Technological developments that could increase risks from nuclear weapons

Michael Aird and Will Aldred

This is a blog post, not a research report, meaning it was produced relatively quickly and is not to Rethink Priorities' typical standards of substantiveness and careful checking for accuracy.

Summary

This post is **a shallow exploration of some technological developments that *might* occur and *might* increase risks from nuclear weapons—especially existential risk or other risks to the long-term future**. This is one of many questions relevant to how much to prioritize nuclear risk relative to other issues, what risks and interventions to prioritize *within* the nuclear risk area, and how that should change in future. But note that, due to time constraints, this post isn't comprehensive and was less thoroughly researched and reviewed than we'd like.

For each potential development, we provide some very quick, rough guesses about how much and in what ways the development would affect the odds and consequences of nuclear conflict (**"Importance"**), the likelihood of the development in the coming decade or decades (**"Likelihood/Closeness"**), and how much and in what ways thoughtful altruistic actors could influence whether and how the technology is developed and used (**"Steerability"**).[1]

These tentative bottom line beliefs are summarized in the table below:

Category	Technological Development	Importance	Likelihood / Closeness	Steerability
Bomb types and production methods	Radiological weapons	Medium	Medium/High	Medium/Low
	Pure fusion bombs	Medium	Medium/Low	Medium
	High-altitude electromagnetic pulse (HEMP)	Medium	Medium/Low	Medium/Low
	Neutron bombs	Low	Medium/Low	Medium
Methods for	Atomically precise manufacturing (APM)	High	Low	Medium

[1] For steerability, we especially have in mind the extent to which *actors associated with the effective altruism community* could affect whether and how the technology is developed and used. In practice, this may typically ultimately involve influencing the US government or to a much lesser extent the UK government (whether this influence occurs directly or via influencing other actors, e.g. advocacy groups or non-EA think tanks). This is because (a) those governments are among the key players for these technological developments and for nuclear risk more broadly, and (b) most of the other key players would probably be far harder for most EA-aligned actors to influence (e.g., the Russian, Chinese, and North Korean governments). That said, for some developments, steering may also or instead involve influencing scientific communities or corporate R&D labs, especially those relevant to AI.

production and design	AI-assisted production/design	Medium/High	Medium/Low	Medium
	Other developments in methods for production/design	?	?	?
Delivery systems	Hypersonic missiles/glide vehicles	Medium/Low	Medium/High	Medium/Low
	More accurate nuclear weapons	Medium	Medium	Medium/Low
	Long-range conventional strike capabilities	Medium/Low	Medium/High	Medium/Low
Detection and defense	Better detection of nuclear warhead platforms, launchers, and/or delivery vehicles	Medium/High	Medium/High	Medium/Low
	Missile defense systems	Medium	Medium	Medium/Low
AI and cyber	Advances in AI capabilities	Medium/High	Medium/High	Medium
	Cyberattack (or defense) capabilities	Medium/High	Medium/High	Medium
	Advances in autonomous weapons	Medium	Medium/High	Medium
	More integration of AI with NC3 systems	Medium	Medium	Medium
Non-nuclear warmaking advances	Anti-satellite weapons (ASAT)	Medium/Low	Medium	Medium/Low
	"Space planes" and other (non-ASAT) space capabilities	Medium/Low	Medium	Medium/Low

Note that:

- **Each "potential technological development" is really more like a somewhat wide area in which a variety of different types and levels of development could occur, which makes the ratings in the above table less meaningful and more ambiguous.**
- "Importance" is here assessed *conditional on the development occurring*, so will overstate the importance of thinking about or trying to steer unlikely developments.
- In some cases (e.g, "More accurate nuclear weapons"), the "Importance" score accounts for potential risk-*reducing* effects as well.
- "Likelihood/Closeness" is actually inelegantly collapsing together two different things, making our ratings of developments on that criterion less meaningful. E.g., one development could be moderately likely to occur quite soon and moderately likely to occur never, while another is very likely to occur in 15-25 years but not before then.
- Some of the topics this post discusses involve or are adjacent to information hazards (especially attention hazards), as is the case with much other discussion of technological developments that might occur and might increase risks. We ask that readers remain mindful of that when discussing or writing about this topic.
    - Additionally, for this reason, we removed some discussion of the above technological developments and some other technological developments.[2] Feel free to reach out to us if you think it would be useful for you to see those additional rough notes.

Epistemic status

In 2021, Michael did some initial research for this post and wrote an outline and rough notes. But he pivoted away from nuclear risk research before having time to properly research and

[2] We also solicited feedback on information hazard levels from reviewers of this post, including highlighting specific sections to specific people.

draft this. We (Michael and Will) finished a rough version of this post in 2022, since that seemed better than it never being published at all, but then didn't get around to publishing till 2023.[3] As such, this is just a very incomplete starting point, may contain errors, and may be outdated.[4] It could be quite valuable for someone to spend more time:

- learning about other possible technological developments worth paying attention to,
- doing more thorough and careful research on the developments we discuss,
- thinking more about their implications for how much to prioritize nuclear risk reduction and what interventions and policies to pursue, and/or
- talking to and getting feedback from various experts

(See also Research project idea: Technological developments that could increase risks from nuclear weapons.)

