is an ideological and mythological text and not an actual historical record, of course, but its mythologized account of ancient Indian proficiency in WMD warfare resonates with modern political narratives of a proudly Hindu country reclaiming a gloriously

imagined past after centuries of disruption by Muslim and British colonial rulers.

India's 1998 nuclear tests occurred against the backdrop of heightened communal tensions, notably the 1992 demolition of the Babri Masjid (Mosque) by Hindu activists, an event that bolstered the BJP's rise – including that of a Gujarati politician named Narendra Modi, who is now the Prime Minister. Critics, including leftist and secular intellectuals, accused the BJP of linking the country's nuclear weapons program to *Hindutva* in order to consolidate domestic support.

India's embrace of nuclear weaponization also fed into longstanding debates about Hindu nationalism in Indian society. As early as 1965, for instance, Nirad C. Chaudhuri, had offered a critical analysis of Indian society, challenging the then-prevalent notion (being promoted then by the Congress Party government) that India was in some sense inherently pacifist. He argued that militarism has been a consistent aspect of Indian (Hindu) civilization, with violence and warfare being deeply embedded in India's cultural and historical fabric and integral to its history and cultural identity, and he purported to trace this trait from ancient times through various dynasties. Chaudhuri cited examples such as Emperor Ashoka's conquest of Kalinga[12] and the military exploits of the Gupta kingdom, suggesting that these instances reflect a broader pattern of militaristic behavior in Indian society. His arguments were further supported by literary evidence from epics such as the *Mahabharata* and *Ramayana*, which, as noted above, depict large-scale wars and valorize martial prowess.

In more recent years, especially with the rise of the BJP, discussions of Hindu nationalism and its influence on India's foreign policy have gained prominence. Some scholars argue that Hindu nationalist ideologies have introduced a more assertive and security-oriented approach to foreign policy, emphasizing India's civilizational identity and strategic autonomy. However, others contend that while Hindu nationalist rhetoric is present, India's foreign policy remains largely pragmatic, driven by geopolitical and economic considerations

rather than ideological imperatives. Nonetheless, Chaudhuri's insights from his 1965 book *The Continent of Circe* provide a historical context for understanding these contemporary debates, highlighting the enduring complexities of militarism and ideology in the shaping of India's national and international identity – complexities that would seem to have become all the more important now that the country openly possesses the most powerful form of weapon in existence.

As noted, India's official stance upon its nuclear weapons tests in 1998 emphasized deterrence, not ideology. Nevertheless, internationally, the "Hindu Bomb" label raised concerns about religious extremism influencing nuclear policy. Pakistan, which conducted its own nuclear weapons tests weeks later in May 1998, dubbed its arsenal the "Islamic Bomb," thereby intensifying the perception of a religiously infused nuclear rivalry in South Asia. Western analysts worried that intertwining nuclear power with religious nationalism could destabilize the region.

Domestic and Global Reactions

Domestically, the 1998 nuclear tests enjoyed broad support across political lines, transcending the BJP's government's political base. Scientists such as A.P.J. Abdul Kalam, who later became India's President, were hailed as national heroes, underscoring the program's technical rather than religious roots. However, dissenters like Arundhati Roy, in her essay "The End of Imagination," condemned the tests as a descent into militaristic chauvinism, arguing that they betrayed alleged "India's Gandhian legacy of nonviolence." Roy's critique implicitly challenged the "Hindu Bomb" narrative by framing acquisition of the weapon as an ethical *failure*, not a religious triumph.

Globally, the 1998 nuclear tests created shock, and the response to the 1998 tests was overwhelmingly negative. The United States imposed economic sanctions, and the United Nations Security Council condemned the tests. The "Hindu Bomb" label fueled fears of proliferation driven by identity politics, though India maintained its no-first-use nuclear declaratory policy and (notwithstanding its open development of nuclear weapons) its professed commitment to global

nuclear disarmament – a stance at odds with the aggressive rhetoric of some nationalists.[13]

Upon India changing its nuclear posture from ambiguity (1974) to clarity (1998), despite two decades of changed geopolitics, the international community reacted strongly. Even at the point of the 1974 explosion, India's actions had generated shock and alarm. In the words of Raja Ramanna, the architect of that test,

> … [the 1974 explosion] came as a surprise to the world. They hadn't expected such an achievement from a developing country … their criterion for measuring success was different in the sense that they judged the success of a country by its material acquisitions and its overt proof of development.… India didn't conform to any of these, and in this context alone it seemed somewhat relevant when the Western world expressed bewilderment, coupled with fear and panic at the success of Pokhran [in 1974].[14]

Some such surprise occurred again 1998, but this time without even the pretense of the tests having been "peaceful" nuclear explosions.[15] India had not openly joined the nuclear weapons world. The international community reacted with sanctions, diplomatic backlash, and strategic recalculations.

India cited multiple justifications for its nuclear tests and bomb program.[16] They included:

- *Threat from China:* Long-standing border disputes and China's nuclear arsenal, compounded by Beijing's 1962 victory over India and its support for India's regional arch-rival Pakistan.

- *Pakistan Conflict*: Ongoing hostility and past wars with Pakistan, including tensions over the Kashmir region since 1947.

- *Kashmir Tensions:* Allegations of terrorism support by Pakistan and persistent regional instability.

- *Regional Prestige:* A show of nuclear strength to command respect from neighboring smaller countries of South Asia.

- *Demonstrated Capability:* Proving India's ability to weaponize its nuclear program through underground tests irrespective of international concerns.

- *National Reassurance:* Boosting public confidence in national security amidst regional nuclear uncertainty.

- *National Pride:* Enhancing a sense of prestige and identity tied to nuclear power.

- *Nuclear Club Aspirations:* Gaining entrance into the elite group of nuclear-armed states.

- *U.N. Security Council Membership Ambition:* Strengthening India's case for permanent membership on the U.N. Security Council.

- *Moral Disarmament Position:* Framing its acquisition of nuclear weapons as a "protest against nuclear apartheid" – that is, against the division of the world into nuclear weapons "haves" and "have not" countries – with India positioning itself morally through a so-called "Nuclear Satyagraha" for eventual global disarmament. Satyagraha refers to applying Mahatma Gandhi's principle of Satyagraha, or nonviolent resistance, to the issue of nuclear weapons. It embodies the belief that

> nonviolence can be a powerful force against nuclear weapons and the threat they pose to humanity.[17]

Senior U.S. officials rejected India's rationales for developing nuclear weapons without offering alternative explanations for its actions. Senator Jesse Helms (R-North Carolina), for example, suggested that India's nuclear actions posed a threat to the United States, while *Insight* magazine's James P. Lucier attributed India's motives to internal ethnic issues and warned of potential tragedy. Some blamed the ruling BJP, with *Time* magazine caricaturing Prime Minister Vajpayee as a "Nuclear Yogi." Prime Minister Vajpayee, however, defended India's nuclear policy in Indian Parliament, citing cultural obligations and quoting from the Hindu spiritual classic, the *Bhagavad-Gita*[18] – though this ironically lent credence to critics who had argued that Hindu nationalist motivations lay behind the country's nuclear weaponization.

As noted, India's 1998 nuclear tests prompted Pakistan to conduct its own tests on May 28, thus escalating the South Asian crisis and creating a new nuclear arms race. *The Bulletin of the Atomic Scientists* set the "Doomsday Clock" to nine minutes before midnight, marking the 16th adjustment since 1947 due to heightened nuclear risks.

Ideology and India's Nuclear Program

> Surprisingly, despite ruling India for two centuries, the British never truly understood – let alone connected with – the Hindu mindset.
>
> Nirad C. Chaudhuri (1965)[19]

The "Hindu Bomb" narrative intertwines strategic imperatives with cultural symbolism. As described earlier, Hindu nationalists often cite ancient texts like the *Mahabharata*, which as we have seen describes the *Brahmastra* – a mythical weapon of mass destruction – as evidence of India's historical scientific legacy. For instance, RSS ideologue M.S. Golwalkar (1906 – 1973) claimed in 1966 published

book *Bunch of Thoughts* that ancient India possessed advanced knowledge suppressed by foreign invaders.

While such claims are historically unverifiable, they resonate with a narrative of reclaiming a lost golden age, a theme echoed in BJP rhetoric from the late 1990s – a narrative that seemed to be reinforced by the tests in 1998. The BJP had in fact actually campaigned on a pro-nuclear weapons platform, promising in its 1998 election manifesto to "exercise the option to induct nuclear weapons,"[20] an agenda that aligned with its muscular nationalism. Senior BJP leader L.K. Advani linked the tests to national pride, stating of the tests' government divinely-derived code-name that "Shakti symbolizes India's strength rooted in its civilization."[21] Critics, however, saw such rhetoric as the BJP's politicization of a nuclear program that had in fact been built by the more secular governments of the past and created by scientists of diverse religious background, including A.P.J. Abdul Kalam (1931-2015), a secular Muslim who had overseen the 1998 tests as chief of India's Defense Research and Development Office (DRDO).[22]

Broader Implications and Legacy

The "Hindu Bomb" concept and its apparent embrace by at least some segments of the Indian political community pose enduring questions about identity and power. Strategically, the 1998 tests bolstered India's deterrence against Pakistan and China, and set in motion an Indian effort to develop explicit nuclear deterrence policies, as evidenced by its formalization in 2003 of a nuclear doctrine emphasizing credible minimum deterrence, as articulated by the Cabinet Committee on Security in January 2003. Yet the 1998 tests also presented a challenge for the global nonproliferation regime and its flagship instrument, the Treaty on the Non-Proliferation of Nuclear Weapons (NPT)[23] – which India had not signed – prompting debates about what degree of nuclear legitimacy to permit non-signatories and whether or not nonproliferation requirements should somehow be extended to such states.

The "Hindu Bomb" concept serves as a lens for examining how cultural narratives may shape technological milestones. This is a

phenomenon not unique to India, for Pakistan's "Islamic Bomb" and Israel's undeclared arsenal – as well as Iran's potential for a Shia-flavored Islamic bomb – offer interesting parallels. Observers still debate the motivations behind India's nuclear weaponization. Scholars such as George Perkovich argued that India's program reflected a "defiant nationalism" rather than religious zeal.[24] It would be a mistake, however, to dismiss the importance of Hindu themes in India's nuclear trajectory.

Western countries, I submit, have on the whole failed to understand India's true motives. Their recurring surprise and shock and India's nuclear steps suggest a deeper gap in understanding India – perhaps, in particular, reflecting too limited a grasp of Hinduism, which shapes India's national psyche. The following section explores the ideological forces that appear to lie behind India's pursuit of nuclear weapons, particularly the idea of a "Hindu bomb."

Evolution of Modern Hinduism

> The Hindu dreams that he will eventually be able to hoist the West with its own petard, and he is not such a fool as many might imagine.
>
> Nirad C. Chaudhuri (1965)[25]

The evolution of modern Hinduism was significantly shaped by British colonial rule, Christian missionary efforts, and Western education. Colonial policies – especially those led by figures like Thomas Macaulay – displaced traditional Hindu and Islamic education systems, introducing English and Western sciences as replacements. According to Lord Macaulay,

> We must at present do our best to form a class who may be interpreters between us and the millions whom we govern ... a class of persons Indian in blood and colour, but English in taste, in opinions, in morals and in intellect.[26]

This effort may not have been successful in producing the *de facto* Englishmen that Macaulay intended, but it did lead to greater

nationwide communication among Indians, particularly high-caste Hindus – and this led, in turn (if ironically, given the aims of Macaulay's project) to a resurgence of Hindu cultural values.

The translation and much wider spread and availability of Hindu texts during the British colonial period, notably the *Bhagavad Gita*, played a pivotal role in this revival. Though virtually unknown in India before its 1785 English translation, the "Gita" gained prominence abroad and then eventually "returned" to India to inspire a new wave of modern Hindu thought and nationalism.[27] The resulting fusion of Western science and Hindu ideology gave rise to reinterpretations to such as "Hindu Physics" and "Hindu Chemistry," promoting the idea of Hinduism as a rational, scientific faith. This ideological shift also alienated the Muslim minority, contributing to rising tensions and eventually helping lead to the creation of Pakistan after the British withdrew and the country split. It eventually also thus contributing to the pursuit by Pakistan of an *Islamic* bomb in response to what became known as the Hindu bomb.

Age of The Atom Bomb

> Modern Hindus had not forgotten their *Mahabharata*. They presented Colonel Rusk who commanded the task force of the 32[2nd] American Air Division in India with a model of the [ancient] battle of Kurukshetra![28]
>
> Nirad C. Chaudhuri (1965)[29]

Hindu concepts in post-independence India seem to have continued to fuel nationalist pride and ideological revival in ways linked to scientific achievements such as nuclear weapons development. In fact, even Mahatma Gandhi (1869-1948) – something of a prophet for modern Hinduism but a notable practitioner of nonviolence (*ahimsa*) within the Indian independence movement, who was ultimately assassinated by a Hindu nationalist zealot – gave a speech in 1947 in which he proclaimed that

> … today they [Sikhs] are thinking of the sword. They do not realize that the age of the sword is past. They do not realize that no one can be saved by the strength of the sword. *This is the age of the atom bomb*.[30]

(Later, Gandhi also asserted that in that "age of the atom bomb, "the sword was a rusty weapon.")[31]

Gandhi's seeming conclusion that, in contrast to the "sword," nuclear weaponry had some potential to "save" the country was not lost on later Hindu nationalists. Decades later, with nuclear weapons stockpiled, Indian leaders – particularly the staunch nuclear hawks – did not hesitate to invoke Gandhi's name to justify their stance. Drawing strength from Gandhi's writings, they also cited his somewhat paradoxical essay titled "The Doctrine of the Sword," penned in the pre-atomic era of the 1920s, to support the development of more atomic bombs.

This work is worth examining in more detail. Gandhi's lengthy article "The Doctrine of the Sword" was published on August 11, 1920, in the newspaper *Young India*. A relevant portion of that essay says:

> I do believe that where there is only a choice between cowardice and violence, I would advise violence. Thus, when my eldest son asked me what he should have done had he been present when I was almost fatally assaulted in 1908 – whether he should have run away and let me be killed, or used his physical strength to defend me – I told him it was his duty to defend me, even by using violence.
>
> Hence, I took part in the Boer War, the so-called Zulu Rebellion, and the Great War. I also advocate arms training for those who believe in violence. I would rather have India take up arms to defend her honor than remain a helpless witness to her own dishonor in cowardice. Let me not be misunderstood. Strength does not come from physical capacity; it comes from an indomitable will. The average Zulu, in terms of bodily strength, may be superior

> to the average Englishman, but he still fears the English boy's revolver. He fears death, and that fear renders him powerless, despite his strong physique.
>
> It is better to use brute force than to betray cowardice. It is better for India to arm itself and take the risk than to avoid arms out of fear. That was why I joined the Boer War and aided the government during the Zulu Rebellion. During the last World War, I also supported the British, both in England and India, including recruitment efforts. Forgiveness is the virtue of the brave. Only the strong can truly forgive. Likewise, only one who is capable of enjoying pleasures can qualify to be a *brahmachari*[32] by restraining desires. There is no such thing as the mouse forgiving the cat. India's soul-force will be proven only when it refuses to fight despite having the strength to do so. This "strength to fight" does not mean physical might alone. Anyone with courage and who has overcome the fear of death possesses such strength.[33]

Today when we judge Gandhi's above comments from 1920, his alleged pacifism might come across curious and problematic. Whether or not these comments were actually intended somehow to invoke moral strength against British force or in fact to endorse violence and power, this passage has proven very helpful to modern Hindu nationalists seeking to build up the country's military might and seek virtue in such power.