How to engage with this post

The full post can be found here. It's ~23,000 words, much of which is extensive quotes without added commentary from us. Some quotes are relevant to multiple sections and hence are repeated. **But each section or subsection should make sense by itself, so readers should feel free to read only the sections that are of interest to them, to skim, and to skip repeated quotes and "Additional notes" sections.** You can navigate to sections using the links in the summary table above.

Scope of this post

This post is focused on what *potential technological developments* could *increase* risks *from nuclear weapons*. As such, this post is *not* necessarily claiming that these technologies will be net harmful overall, nor that nuclear risk will increase in future overall; both of those claims are plausibly true and plausibly false, and we don't assess them here.[5] Here are some further notes on what this post is vs isn't focusing on and claiming:

- **We're not claiming any of these developments are guaranteed** - in fact, we think several are unlikely or would only happen a long time from now.
- We don't address **things that could increase risk from nuclear weapons but that aren't technological developments.**[6]

[3] Where this post says "I" or other first person singular words, that refers to Michael.
[4] E.g., we didn't try to think about ways in which the Russian invasion of Ukraine or developments in US-China relations should update our views.
[5] For some very initial thoughts on the latter question, see "How might nuclear risk change over time? Why? What can be done?"
[6] Some other things that might increase risk from nuclear weapons include:

- Increases in arsenal sizes
- Increases in yields (of regular nuclear weapons)

- We mostly focus on possible new technological developments, **setting aside proliferation of existing technologies or changes in how those technologies are deployed**.[7]
    - However, sometimes these lines are blurry, and we did end up discussing some things that may be more like deployment than development (e.g., integration of AI with NC3).
- We don't focus on discussing ways **these technological developments might *also decrease* nuclear risks, whether their *net effect* might be a decrease in nuclear risk (even if they could also have important risk-increasing effects), or *other* technological developments that could decrease nuclear risks**.
    - In a few places we do touch on those points, but we didn't set out to do so, and thus there's far more that could be usefully said than what we've said in this post.
- We had hoped to discuss **what could and should be done to influence whether and how these technological developments occur and are used, and what other implications these potential developments might have** for what risks and interventions to prioritize in the nuclear risk space. But ultimately we ran out of time, and hence this post only contains extremely preliminary and patchy discussion of those questions.
    - As expressed here, we think those questions would likely be one of the most valuable directions for further research building off this post, and potentially one of the most valuable directions for further research on nuclear risk in general.
    - For more general discussion of possible goals and interventions related to nuclear risk, see 8 possible high-level goals for work on nuclear risk and Shallow review of approaches to reducing risks from nuclear weapons.

- Changes in targeting policies (e.g., toward an increase chance of large numbers of countervalue strikes)
- Other changes in policy/diplomacy that increase the chance of nuclear war
- Increases in geopolitical tensions or
- Population increase, urbanization, and wealth increases (which will tend to increase fuel density in targeted areas)

For more, see our rough notes on "How might nuclear risk change over time? Why? What can be done?"

[7] In other words, I mostly focus on ways people might expand the technological frontier, as opposed to countries or people "catching up to" that frontier or changing how they use the technologies within the current frontier.

Bomb types and production methods

Radiological weapons / Salted bombs

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium/High	Medium/Low

What this is

- "**Radiological warfare** is any form of warfare involving deliberate radiation poisoning or contamination of an area with radiological sources." (Wikipedia)
- "A **salted bomb** is a nuclear weapon designed to function as a radiological weapon, producing enhanced quantities of radioactive fallout, rendering a large area uninhabitable. [...] A salted bomb should not be confused with a "dirty bomb", which is an ordinary explosive bomb containing radioactive material which is spread over the area when the bomb explodes. A salted bomb is able to contaminate a much larger area than a dirty bomb." (Wikipedia)
- "A **cobalt bomb** is a type of "salted bomb": a nuclear weapon designed to produce enhanced amounts of radioactive fallout, intended to contaminate a large area with radioactive material, potentially for the purpose of radiological warfare, mutual assured destruction or as doomsday devices. [...] A cobalt bomb could be made by placing a quantity of ordinary cobalt metal (^{59}Co) around a thermonuclear bomb. When the bomb explodes, the neutrons produced by the fusion reaction in the secondary stage of the thermonuclear bomb's explosion would transmute the cobalt to the radioactive cobalt-60, which would be vaporized by the explosion. The cobalt would then condense and fall back to Earth with the dust and debris from the explosion, contaminating the ground." (Wikipedia)

What developments might occur? How likely are they?