Preparing for the Age of the Hindu Bomb

> Modern defense as well as modern industry require scientific research both on a broad scale and in highly specialized ways. If India has not got highly qualified scientists and up-to-date scientific institutions in large numbers, it must remain a weak country incapable of playing a primary part in a war.
>
> Jawaharlal Nehru (1956)[34]

For all his famous devotion to nonviolence, Gandhi seems to have had a degree of ambivalence about nuclear weapons. At a prayer speech on June 16, 1947, for example, he observed that: "If we had the atom bomb, we would have used it against the British."[35] This was not an unambiguous endorsement of nuclear weaponization, of course, for he did not exactly say that it was a bad thing that Indian nationalists had lacked the bomb, and by June 1947 India was clearly about to win independence without it. Nevertheless, India's post-independent leaders seem to have felt India needed such weaponry.

In 1942, Gandhi had named Jawaharlal Nehru as his successor, confident in his loyalty. And after British left the Indian subcontinent, Nehru, a high-caste Brahmin, would lead India into the nuclear era despite the nation's poverty and lack of infrastructure. Under the guise of peaceful development and Western-style progress, India quietly pursued nuclear ambitions, projecting a non-violent, democratic image to the world even as militant groups such as the RSS and associated modern Hindu[36] groups developed their thinking on the margins of the Indian political community and prepared to transform such ideas into a new national ideology.

Various legislative and institutional steps were taken that helped prepare India not only for nuclear power production but also for the "age of Hindu Bomb." In 1948, for instance, Prime Minister Nehru introduced the Atomic Energy Bill, seeking to ensure the secrecy of (and central government control over) nuclear technology, and limiting nuclear policymaking to a select few government officials without legislative oversight. It also contained severe penalties against those who would violate the law.

This law faced little opposition in the Constituent Assembly (India's provisional parliament at the time). During debates on this measure, a number of legislators from various portions of the political spectrum framed the issue in terms suggesting that they associated nuclear questions with Hindu religious themes.

H.V. Kamath framed atomic energy policy through a Hindu nationalist lens, for example, linking it to ancient Hindu wisdom, while Nehru himself emphasized a global "world-time" perspective with clear Hindu resonances, urging India to harness atomic power to avoid historical backwardness. S.V. Krishnamurthy Rao questioned the restrictive controls on nuclear technology to ensure its use for peaceful purposes, comparing India's approach unfavorably to less restrictive laws in Britain and the United States. In responding to these complains, Nehru deflected, stressing timing and geopolitical opportunity, but thereby apparently also revealing the bill's underlying strategic motives, hinting at a "Hindu Bomb" agenda. The heated exchange between these two personalities is worth quoting, for it suggests the true nature of what was being debated in the parliament:

> **Rao:** May I know if secrecy is insisted upon even for research for peaceful purposes?
>
> **Nehru:** Not theoretical research. Secrecy comes in when you think in terms of the production or use of atomic energy. That is the central effort to produce atomic energy.
>
> **Rao:** In the Bill passed in the United Kingdom secrecy is restricted only for defense purposes.
>
> **Nehru:** I do not know how to distinguish between the two [that is, peaceful and defense purposes].[32]

Building a Nuclear Bureaucracy

From the start, India's atomic program had a military aspect. Nehru's stance reflected Modern Hinduism's blurring of lines between violence and nonviolence. Despite some debate, the bill passed with support for its military intent. The industrial revolution in India at its most disinterested is an expression of anti-European and anti-Western nationalism. It is the realization of the desire, and now the policy, of the Hindus to get even with the

> West and take revenge for the dead European imperialism by adopting its technology and organization.
>
> Nirad C. Chaudhuri (1965)[37]

On August 15, 1948, the Atomic Energy Commission (AEC) of India was established under the Atomic Energy Act of 1948. It was led by three prominent scientists – H.J. Bhabha, K.S. Krishnan, and S.S. Bhatnagar – who had also been appointed a month earlier to the Scientific Advisory Committee of the Ministry of Defense.

Dr. Homi Jehangir Bhabha, a nuclear physicist trained in Europe, became the AEC chairman. Before Indian independence in 1947, he was already heading key scientific institutions such as the Tata Institute of Fundamental Research (TIFR) and bodies within the Council of Scientific and Industrial Research (CSIR). He had also secured international support from Canada, France, and the United Kingdom for nuclear energy collaboration. With Prime Minister Nehru's backing, Dr. Bhabha helped push the Atomic Energy Act through Parliament with minimal scrutiny, and India began building a nuclear infrastructure.

By the late 1950s and early 1960s, India had developed a substantial and increasingly complex nuclear infrastructure, exemplified by the Trombay Atomic Reactor Center. This facility operated through six specialized divisions covering education and training, uranium production and plutonium extraction, scientific research, engineering, biological and medical research on radiation effects, and atomic minerals exploration.

In 1958, the Atomic Energy Commission (AEC) was restructured, increasing its membership from three to seven, with Bhabha continuing as chairman. To further strengthen state control, the Atomic Energy Act of 1962 replaced the earlier 1948 Act. The new law granted sweeping powers to the government, even allowing it to override any other national legislation (Clause 28) that could hinder

atomic energy activities. This marked a significant escalation in the centralization and secrecy surrounding India's nuclear program.

In its public pronouncements, the Indian government constantly proclaimed that the entire nuclear energy program existed exclusively for peaceful purposes, and that the atomic energy it would produce would provide an incredibly cheap and abundant source for electricity and for other forms of economic and industrial development. Over time, however, the government slowly changed its messaging by introducing themes identifying India's neighbors as enemies thus gradually altering the context in which comments about "peaceful" nuclear energy were made.

In the early years, China was portrayed a friend, but from the late 1950s it came to be depicted as a great enemy of India, especially after Sino-Indian War of 1962. In time, not just India's initial rival Pakistan, but also China, Nepal, Myanmar (Burma), Bangladesh, and Sri Lanka were all portrayed as potential threats. Even as they helped India develop its nuclear infrastructure, the United States and other Western countries were also regarded with deep suspicion. Amid growing perceptions of "enemies all around," nuclear non-proliferation must have seemed less and less attractive.

And indeed, even though India's scientific and technological efforts had long been justified as aiming for post-colonial "self-reliance," critics of India's nuclear power program such as Dhirendra Sharma have long argued that the country's nuclear industry is bloated and corrupt – and nuclear power generation remains "an unfulfilled promise in India." This suggests that other motives may have been present all along, and indeed declassified documents now reveal that at least as early as 1968, Western visitors to the Trombay facility were "unsettled" by "data suggesting that India was heading toward the 'development of a nuclear device.'"