Some historical info:

"No intentionally salted bomb has ever been atmospherically tested, and as far as is publicly known, none has ever been built. However, the UK tested a one-kiloton bomb incorporating a small amount of cobalt as an experimental radiochemical tracer at their Tadje testing site in Maralinga range, Australia, on September 14, 1957. The triple "taiga" nuclear salvo test, as part of the preliminary March 1971 Pechora–Kama Canal project, converted significant amounts of stable cobalt-59 to radioactive cobalt-60 by fusion-generated neutron activation and this product is responsible for about half of the gamma dose measured at the test site in 2011. The experiment was regarded as a failure and not repeated." (Wikipedia)

Info on Russia's in-development Poseidon system:

Ladish (2019) writes:

- "Status 6 / Poseidon is an underwater, high yield [see nuclear weapon yield], autonomous, radiological weapon being developed by Russia. There is a lot to unpack here. First, this may be the first time a nuclear state has announced a weapons system aimed explicitly at radiological area denial.
- Status 6 may also be the first autonomous nuclear weapons system with the ability to do anything more complex than simply reach its target. ICBMs are 'autonomous' in a sense but they aren't very smart."

Kristensen & Korda (2022) write:

- "Moreover, the fact that Russian military planners are pursuing a broad range of upgraded and new versions of nuclear weapons suggests that the real doctrine goes beyond basic deterrence and toward regional war-fighting strategies, or even weapons aimed at causing terror. One widely-cited example involves the so-called **Status-6—known in Russia as "Poseidon" and in the United States as "Kanyon"—a long-range nuclear-powered torpedo that a Russian government document described as intended to create "areas of wide radioactive contamination that would be unsuitable for military, economic, or other activity for long periods of time"** (Podvig 2015). A diagram and description of the proposed weapon, first revealed in a Russian television broadcast, can still be seen on YouTube (YouTube 2015). The weapon, which is under development, **appears designed to attack harbors and cities to cause widespread indiscriminate collateral damage in violation of international law**." (emphasis added)
- "The Russian Navy is also developing the Status-6 Poseidon mentioned above—a nuclear-powered, very long range, nuclear-armed torpedo. Underwater trials began in December 2018. The weapon is **scheduled for delivery in 2027** and will be carried by specially configured submarines (TASS 2018f). The first of these special submarines—the Project 09852 Belgorod (K-329)—was launched in April 2019 and was originally scheduled for delivery to the Navy by the end of 2020; however, it only began sea trials in June 2021 and returned to dry dock in October 2021 (Sutton 2021a; Sutton 2021b). State trials are scheduled for 2022, which could indicate that delivery to the Navy could be delayed until late 2022 (TASS 2021o). Belgorod will become Russia's largest submarine and reportedly will be capable of carrying up to six Poseidon torpedoes (TASS 2019d). The launch of the second Poseidon-capable submarine—Project 09851 Khabarovsk—was expected to take place in the autumn of 2021, but appears to have been delayed until 2022 (TASS 2021g). Khabarovsk will reportedly also be capable of carrying up to six Poseidon torpedoes (TASS 2020b)." (emphasis added)

From Wikipedia's article on the Status-6 Oceanic Multipurpose System:

- "It is 1.6–2 metres in diameter and 24 metres long. The warhead shown in the leaked figure is a cylinder 1.5 metres in diameter by 4 metres in length, giving a volume of 7 cubic meters. Comparing this to the volumes of other large thermonuclear bombs, the

1961 Soviet-era Tsar Bomba itself measured 8 metres long by 2.1 metres in diameter, indicating that the yield is at least several tens of megatons, generally consistent with early reports of 100 megatons. Some reports suggest the yield of the Poseidon's warhead is as low as 2 Mt."

- "Oscar-class submarines could carry six Poseidon torpedoes at the same time for a total yield of up to 400 megatons."

Russia's Exotic Nuclear Weapons and Implications for the United States and NATO states:

> "The Poseidon, a nuclear-armed underwater drone, is the final exotic Russian strategic system. Eight Poseidon drones would be carried by, and launched from the torpedo tubes of, a nuclear-powered, guided-missile submarine (SSGN). While the Poseidon can be armed with conventional or nuclear payloads, **its ability to carry a large-yield nuclear warhead has attracted much attention. Indeed, there is even speculation that the Poseidon would be laden with a multi-megaton warhead seeded with cobalt—which would result in particularly deadly nuclear fallout. Given that the Poseidon operates deeply underwater, it is unlikely that it could be guided by satellite navigation; therefore, its delivery would probably be inaccurate.** Accordingly, when targeted at the US eastern seaboard, for example, it could be expected to hit "somewhere between Charleston, SC and Charlestown, MA," as one participant claimed. Like the Burevestnik, this system may possess a loitering capability. **The weapon could also be used at closer ranges in a counterforce capability against large fleet formations. Notably, some Russian commentators, including former Russian military officers, have criticized this weapon for being too noisy and slow, and, thus, vulnerable to interception.**" (emphasis added)

But the same report also notes that it remains somewhat unclear whether and when Poseidon and other new Russian nuclear systems will actually be deployed:

> "**Some have questioned whether these are real capabilities, and whether Russia is making genuine progress on their development and deployment. Indeed, there is clear evidence that Russia is having difficulties with at least some of these systems.** It was widely reported, for example, that Russia experienced an accident while testing the Skyfall system. Further, the estimated deployment times listed above are those claimed in the Russian press, and have not been verified by Western sources. **Still, it is clear that the announced systems are not merely design studies.** Russia is "spending money and bending metal," as one participant stated, demonstrating that it is actively working toward these capabilities. Moreover, **Russia has always been more comfortable rushing weapons systems into the field at a pace that would not be possible in the United States**, given the current state of its nuclear complex, acquisition process, and other issues. **It is likely, therefore, that most, if not all, of these systems will be deployed in the coming years.**
>
> [...] These weapons may also be used for nuclear signaling in a crisis. With its near-inexhaustible power source, the nuclear-powered, nuclear-armed cruise missile would have the ability to loiter over and around targets in Western Europe or the United

States. Imagine in a crisis, for example, a nuclear-powered cruise missile overflying Europe and back. The Poseidon submarine drone could lie in wait in or near a major US or allied harbor. Upon their detection, or after Russia announced their presence, these loitering systems could provide a tangible and proximate reminder of escalation risks. They may also blur the lines between crisis and war, as the United States and its allies would debate whether attacking a loitering system constituted an act of war. Moreover, given that several of these systems are dual-capable, their deployment, even in conventional conflicts, could contribute to uncertainty and perceptions of nuclear escalation risks.

On the other hand, there are practical hurdles to such employment, and it may be difficult to imagine Russia conducting its nuclear command and control in this way. Autocratic systems in general, and Russia in particular, tend to prefer strict, top-down control of nuclear weapons. Russia may not be willing to run the risks associated with losing control of loitering weapons or of having them shot down." (emphasis added)

Why might this increase nuclear risk?

- It seems plausible that the use of sufficient numbers and sizes of salted bombs could create a pathway to global or existential catastrophes *other than* the more commonly considered (and probably more important) pathways of (a) nuclear winter and (b) the more immediate effects of large numbers of regular nuclear weapons.
- It also seems plausible that development of this technology would change how much and in what ways nuclear weapons (broadly construed) or nuclear conflict act as an existential risk factor.
- But we didn't have time to think through the details of those arguments or figure out how important they are.
- See also "Research project idea: Direct and indirect effects of nuclear fallout"
- One weak indication (via deference) that this might indeed by important is that Shulman (2020) writes:
    - "I agree it's very unlikely that a nuclear war discharging current arsenals could directly cause human extinction. But the conditional probability of extinction given all-out nuclear war can go much higher if the problem gets worse. Some aspects of this:

      [Many points raised]

      radiation effects could be intentionally greatly increased with alternative warhead composition"
    - (This doesn't include any information we didn't already have except that Shulman appears to see this sort of thing as worth paying attention to as a contributor to extinction risk.)

- Another thing we haven't looked into: Would salted bombs also add much to the smoke that creates a risk of nuclear winter? Or would they actually create *less* of that particular risk than regular nuclear weapons do?

What could be done about this?

Some very rough, speculative ideas:

- Perhaps reduce the extent to which people think about salted bombs at all, or the extent to which they think salted bombs would be quite useful for deterrence or other reasons?
- Perhaps increase the extent to which people consider salted bombs risky and/or immoral and hence bad to create or *at least* bad to create in large numbers and sizes?
    - But note that actions aimed at this goal could also pose attention hazards.
- Perhaps try to somehow stop, slow, or limit Russia's development and deployment of the Poseidon system?
    - But that's probably very hard to do for anyone who isn't a high-level Russia government or military official?
    - And actions aimed at this goal could perhaps contribute to the Russian government, or other actors, feeling that people are scared of these weapons and hence that developing them gives them leverage/power.

Additional notes

We found this quote from Wikipedia interesting in relation to the idea of attention hazards from EA analysis and advocacy on global catastrophic risks:

> "A cobalt bomb was first suggested by Leo Szilard, who publicly sounded the alarm against the possible development of a salted thermonuclear bombs that might annihilate mankind in a University of Chicago Round Table radio program on February 26, 1950. His comments, as well as those of Hans Bethe, Harrison Brown, and Frederick Seitz (the three other scientists who participated in the program), were attacked by the Atomic Energy Commission's former Chairman David Lilienthal, and the criticisms plus a response from Szilard were published. Time compared Szilard to Chicken Little while the AEC dismissed his ideas, but scientists debated whether it was feasible or not. The Bulletin of the Atomic Scientists commissioned a study by James R. Arnold, who concluded that it was. Clark suggested that a 50 megaton cobalt bomb did have the potential to produce sufficient long-lasting radiation to be a doomsday weapon, in theory, but was of the view that, even then, "enough people might find refuge to wait out the radioactivity and emerge to begin again.""

This seems like it *might* be an example of a well-intentioned, thoughtful, foresighted person sounding alarm bells too publicly and prominently and hence plausibly increasing risks overall or at least increasing risks relative to what they'd be if the person sounded alarm bells to smaller audiences or with different framings. Though we haven't looked into this enough to be confident.