The Rise of Ethno-Religious "Bombs"

India's nuclear weapons program does not exist in a vacuum, but rather is but one example of a disturbing trend in nuclear

proliferation to states whose possession of such weapons is taken to represent their acquisition on behalf of and for the purposes of advancing a specific ethnic or religious group. Alongside the "Hindu Bomb," in other words, there is also perceived to be an "Islamic Bomb" and a "Jewish Bomb."

India's nuclear development is often seen as a strategic response to China – which tested its first nuclear weapon in 1964 – and other regional threats. In my view, however, another powerful force in its development are ideological imperatives of Indian greatness tied to and fueled by modern Hindu nationalism. These ideological resonances have alarmed the Islamic world – especially Pakistan – triggering its own nuclear response. These ongoing dynamics risk wider proliferation, especially in the Middle East, where there is already said to exist an Israeli nuclear arsenal, and where both the sectarian religious state of Shi'ite Iran and the Sunni kingdom of Saudi Arabia are also envisioned as potential future proliferators. Meanwhile, in Russia, the Putin government has presided over increasing connections between the Russian Orthodox Church and Russia's own nuclear weapons establishment.

Competitive ethno-religious identity politics between such nuclear-armed countries or groupings risks spurring both further proliferation and escalation to a nuclear conflict. As Nigel Calder put it back in 1979, "even the early phase of the nuclear epidemic is dangerous, and the Israeli and Pakistani bombs could be the death of us."[38]

Emergence of the Pakistani Nuclear Program

> If India builds the bomb, we will eat grass or leaves, even go hungry. But we will get one of our own. We have no alternatives.
>
> Zulfikar Ali Bhutto (1965)[39]

While India portrayed its nuclear program as peaceful, Pakistan tended to see it as a threat, its views shaped by centuries of Islamic rule, Hindu-Muslim conflict, the trauma of Partition, and multiple

wars with India. In particular, the 1971 war, leading to Bangladesh's independence,[40] marked a major blow to Pakistan and what felt like a symbolic victory for modern Hinduism over Islam. Driven by deep historical memory and religious rivalry, Pakistan suspected India's nuclear ambitions early on, and viewed them as a serious threat.

> We know that Israel and South Africa have full nuclear capability. The Christians, Jewish and Hindu civilizations have this capability. The communist powers also possess it. Only the Islamic civilization was without it, but that position was about to change.
>
> Zulfikar Ali Bhutto (1977)[41]

> It is our right to obtain this [nuclear] technology. And when we acquire this technology, the entire Islamic world will possess it with us.
>
> General Zia ul-Haq (1986)[42]

The United States shared nuclear power-generation knowledge with both India and Pakistan under the "Atoms for Peace" program, and Pakistan accelerated its efforts to develop a nuclear technology base after its 1971 defeat. In 1972, then-Prime Minister Zulfikar Ali Bhutto (1928-1979) launched Pakistan's nuclear weapons program, later supported by Islamic nations and other foreign suppliers. Dr. Abdul Qadeer Khan, using insider access in Europe, stole European enrichment technology and thereafter smuggled key centrifuge designs to Pakistan (and thereafter other countries as well, along with Chinese nuclear weapons designs). By the late 1980s, with China's help, Pakistan had a complete nuclear arsenal. As a result of U.S. pressure, Pakistan refrained from testing until India's 1998 nuclear tests forced its hand – at which point it, too, openly weaponized.

The nuclear weaponization of both India and Pakistan from 1998 created a dangerous arms race that has alarmed international onlookers ever since. As William E. Burrows and Robert Windrem had put it in 1994, for instance, even before both countries had openly tested weapons,

> … [t]he Indian subcontinent is [already] the most dangerous place on Earth. It is the incubator of racial and religious hatred that is more virulent and persistent than any biological epidemic (though it, too, could be unleashed in a war). The slum of every city of consequence is a purgatory in which rampaging Hindu and Muslim fundamentalists search for their opposite number and kill them. The race for superweapons is therefore driven as much by sheer hatred as by geopolitical considerations. While both sides have invented elaborate excuses for developing nuclear weapons – strategic deterrence, for example – their real purpose is genocide.[43]

Such concerned heightened further once both India and Pakistan began their arms race.

Zionism and Israel's Nuclear Program

The idea of Israel emerged in the late 19th century under Theodor Herzl as a response to rising antisemitism in Europe, advocating for a Jewish homeland, eventually centered on Palestine. Its core ideology emphasized Jewish nationhood and the necessity of a sovereign state for survival. Growing nationalism and events like the Balfour Declaration (1917) strengthened the movement. The development of Israel involved early Jewish migration, institution-building, and then considerable international support following the Holocaust, which seemed to illustrate that the Jews had no safe alternative but to create a distinct, separate national home. Israel declared independence in 1948, leading to immediate conflict with its Arab neighbors and long-term issues associated with the occupation of formerly Palestinian-occupied land after the 1967 war.

Zionist security ideology, shaped by Holocaust trauma and regional hostilities, focused on survival, perceived existential threats, and the doctrine of self-reliance, prompting Israel to seek military and nuclear superiority to ensure its existence. Israel, though never officially confirming its possession of a nuclear arsenal, is widely

believed to possess nuclear weapons. And Zionism, as both a political and religious ideology, played a key role in shaping Israeli security doctrine. Israel's nuclear strategy, including the policy of strategic ambiguity (*i.e.*, not confirming it has nuclear weapons but benefiting from widespread assumptions to this effect),[44] is influenced by a belief in existential threat from its neighbors – all of which are predominantly Muslim, and which have repeatedly attacked it in the past – and by a theological commitment to the survival and protection of the Jewish people in what they view as having been their God-given homeland since Biblical times.

In the 1950s, Prime Minister David Ben-Gurion secretly authorized efforts to develop Israel's nuclear capabilities. His motto was "Never again will we be helpless." France, after the Suez Crisis of 1956, collaborated with Israel, helping construct the Dimona nuclear facility in the Negev desert that is today widely believed to be the center of Israel's secret weapons program. (France provided the nuclear reactor, heavy water, and technological expertise.)

As noted, pursuant to its policy of nuclear ambiguity ("Amimut"),[45] Israel neither confirms nor denies having nuclear weapons. This strategy allows it to benefit from a degree of nuclear deterrence while minimizing the risk of formal retaliation, international sanctions, or reciprocal weaponization by a hostile Muslim neighbor. Possessing nuclear arsenal also provides a "last resort" retaliatory capability in case of national catastrophe.

By the late 1960s, Israel is believed to have acquired its first operational nuclear weapons.[46] (During the Yom Kippur War of 1973, Israel's nuclear preparedness reportedly influenced the U.S. decision to resupply Israel with conventional arms against its Arab adversaries.) The size of Israel's nuclear arsenal is the subject of much speculation, with guesses about its total number of warheads ranging from 60 to "over 400."[47] Its delivery systems include aircraft, land-based missiles, and cruise-missile submarines (which offer some degree of survivable second-strike capability in the event of a catastrophic attack upon Israel itself).

Despite its deliberate opacity, Israeli nuclear capabilities have contributed to regional arms races, such as the Iraqi and Syrian pursuit of a nuclear capability, resulting in the Iraqi reactor project at Osirak being destroyed by Israel in an aerial attack in 1981 and the Syrian reactor at Dair Alzour being similarly bombed by the Israelis in 2007, as well as ongoing concerns over Iran's nuclear program. This opacity has also led to criticism from Arab states, who view Israel's exemption from the Treaty on the Non-Proliferation Non-Proliferation of Nuclear Weapons (NPT) – a treaty that, like India and Pakistan, it never signed – as a clear double standard. Calls for the establishment of a nuclear weapons-free "Zone" in the Middle East have become loud, particularly from Arab states who perceive the current *status quo* as unsustainable.