Pure fusion bombs

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium/Low	Medium

What this is and why it might increase risk

From Wikipedia:

- "A **pure fusion weapon** is a hypothetical hydrogen bomb design that does not need a fission "primary" explosive to ignite the fusion of deuterium and tritium, two heavy isotopes of hydrogen used in fission-fusion thermonuclear weapons. **Such a weapon would require no fissile material and would therefore be much easier to develop in secret than existing weapons.** Separating weapons-grade uranium (U-235) or breeding plutonium (Pu-239) requires a substantial and difficult-to-conceal industrial investment, and blocking the sale and transfer of the needed machinery has been the primary mechanism to control nuclear proliferation to date. Due to its not requiring a fission primary explosive to initiate a fusion reaction, the pure fusion weapon **would also have greatly increased potential yield over current thermonuclear weapons**." (emphasis added)
- "These weapons would be lethal not only because of their explosive force, which could be large compared to bombs based on chemical explosives, but also because of the neutrons they generate."[8]
- "many have expressed concern that pure fusion weapons research and development would subvert the intent of the Nuclear Non-Proliferation Treaty and the Comprehensive Test Ban Treaty."

So it seems plausible that such weapons would increase arsenal sizes and proliferation, and also separately increase nuclear winter risk via increased yields. But we haven't tried to assess any of those claims carefully.

What developments might occur?

Wikipedia indicates that nuclear weapons designers in nuclear weapon possessor states, including or at least the US, have put substantial effort into creating pure fusion weapons.

- **But we don't know if those claims are accurate.**
    - The page has few citations, and we haven't looked into them.

[8] "Due to the high kinetic energy of neutrons, [neutron] radiation is considered the most severe and dangerous radiation to the whole body when it is exposed to external radiation sources" (Wikipedia). The 3 more commonly occurring types of radiation are alpha, beta, and gamma (source).

    - We spent a few minutes googling and found some other things that looked like maybe credible claims that the US had made efforts towards developing pure fusion weapons, but nothing clearly decisively stating that.
    - It would probably be good if someone tried to find and talk to a relevant expert about this
- The relevant passages from the page are:
    - “For many years, nuclear weapon designers have researched whether it is possible to create high enough temperatures and pressures inside a confined space to ignite a fusion reaction, without using fission.”
    - “Despite the many millions of dollars spent by the U.S. between 1952 and 1992 to produce a pure fusion weapon, no measurable success was ever achieved. In 1998, the U.S. Department of Energy (DOE) released a restricted data declassification decision stating that even if the DOE made a substantial investment in the past to develop a pure fusion weapon, ‘the U.S. is not known to have and is not developing a pure fusion weapon and no credible design for a pure fusion weapon resulted from the DOE investment’.”
        - Note: “many millions” actually sounds pretty small by US military standards?
- **If the above claims are true, that seems crazy and horrible, and would also make us more more worried about anthropogenic existential risk more broadly by providing negative evidence about the competence and caution of powerful governments and militaries in relation to risky technology development.**
    - It seems like it would be very much against the interests of the US to design a new type of nuclear weapon whose proliferation would be harder to control and which could more easily be developed by other actors in secret.
    - It also seems bad if the potential yield was indeed greatly increased. But that’s less obviously against the US’s own interests (even if bad from an impartial longtermist perspective), so wouldn’t provide as strong of a negative update regarding the overall *sanity/rationality* of powerful institutions.

The same Wikipedia article also says that a pure fusion weapon doesn’t yet exist and may be hard/infeasible to create.

- “While various neutron source devices have been developed, some of them based on fusion reactions, none of them are able to produce an energy yield, either in controlled form for energy production or uncontrolled for a weapon.”
- “The power densities needed to ignite a fusion reaction still seem attainable only with the aid of a fission explosion, or with large apparatus such as powerful lasers like those at the National Ignition Facility, the Sandia Z-pinch machine, or various magnetic tokamaks. Regardless of any claimed advantages of pure fusion weapons, building those weapons does not appear to be feasible using currently available technologies”.
- That said, the article also says: “It has been claimed that it is possible to conceive of a crude, deliverable, pure fusion weapon, using only present-day, unclassified technology. The weapon design weighs approximately 3 tonnes, and might have a total yield of approximately 3 tonnes of TNT. The proposed design uses a large explosively pumped

flux compression generator to produce the high power density required to ignite the fusion fuel. From the point of view of explosive damage, such a weapon would have no clear advantages over a conventional explosive, but the massive neutron flux could deliver a lethal dose of radiation to humans within a 500-meter radius (most of those fatalities would occur over a period of months, rather than immediately)."

The same article also states: "Nanotechnology can theoretically be used to develop miniaturized laser-triggered pure fusion weapons that will be easier to produce than conventional nuclear weapons."

One source claims that Russia may be interested in developing pure fusion weapons.

- "One emerging but realistic technology that Russia may be inclined to develop, which could offer significant military advantages and disrupt existing deterrence and arms control paradigms, is low-yield, pure fusion fourth generation nuclear weapons (FGNWs). This new class of weapons could be designed with a highly-tailorable range of yields and would produce significantly less residual radiation and collateral damage, making them well-suited for close integration with maneuver forces in regional conflicts."
- We haven't tried to verify this claim.
- As with the claim at the start of this subsection, *if true*, that would seem *probably* crazy and horrible and to warrant a negative update regarding the sanity/rationality of powerful institutions (since Russia, like the US, has an interest in ensuring nuclear proliferation stays difficult and controllable).