Israel's nuclear posture remains a highly sensitive issue in diplomatic discussions involving the United Sates, European nations, and the United Nations. Israel's nuclear capability, however, is embedded within its strategic partnership with the United States, rooted in mutual security interests. For Washington, Israel's nuclear deterrent is often viewed as a stabilizing force that supports U.S. hegemony in the Middle East.

Shi'ite Ideology and Iran's Nuclear Program

Iran's nuclear ambitions are shaped by a complex interplay of Shia Islamic ideology, nationalism, and strategic considerations. Since the 1979 Islamic Revolution, Iran has been governed by a theocratic regime that integrates religious authority into all aspects of governance. It has been reported in some circles that supreme leader Ayatollah Ali Khamenei has issued a *fatwa* (authoritative pronouncement under Islamic law) declaring nuclear weapons un-Islamic, yet suspicions persist internationally both regarding the existing of such a *fatwa* and more generally regarding Iran's ultimate intentions. (One recent account from Iran, in fact, suggests that the purported fatwa only bars the *deployment and use* of nuclear weapons, but would permit their production!) Iran's compliance with its NPT and nuclear safeguards obligations has been poor for many years, and the International Atomic Energy Agency has documented many

aspects of Iran's nuclear work that have appear to be related to weaponization.[48] The tension between Iran's reported religious prohibition and its pursuit of nuclear technology reflects deeper ideological divisions within the regime and illustrates how religious doctrine can both constrain and justify nuclear development.

Iran's nuclear power program started under Shah Mohammad Reza Pahlavi, with U.S. support as part of the Atoms for Peace Program. The Shah's aim was energy diversification and prestige; he wanted up to 20 nuclear reactors and even hinted at possible weaponization. After the 1979 Islamic Revolution, Ayatollah Khomeini initially halted the program, reportedly seeing nuclear power as unnecessary and nuclear weapons as un-Islamic and extravagant[49]. But the long and bloody Iran-Iraq War of 1980–88 changed Iranian perceptions: chemical weapons attacks by Iraq and regional insecurity reignited Iranian interest in nuclear capabilities as a way to help preserve Iran's Islamic Revolution.

Shia political thought strongly emphasizes resisting injustice and protecting the rights of the *Mazlum* (oppressed).[50] This is rooted in their belief that the only legitimate government is one that follows God's righteous will, advocating for social justice and equality, as well as by a long and painful Shi'ite history as an oppressed minority even within Islam. The Battle of Karbala in the year 860 – at which Husayn ibn Ali, grandson of the Prophet Mohammed and head of the Shi'ite community, was slain by Sunni Muslim rivals of the Umayyad caliphate – is seen as a key example of standing up against injustice, even at the cost of martyrdom. Shia political thought sees resistance against unjust rulers and systems as a moral and spiritual duty, drawing from the Quran and teachings of the Prophet Muhammad and his descendants.

In the modern era, Iran's lack of advanced technology (including nuclear technology) had been seen as a symbol of Western domination, and the development of nuclear capabilities in Iran as a symbol of standing up against oppression in ways that draw upon these Shi'ite traditions. Shia Islam encourages independence from foreign powers, aligning with Iran's push for indigenous nuclear technology. As

Shafat Yousuf and Syed Jaleel Hussain have noted, Iran frames its nuclear program as strictly peaceful, justified by Islamic law forbidding weapons of mass destruction but allowing defensive science. Some, however, doubt its sincerity in this respect, noting that in Shia tradition, the doctrine of *taqiyya* permits concealment of one's true intentions under threat – a concept that developed during the long years in which Shi'ites had to conceal themselves against Sunni Muslim oppression.

There are several strategic and theological themes upon which Iran may be drawing in justifying the pursuit of nuclear weapons:

- **"Science as worship"**: Some Islamic scholars in Iran argue that scientific advancement, including nuclear technology, is a form of *religious duty*.

- **Defensive Deterrence**: Building robust scientific and possibly latent nuclear capability is seen as deterrence against existential threats (*e.g.*, Israel and the United States).

- **Imam Mahdi's Return**: Some Shi'ite religious leaders believe that building a powerful Islamic society (including demonstrating Iran's technological mastery) is necessary for the eventual return of the Mahdi, a messianic figure in Shia eschatology who is felt to have become "occulted" after the Battle of Karbala. (It is also conceivable that Sh'ite traditions valorizing glorious martyrdom in standing up against injustice may make Iran more willing to contemplate nuclear escalation.)

Iran's nuclear program thus blends national security needs, religious justifications, and Shia political philosophy. Shia Islam provides both moral *restrictions* and *motivations* for Iran's scientific nuclear pursuits while encouraging it to resist Western hegemony and

promote national sovereignty as a theocracy representing the world's only Shi'ite government.[51]

Ideology and Saudi Arabia's Nuclear Program

Saudi Arabia's interest in nuclear technology began in the 1970s, primarily for peaceful purposes such as energy and desalination, given its rapidly growing population and water scarcity. However, regional tensions especially with Iran's nuclear advancements and Israel's undeclared nuclear arsenal have pushed Sunni-ruled Saudi Arabia to consider a more strategic dimension to its nuclear program. In 2015, Saudi Arabia launched a major initiative called the King Abdullah City for Atomic and Renewable Energy (KACARE) to formally push nuclear energy development. Riyadh signed multiple agreements with countries offering nuclear technology, but has not yet accepted the strict "Gold Standard"-style civil-nuclear cooperation agreement with the United States, which would restrict the kingdom's ability to produce its own fissile material through uranium enrichment or plutonium reprocessing.

> If Iran developed a nuclear bomb, Saudi Arabia would follow suit as soon as possible.
>
> Crown Prince Mohammed bin Salman (2018)

Saudi Arabia's nuclear ambitions are not just technical or strategic; they are deeply ideological:

- **National Security and Regional Balance.** The monarchy sees nuclear capability (even latent capability) as essential to maintaining a balance of power with Iran and Israel. Nuclear technology symbolizes modern sovereignty and strategic independence.

- **Preservation of Regime Stability.** In Saudi political ideology, maintaining the monarchy's survival is paramount.[52] A nuclear program is viewed as a

deterrent against both external threats and internal destabilization caused by regional conflicts.

- **Islamic Leadership.** As the Custodian of the Two Holy Mosques (at Mecca and Medina), Saudi Arabia's leadership feels responsible for defending Islamic lands. Some ideological narratives frame nuclear capability as necessary to protect Islam from external aggression, especially given Iran's Shia-majority regime and its counterpart Saudi Arabia's Sunni-Wahhabi orientation.

Religious Extremism and Non-State Actors

The most alarming intersection of religion and nuclear weapons arises from the potential acquisition of nuclear technology by extremist groups. Religious militancy, especially where groups interpret holy texts to justify mass destruction, poses a unique and urgent threat.

One obvious potential threat comes from terrorist groups such as Al-Qaeda, ISIS, and others who have openly expressed interest in acquiring weapons of mass destruction. Should they acquire nuclear weapons, their pre-existing religious justifications for mass violence could make them uniquely dangerous – not least since traditional deterrence models that have traditionally helped restrain *state* nuclear use may not apply to these terrorists, as such actors are not necessarily bothered by the prospect of mutual destruction and may indeed prize martyrdom.

Nor is it impossible to imagine that a nuclear weapon could be delivered by terrorist means. A small tactical nuclear device, for example, could perhaps be made portable enough to be secretly transported across borders and placed near strategic targets, or used to render areas uninhabitable.

Following an interview with CBS newsmagazine *Sixty Minutes* on September 7, 1997, late governor of Krasnoyarsk Krai and former Russian Security Council Secretary, General (Ret.) Alexander Lebed

claimed that the Russian military had lost track of 80 or so "suitcase"-sized atomic demolition munitions (ADMs). Lebed stated that these devices were made to look like suitcases, and that he had learned of their existence only a few years earlier. His account may indeed have been fanciful, and on September 10, the Ministry for Atomic Energy of the Russian Federation (MINATOM)[53] rejected Lebed's claims as baseless. Russian Prime Minister Viktor Chernomyrdin also ridiculed Lebed's account as "absolute stupidity" and said that "all Russian nuclear weapons are under the total and absolutely reliable control of the Russian armed forces." Most Western observers today think that no such loss of "suitcase nukes" in Russia actually occurred.

In another instance of possible terrorist nuclear threats, the U.S. Central Intelligence Agency (CIA) was told by an intelligence source code-named "Dragonfire" that al-Qaeda had smuggled a 10-kiloton Russian nuclear device into the United States, specifically targeting New York City. This alarming report caused considerable concern within the U.S. government, leading President George W. Bush to order Vice President Dick Cheney to leave Washington for an undisclosed location to ensure continuity of the presidency in the event such a device was detonated in Washington. Subsequent investigations found no concrete evidence to support the existence or presence of such a weapon, but the incident underscores the challenges intelligence agencies face in assessing threats based on human sources and the importance of corroborating information before acting – as well as the potential catastrophe that could result if indications of a *real* threat were overlooked.

Nevertheless, such small devices *are* surely possible. Even though their yield would likely be low, if detonated in a populated area, even a low-yield "suitcase"-scale nuclear weapon could cause catastrophic damage. A one-kiloton explosion, for example, could destroy structures within a half-mile radius and result in tens of thousands of casualties, depending on the population density. Additionally, the radioactive fallout would pose long-term health and environmental risks. An act of nuclear terrorism could thus rip the heart out of a major city, and cause ripple effects throughout the world, producing not just local damage but widespread fear elsewhere, flight

from major cities in a large-scale uncontrolled evacuation in response to any further terrorist threats (even false ones), and widespread havoc and economic chaos.

A bomb in Washington, D.C., for example, might kill the President, the Vice President, and many of the members of Congress and the Supreme Court. The explosion would also destroy much of the city's ability to respond. Hospitals would be leveled, doctors and nurses killed and wounded, and ambulances destroyed. (In the Japanese city of Hiroshima – attacked by the United Staes with an atomic bomb in August 1945 – 42 of 45 hospitals were destroyed or severely damaged, and 270 of 300 doctors were killed.) Resources that survived outside the zone of destruction would be utterly overwhelmed. (Hospitals have no ability to cope with tens or hundreds of thousands of terribly burned and injured people all at once; the United States, for example, has 1,760 burn beds in hospitals nationwide, of which only a third are available on any given day.)

Russian Nuclear Orthodoxy

The prospect of terrorist nuclear use has concerned Western national security planners for many years, but such worries may be all the more acute in connection with the possibility that religiously motivated elements within a nuclear weapons possessor state could facilitate proliferation. This risk is particularly relevant in countries where the military, intelligence, or nuclear establishment has strong ties to religious and ideological extremist groups or movements.

Connections between religious movements and nuclear weaponization can be seen even in the Russian Federation, one of the two "nuclear superpowers" left over from the Cold War and a country that currently possesses the world's largest nuclear arsenal. In a phenomenon that the Israeli scholar Dmitry Adamsky has termed "Russian Nuclear Orthodoxy," the Russian Orthodox Church (ROC) – once persecuted by the Soviet regime – has emerged as a co-guardian of national security in a close alliance with the Putin government, shaping the values, behavior, and identity of nuclear weapons-related personnel and institutions. The ROC has become deeply intertwined

with the Russian nuclear forces, influencing their symbols, practices, and even strategic thinking. This relationship, encouraged by the Putin regime, sees the Church not only legitimizing but also actively shaping Russia's assertive national security strategy, including its nuclear doctrine; it has been making ROC theology an increasingly important factor in Russian nuclear thinking.[54]

Analysis and Recommendations

The examples above reveal recurring themes in how religion can influence nuclear proliferation. First, religion can enhance perceptions of existential threat. Officials in states such as India, Israel, and Pakistan have invoked religious narratives of survival or martyrdom to justify their nuclear programs, reinforcing the idea that nuclear arms are essential to national and spiritual survival.

Second, religion can provide a moral and ideological justification for the possession or potential use of nuclear weapons. In Pakistan, the concept of defending the Muslim world gave nuclearization a pan-Islamic moral weight. In India, references to Hindu civilization and divine power bolstered domestic support for nuclear tests.

Third, religious ideologies can shape strategic culture in ways that may undermine logic of conventional deterrence. For example, apocalyptic or martyrdom-based belief systems, where death in defense of faith is valorized, complicate rational cost-benefit calculations assumed in deterrence theory and may dangerous special tolerance for escalation risks.

Fourth, however, religion may also sometimes constrain nuclear ambitions. If indeed it exists, Iran's reported *fatwa* against nuclear weapons illustrates how religious doctrine could perhaps act as a limit on military policy. However, such constraints are often contested or reinterpreted within the political-religious elite.

Finally, the symbolic role of nuclear weapons as embodiments of divine favor or civilizational prestige can contribute to a narrative of spiritual and national power and create incentives for proliferation. In

general, nuclear weapons can serve not only as tools of security policy (*i.e.*, deterrence) but also as markers of ideological and religious identity and affirmation in ways that may create risks and challenges not hitherto considered in international security planning.

To address the role of religion in nuclear proliferation and escalation management, policy responses must be sensitive to religious practices while also recognizing the importance of the institutions, values, and practices of the existing nuclear nonproliferation regime. The following are key recommendations:

1) **Promote Interfaith Dialogue and Confidence-Building Measures:** International organizations and regional fora should invest in interfaith dialogue initiatives that include discussions on peace, disarmament, and ethical responsibilities related to weapons of mass destruction. Faith leaders can serve as influential actors in tempering extreme narratives and advocating for restraint.

2) **Depoliticize Religion in National Security Discourses:** States should strive to separate religious rhetoric from strategic decision-making. This involves promoting secular policy frameworks, discouraging the use of religious language in defense policies, and ensuring that military doctrines are grounded in rational, non-theological terms.

3) **Engage Religious Authorities in Non-Proliferation:** Religious leaders and institutions can play a vital role in reinforcing global nuclear non-proliferation norms. Their moral authority can legitimize and amplify calls for restraint. The alleged *fatwa* by Iran's Supreme Leader prohibiting nuclear weapons, for example, may represent a religious-based commitment to non-proliferation. Such declarations should be publicly supported and promoted through

diplomatic channels to strengthen global norms against nuclear weapons development and use.

4) **Address Root Causes of Insecurity:** Often major religions consider specific geographical locations under their control as sacred spaces that may be felt to deserve protection at essentially any cost. In addition, religious movements may seek territorial expansion for religious reasons, at the expense of others. Some religious justifications for nuclear weapons may stem from non-religious insecurities, such as disputes over territory or resources, historical traumas, or identity crises. To treat religious ideology as merely epiphenomenal, however, would be to overlook its real, generative power in politics, conflict, and global affairs – including in nuclear policy. We must recognize that such ideologies can possess independent influence through their moral authority, emotional power, institutional autonomy, and historical impact. They can legitimize political actions, mobilize identity, shape public sentiment, and act independently of the state. Long-term peacebuilding efforts must address these root causes through comprehensive development, equitable conflict resolution, and justice mechanisms.