High-altitude electromagnetic pulse (HEMP)

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium/Low	Medium/Low

What this is

High-altitude electromagnetic pulse (HEMP) or nuclear EMP essentially refers to a particular way of using regular nuclear bombs, rather than a new type of bomb or production method. It might therefore not belong in this post at all. We included it anyway partly because we felt there *might* be plausible relevant technological developments, like the design of new bomb types more suited to usage as HEMPs, but that might be inaccurate.

HEMP is an electromagnetic energy field produced in the atmosphere by a nuclear detonation, and which is damaging to electronic equipment over a very wide area (CRS Report RL32544, 2008).[9]

From CRS Report RL32544 (2008):

> "The effects of HEMP became fully known to the United States in 1962 during a high-altitude nuclear test (code named 'Starfish Prime') over the Pacific Ocean, when radio stations and electronic equipment were disrupted 800 miles away throughout parts of Hawaii. The HEMP effect can span thousands of miles, depending on the altitude and the design and power of the nuclear burst (a single device detonated at an appropriate altitude over Kansas reportedly could affect all of the continental United States), and can be picked up by metallic conductors such as wires, or overhead power lines, acting as antennas that conduct the energy shockwave into the electronic systems of cars, airplanes, or communications equipment."

Why might this increase nuclear risk?

- See some work by ALLFED, e.g. Feeding Everyone if Industry is Disabled
- See Michael's [rough notes] Harms from nuclear conflict via EMPs, fallout, or ozone depletion
- Baum (2015a) writes: "An important question is whether command and control systems can be shielded. If they cannot be, then electromagnetic weapons could be used in a first strike, thereby reducing strategic stability and rendering electromagnetic weapons less suitable for deterrence."

Relevant forecasts

(These are all from Metaculus.)

- Will there be at least one HEMP attack by 2024?
- Will at least one HEMP attack occur by 2024, if an offensive nuclear detonation occurs by 2024?
- If one or more HEMP attacks occur by 2030, will that lead to >10 million fatalities?
- How many HEMP attacks will occur by 2030, if at least one does?

---

[9] We would assume (based on thinking about this from a physics perspective) that the effects of a HEMP depend on yield of the nuclear bomb and altitude of the detonation. In particular, we expect that damage is greater the greater the yield of the nuclear bomb, and that, broadly speaking, the area affected is greater the higher the altitude of the detonation, up to a point, while damage intensity on the ground decreases with altitude of detonation. (We haven't looked for sources to validate these expectations.)

It's unclear what the point where this stops being the case is. U.S. Air Force Civil Engineer Center (2020) says "Electromagnetic pulses (EMP) from nuclear weapon detonations at altitudes from 30 to 400 kilometers (18 to 50 miles) can damage or destroy sensitive electronic equipment at ground level", but there must be a typo somewhere there since 400 kilometers does not equal 50 miles, and we couldn't quickly find another source to resolve the uncertainty/contradiction here.

Neutron bombs

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Low	Medium/Low	Medium

Beyond reading the Wikipedia page, we've spent no time looking into whether there are versions of neutron bombs that could plausibly be created and would pose major existential risks; whether any actors have tried, are trying, or might try to develop such weapons; or whether and how to influence developments in this area. But we think it could be valuable for someone to look into that at least briefly (e.g., at least asking a couple experts whether there's any plausible way this could be important).

What this is and what developments might occur

Wikipedia states:

> "A **neutron bomb**, officially defined as a type of enhanced radiation weapon (ERW), is a low-yield thermonuclear weapon designed to maximize lethal neutron radiation in the immediate vicinity of the blast while minimizing the physical power of the blast itself. The neutron release generated by a nuclear fusion reaction is intentionally allowed to escape the weapon, rather than being absorbed by its other components. The neutron burst, which is used as the primary destructive action of the warhead, is able to penetrate enemy armor more effectively than a conventional warhead, thus making it more lethal as a tactical weapon.[10]
>
> The concept was originally developed by the US in the late 1950s and early 1960s. It was seen as a "cleaner" bomb for use against massed Soviet armored divisions. As these would be used over allied nations, notably West Germany, the reduced blast damage was seen as an important advantage.
>
> ERWs were first operationally deployed for anti-ballistic missiles (ABM). In this role the burst of neutrons would cause nearby warheads to undergo partial fission, preventing them from exploding properly. For this to work, the ABM would have to explode within approximately 100 metres (300 ft) of its target. The first example of such a system was the W66, used on the Sprint missile used in the US's Nike-X system. It is believed the Soviet equivalent, the A-135's 53T6 missile, uses a similar design.
>
> The weapon was once again proposed for tactical use by the US in the 1970s and 1980s, and production of the W70 began for the MGM-52 Lance in 1981. This time it led

[10] Our note: "A tactical nuclear weapon (TNW) or non-strategic nuclear weapon is a nuclear weapon which is designed to be used on a battlefield in military situations mostly with friendly forces in proximity and perhaps even on contested friendly territory" (Wikipedia).

to protests as the growing anti-nuclear movement gained strength through this period. Opposition was so intense that European leaders refused to accept it on their territory. President Ronald Reagan ordered the production of the W70-3, which remained stockpiled in the US until they were retired in 1992. The last W70 was dismantled in 2011."