5) **Monitoring Religious Extremism in Nuclear States:** Intelligence cooperation, open-source collection, and analysis by area specialists and civil-society organizations should focus upon detecting and understanding the role of religious extremism within nuclear weapons bureaucracies and nuclear policy-making establishments. Close monitoring of nuclear states can help to identify and prevent radicalization within the scientific, political and military personnel who have access to sensitive technologies.

6) **Religious Education Reform:** Religious curricula should emphasize peace, coexistence, and ethical responsibility. Interpretations that justify violence should be morally scrutinized and responsibly challenged by both religious and non-religious scholars and varied institutions. Not all violent interpretations of holy texts are misinterpretations, and some may have historical and theological legitimacy. However, in today's context, if such interpretations go unchallenged, they can undermine peace, human rights, and nuclear nonproliferation objectives.

7) **Theological education, scrutiny, and its promotion:** To the extent that major mythology-based religions – such as those based upon the Bible, the Quran and Hadiths, and Hindu texts – have produced offshoots in the form of political movements that promote violence, including the use of weapons of mass destruction, they can pose a significant threat to the existing human-centered world order. As with other elements of terrorist radicalization and propaganda, efforts should be made to stop the spread of hate and counter such provocative messages. Today, multiple exchanges and dialogues are taking place through various media (radio, TV, YouTube, podcasts, etc.) to track and address the destructive nature of such religious ideologies. This work must continue in order to reduce the dangerous impetus that extremist religious movements create.

8) **National Security Studies:** U.S. military as well as Intelligence Community schools and research centers must adjust their curricula and scholarly research in order to help better identify, understand, and address the burgeoning threats that can stem from the intersection of religious movements and WMD policy.

9) **Military Force:** As a last resort, it may indeed become necessary to take direct action to neutralize the threat presented by nuclear-armed and religiously-motivated fanatics. Strikes against nuclear sites in "rogue states" are rarely advisable due to their high risks, including those of potential escalation into a broader war, violations of international law, undermining global norms, strengthening of hardline elements within the targeted state with the possible side-effect of actually encouraging more rapid proliferation, or even spreading radioactive contamination (*e.g.*, if an operating nuclear reactor is hit). Diplomacy, economic sanctions, cyber operations, and strengthening non-proliferation frameworks are usually far preferable alternatives.

   Notably, Israel and the United States jointly resorted to surgical strikes against Iran's nuclear infrastructure in June 2025. How far they were successful is not clear at this time, though past Israeli strikes on reactor projects in Iraq (1981) and Syria (2007) do seem to have been successful in preventing proliferation. Overall, short of imminent threats, such military strikes should thus remain a last resort, and should occur under strict international oversight if possible.

Conclusion

Religious ideologies, while not sole determinants, can play a significant role in the motivations and justifications behind nuclear proliferation. As indicated by the case studies above, religion can be (and sometimes is) used to sanctify and encourage nuclear ambitions, reinforce existential fears, and increase nuclear weapons-related risk acceptance. At the same time, it can also be used to constrain weapons

development. Understanding the interplay between theology, identity, and security is thus essential for policymakers, diplomats, and non-proliferation advocates alike.

Efforts to curb nuclear proliferation must move beyond traditional state-centric and rational-actor focused models and incorporate ideological and cultural factors into intelligence analysis and national security strategy. Engaging religious leaders, promoting inclusive narratives of peace, and addressing underlying sources of insecurity are crucial steps in building a more stable, less nuclearized world. By integrating these insights into international policy frameworks, the global community can better navigate the complex relationship between faith and force in the atomic age.

* * *

About the Author

Inderjit S. Grewal *is a graduate student in the School of Defense and Strategic Studies at Missouri State University, and a technology professional with over three decades of experience in the IT industry, specializing in automation and cybersecurity. The views expressed herein are entirely his own and do not necessarily represent those of any other institution or individual.*

Notes

1 *See generally, e.g.*, Shiv Kunal Verma, *1962: The War That Wasn't* (Aleph Book Company, 2016); Gerry van Tonder, *Sino-Indian War – Border Clash: October-November 1962* (Pen & Sword, 2018).

2 George Perkovich, *India's Nuclear Bomb: The Impact on Global Proliferation* (University of California Press, 1999), 163–170; William Burr & Jeffrey Richelson, "'Whether to 'Strangle the Baby in the Cradle': The United States and the Indian Nuclear Program, 1974–1976," *International Security*, vol. 25, no. 4, 2001, 54–99.

3 Perkovich, *India's Nuclear Bomb*, 447-52.

4 Paul S. Kapur, "India and Pakistan's Unstable Peace: Why Nuclear South Asia Is Not Like Cold War Europe," *International Security*, vol. 30, no. 2 (Fall 2001), 127–152.

5 Perkovich, *India's Nuclear Bomb*, 17.

6 *India Today*, May 25, 1998, 12. L.K. Advani stated, "The BJP government has shown the will to display our strength. India is now a nuclear weapons state. It has the capacity to defend itself. This has changed the strategic equation. It has added a new dimension to our national self-confidence. India is now a strong nation — and strength respects strength."

[7] The Communist Party of India (Marxist–Leninist) (CPI(ML)) was an Indian communist party formed by the All-India Coordination Committee of Communist Revolutionaries (AICCCR) at a congress in Calcutta in 1969. CPI (ML) saw Naxalbari as the spark that would start a new Indian revolution, and the movement came to be known as "Naxalites."

[8] In a letter to President Bill Clinton, Prime Minister Vajpayee declared that "We have an overt nuclear weapon state on our borders, a state which committed armed aggression against India in 1962. Although our relations with that country have improved in the last decade or so, an atmosphere of distrust persists mainly due to the unresolved border problem. To add to the distrust that country has materially helped another neighbor of ours to become a covert nuclear weapons state."

[9] Hindutva is a political ideology encompassing the cultural justification of Hindu nationalism and the belief in establishing Hindu hegemony within India.

[10] Chidanand Rajghatta, "The Hindu Bomb," *Indian Express*, May 21, 1998, https://indianexpress.com/article/news-archive/the-hindu-bomb/?ref=archive_pg.

[11] Nirad C. Chaudhuri, *The Continent of Circe: An Essay on the Peoples of India* (Oxford University Press, 1966), 104–105.

[12] Ashoka is remembered as a wise and generally peaceable Buddhist emperor, but his conversion to that faith is recounted as having occurred out of revulsion at witnessing the bloody results of his war of conquest against the rival kingdom of Kalinga.

[13] *See, e.g.*, Robert Marquant, "India Bomb Hits Chord for Hindus," *Christian Science Monitor*, June 4, 1998, https://www.csmonitor.com/1998/0604/060498.intl.intl.3.html; Bill Drexel, "How Competing Hindu Theologies Drove India's Nuclear Decision-Making – in Opposite Directions," Hudson Institute, April 4, 2025, https://www.hudson.org/missile-defense/how-competing-hindu-theologies-drove-indias-nuclear-decision-making-opposite-bill-drexel.

[14] Ramanna, Raja. *Years of Pilgrimage* (Viking, 1991), 93.