Why might this increase risk

We haven't actually heard anyone raise the idea that these weapons might increase existential risk, and we don't immediately see a reason to expect they would, so **our best guess is that this technology actually isn't worth much attention**. The only reason we mention this technology in this post at all is that in general it seems worth at least briefly considering the potential significance of *any* nuclear-weapon-like system whose capabilities or production method are significantly different from that of other systems.

Baum (2015a) seems mostly sanguine about neutron bombs.

Methods for production and design

Atomically precise manufacturing (APM)

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
High	Low	Medium

What is this, and what developments might occur?

Open Philanthropy write:

- "Atomically precise manufacturing is a proposed technology for assembling macroscopic objects defined by data files by using very small parts to build the objects with atomic precision using earth-abundant materials. **There is little consensus about its feasibility, how to develop it, or when, if ever, it might be developed.**" (emphasis added)
- "Unless APM is developed in a secret "Manhattan Project"—and there is disagreement about how plausible that is —**the people we spoke with believe it would be extremely unlikely for an observer closely watching the field to be surprised by a sudden increase in potentially dangerous APM capabilities.**" (emphasis added)
    - That said, Open Philanthropy's list of "Questions for further investigation" includes: "How confident can we be that there will be substantial lead time between early signs that APM is feasible and the deployment of APM?"

Why might this increase nuclear risk?

Open Philanthropy note that, if created, APM "would likely make it substantially easier to develop new weapons and quickly and inexpensively produce them at scale with an extremely small manufacturing base."

The concern related to nuclear risk is that APM could potentially make proliferation and/or huge nuclear arsenal sizes *much* more likely. Specifically:

- APM could make it *much* harder to monitor/control who has nuclear weapons, including even non-state groups or perhaps individuals.
- APM could make it *much* more likely that non-state groups or individuals would not only attain *one or a few* but rather *many* nuclear weapons.
    - This could land us in the very risky-seeming "easy nukes" scenario described in Bostrom's (2019) The Vulnerable World Hypothesis paper.
- APM could make it so that nuclear-armed states can more cheaply, easily, and quickly build hundreds, thousands, tens of thousands, or even millions of nuclear weapons.
    - The enhanced ease and speed of doing this could undermine arms race stability.
        - This might also mean that even just the perception of APM being on the horizon could undermine strategic stability, even before APM has arrived.
    - And nuclear conflicts involving huge numbers of warheads are *much* more likely to cause an existential catastrophe, or otherwise increase existential risk, than nuclear conflicts involving the tens, hundreds, or thousands of nuclear weapons each nuclear-armed state currently has.
    - A counterpoint is that at least some nuclear-armed states already have the physical capacity to make many more nuclear warheads than they do (given enough time to build up nuclear manufacturing infrastructure), but having many more warheads doesn't appear to be what these states are aiming for. This is some evidence that nuclear weapon production being easier to make might not result in huge arsenal sizes. But:
        - This is only *some* evidence.
        - And *some* nuclear-armed states (e.g., North Korea, Pakistan, and maybe India) may be developing nuclear weapons about as fast as they reasonably can, given their economies. As such, perhaps having access to APM would be likely to substantially affect at least *their* arsenal sizes.

(Note that we haven't spent much time thinking about these points, nor seen much prior discussion of them, so this should all be taken as tentative and speculative.)

Finally, **we should flag that APM could of course pose many risks (and benefits) *unrelated to* nuclear risk**, as discussed in (among other places) the above-linked Open Philanthropy post. This post focuses on the connection to nuclear risk merely because that's the scope of this post, not because we expect that to be the most important effect of APM.

AI-assisted production/design

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/High	Medium/Low	Medium

- There are presumably various ways advances in narrow or general AI systems could make production of nuclear warheads and delivery systems cheaper, easier, or harder to control.
    - However, we haven't really looked into any specifics.
    - It seems potentially worth someone shallowly investigating this (e.g., at least talking to a few experts and doing some googling).
- At the extreme end, AGI or other transformative artificial intelligence systems might have similarly huge effects on arsenal sizes and proliferation as some versions of APM could.
- Narrower or less advanced AI systems might have effects that are smaller but somewhat similar in kind.

Other developments in methods for production/design

- Presumably there might be other ways (besides APM and AI) to change the methods by which nuclear warheads and delivery systems are produced and to thereby make the process cheaper or make proliferation more likely.
    - However, we haven't looked into or heard of specific possible developments.
    - It seems potentially worth someone shallowly investigating this (e.g., at least talking to a few experts and doing some googling).
    - Perhaps 3D printing could be important here?
        - We base this idea solely on the "presentation notes" slide 10 of Robichaud (2022)
- If such developments occur, we imagine the consequences would be similar in kind to the consequences discussed for APM and AI, though probably less extreme than those for APM and for some AI developments.
- Shulman (2020) notes:
    - "[...] so with 1950s proportional military expenditures, half going to nukes, the US and China could each produce 20,000+ ICBMs, each of which could be fitted with MIRVs and several warheads, building up to millions of warheads over a decade or so; **the numbers could be higher for cheaper delivery systems**
    - **economies of scale and improvements in technology would likely bring down the per warhead cost**" (emphasis added)

Delivery systems

Hypersonic missiles/glide vehicles

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/Low	Medium/High	Medium/Low

What this is

(This section is somewhat repetitive.)