[15] The idea of a "peaceful nuclear explosion" (PNE) enjoyed at least some legitimacy at the time, having been mentioned in Article V of the Treaty on the Non-Proliferation of Nuclear Weapons (a.k.a. NPT) discussing "peaceful applications of nuclear explosions" – which held out the theoretical possibility that nuclear weapons possessors might perform such explosions *as a service* for non-possessors (*e.g.*, for large scale excavations of canals or reservoirs). No one today takes the idea of a PNE seriously.

[16] *See, e.g.*, Krishnaswamy Subrahmanyam, *"India and the Nuclear Bomb," Foreign Affairs* 77, no. 4, 1998, 33–44; Perkovich, *India's Nuclear Bomb*, 405.

[17] G.B. Singh, *Gandhi: Behind the Mask of Divinity* (Prometheus Books, 2004), 193-73.

[18] Atal Bihari Vajpayee, Speech in Lok Sabha, May 1998, in *Lok Sabha Debates* (Parliament of India), vol. 148, No. 11, May 1998.

[19] Chaudhuri, *The Continent of Circe*, 87.

[20] John Cherian, "The BJP and the Bomb," *Frontline*, April 11, 1998, https://frontline.thehindu.com/cover-story/article30161203.ece.

[21] L.K. Advani, remarks reprinted in *India Today*, May 25, 1998, 12.

[22] A.P.J. Abdul Kalam & Arun Tiwari *Wings of Fire: An Autobiography* (Universities Press, 1999), 158. The Defence Research and Development Organisation (DRDO) is an agency under the Department of Defence Research and Development in the Ministry of Defence of the Government of India, charged with the military's research and development, headquartered in Delhi, India.

[23] The objective of the Treaty on the Non-Proliferation of Nuclear Weapons is to prevent the spread of nuclear weapons and weapons technology, to promote cooperation in the peaceful uses of nuclear energy, and to further the goal of achieving nuclear disarmament and general and complete disarmament.

[24] Perkovich, *India's Nuclear Bomb*, 104.

[25] Chaudhuri, *The Continent of Circe*, 79.

[26] Thomas Babington Macaulay, "Minute on Indian Education" (1835), *Selections from Educational Records,* Part I (1781–1839), quoted by Bruce Tiebout McCully, "English Education and the Origins of Indian Nationalism," *Studies in History, Economics, and Public Law*, no. 473 (Columbia University Press, 1940).

[27] Richard H. Davis, *The Bhagavad Gita: A Biography* (Princeton University Press, 2014), 5.

[28] Kurukshetra is a city in the north Indian state of Haryana. It is known as the setting of the Hindu epic the "Mahabharata."

[29] Chaudhuri, *The Continent of Circe*, 114.

[30] M.K. Gandhi, *The Collected Works of Mahatma Gandhi*, vol. 88 (May 25-July 31, 1947) (Publications Division, Ministry of Information and Broadcasting, Government of India, 1983), 22 (emphasis added).

[31] S.R. Tikekar, *Epigrams from Gandhiji* (Publications Division, Ministry of Information & Broadcasting, Government of India, 1971), 253).

[32] A person professing sexual abstinence and an ascetic mode of life.

[33] M.K. Gandhi, *The Collected Works of Mahatma Gandhi*, vol. 21 (August 21, 1921 – on or after December 14, 1921) (Publications Division, Ministry of Information and Broadcasting, Government of India, 1966), 133-136

[34] Itty Abraham, *The Making of the Indian Atomic Bomb: Science, Secrecy and the Postcolonial State* (Zed Books, 1998), 49.

[35] Singh, *Gandhi: Behind the Mask of Divinity*, 314.

[36] *Modern Hinduism*, as I use the term here, is a combination of ideologies like Theosophy, Communism, and Fascism. The rise of Hindu nationalism based upon Modern Hinduism represents the politicization and militarization of traditional Hinduism. Religious imagery and new temples like the Bharat Mata Mandirs symbolize the merging of Hindu devotion with nationalist propaganda. Modern Hinduism often fosters violence against non-Hindu communities. The RSS and its global networks are also active in North America and pose serious threats to religious pluralism and social harmony.

[37] Chaudhuri, *The Continent of* Circe, 78.

[38] Nigel Calder, *Nuclear Nightmares: An Investigation into Possible Wars* (Penguin Books, 1979), 83.

[39] Steven Weissman & Herbert Krosney, *The Islamic Bomb: The Nuclear Threat to Israel and the Middle East* (Times Books, 1981), 161.

[40] The Indo-Pakistani war of 1971, also known as the third India-Pakistan war, was a military confrontation between India and Pakistan that occurred during the Bangladesh Liberation War in East Pakistan from December 3, 1971, until the Pakistani capitulation in Dhaka on December 16, 1971.

[41] Weissman & Krosney, *The Islamic Bomb*, 39.

[42] William E. Burrows & Robert Windrem, *Critical Mass: The Dangerous Race for Superweapons in a Fragmenting World* (Simon & Schuster, 1994), 360.

[43] Burrows & Windrem, *Critical Mass*, 351.

[44] *See generally, e.g.*, Avner Cohen, *The Worst-Kept Secret: Israel's Bargain with the Bomb* (Columbia University Press, 2010).

[45] *See generally, e.g.*, Avner Cohen, *Israel and the Bomb* (Columbia University Press, 1998), 277-91.

[46] Dhirendra Sharma, "India's Lopsided Science," *Bulletin of the Atomic Scientists*, May 1991, 36.

[47] See, for example, the range of estimates given by Cohen, *The Worst-Kept Secret*, xxvii.

[48] *See, e.g.*, International Atomic Energy Agency, "Implementation of the NPT Safeguards Agreement and relevant provisions of Security Council resolutions in the Islamic Republic of Iran," GOV/2011/65 (November 8, 2011), https://www.iaea.org/sites/default/files/documents/gov2011-65.pdf.

[49] *Foreign Policy* reports that when asked about the possibility of developing chemical or nuclear weapons, Ayatollah Khomeini responded sharply: "We don't want to produce nuclear weapons," and instead directed nuclear scientists to civilian pursuits at the Atomic Energy Organization. https://foreignpolicy.com/2014/10/16/when-the-ayatollah-said-no-to-nukes/.

[50] The Arabic term "Mazlum" broadly refers to the "oppressed, ill-treated, injured, and sinned-against."

[51] Alfoneh, Ali. "Iran's Nuclear Program and Shiite Political Theology." *Middle East Quarterly* 19, no. 4 (Fall 2012): 45–55.

[52] F. Gregory Gause III, *The International Relations of the Persian Gulf* (Cambridge University Press, 2010), 52.

[53] The Ministry for Atomic Energy of the Russian Federation was established on January 29, 1992, as a successor of the Ministry of Nuclear Engineering and Industry of the former USSR.

[54] *See* Dmitry Adamsky, *Russian Nuclear Orthodoxy: Religion, Politics, and Strategy* (Stanford University Press, 2019).

Defense and Strategic Studies Online (DASSO) *is produced by the School of Defense & Strategic Studies at Missouri State University. The views expressed herein are entirely those of the individual authors, and do not necessarily reflect those of anyone else at Missouri State University or elsewhere.*

The School of Defense and Strategic Studies is fully accredited by the Missouri Higher Learning Commission and is certified by the State Council of Higher Education for Virginia.

Are *you a member or friend of the Missouri State University community with an interest in defense and strategic studies and who has something thoughtful to contribute to public discourse on such topics in this short essay format? We welcome essays, commentary, accounts and interpretations of original research, book reviews, and more. Contact DASSO's Editorial Board at DASSO@MissouriState.edu.*

Missouri State

SCHOOL of DEFENSE AND STRATEGIC STUDIES

No. 4 (Summer 2025)