Reuters/ABC (2021):

> "Hypersonic missiles travel at more than five times the speed of sound in the upper atmosphere — or about 6,200km per hour. This is slower than an intercontinental ballistic missile (ICBM) but the shape of a hypersonic glide vehicle allows it to maneuver toward a target or away from defenses. Hypersonic missiles can also travel for longer [than ICBMs] without being detected by radar."[11]

Congressional Research Service (2021):

> "The key difference between missiles armed with HGVs and missiles armed with ballistic reentry vehicles (i.e., those that travel on a ballistic trajectory throughout their flight) is not their speed, but their ability to maneuver and change course after they are released from their rocket boosters. In addition, although it is not necessary, many concepts for the delivery of HGVs presume that the boosters will launch along a flatter, or depressed, trajectory than standard ballistic missiles, and will release their gliders at a lower altitude of flight.
>
> Taken together, the HGV's novel trajectory and maneuverability in flight would complicate a U.S. effort to detect, track, and defend against an attack. The United States would likely detect the booster's launch, as it would for the launch of any ballistic missile, but it would not be able to predict the HGV's flight path. In addition, although an HGV launched by a rocket booster would reach its target far more quickly than a warhead delivered by an aircraft or subsonic cruise missile (in minutes instead of hours), it would

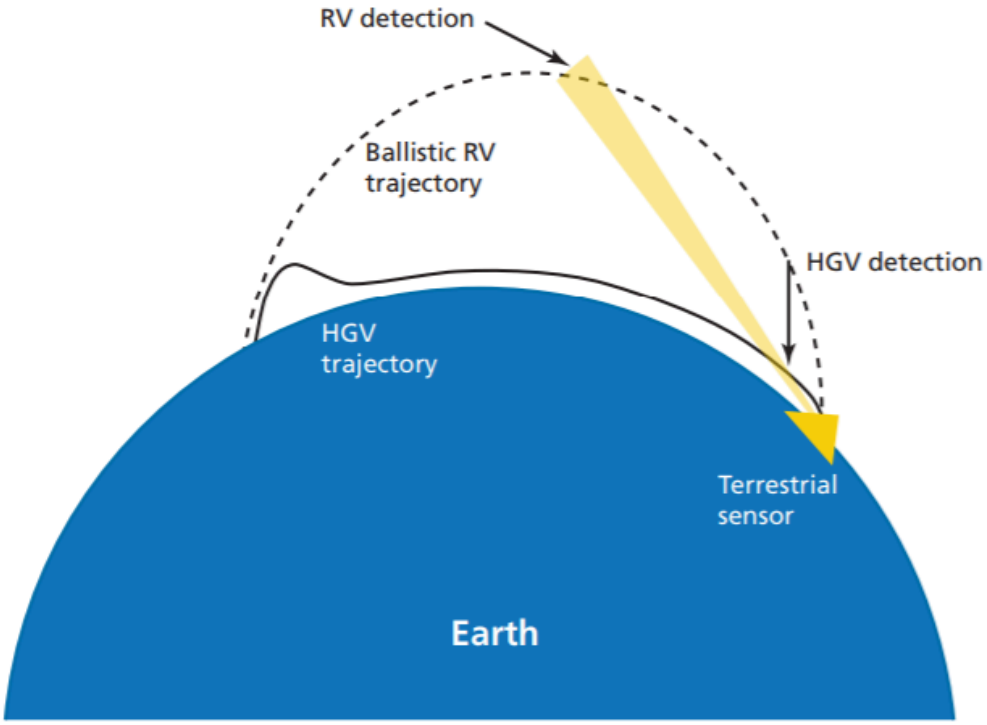
[11] Similarly, SIPRI (2022) states:

> "Wikipedia [...] defines 'hypersonic flight' as 'flight through the atmosphere below about 90 km at speeds ranging between Mach 5–10, a speed where dissociation of air begins to become significant and high heat loads exist.' [...]
>
> The US-based Missile Defense Advocacy Alliance states that 'hypersonic weapons refer to weapons that travel faster than Mach 5 (~3800mph) and have the capability to maneuver during the entire flight.'"

not travel faster than a ballistic reentry vehicle. However, it would be more difficult to predict the intended target and to direct missile defense interceptors toward the attacking HGV."[12]

The points about flatter flight paths and later detection are illustrated in the following diagram from a RAND report cited in Robichaud (2022), in which "RV" refers to regular ICBMs (RV is an abbreviation for "reentry vehicle"):

Figure 2.2
HGV Versus RV Terrestrial-Based Detection



Partyard Military (2019) states:

"Hypersonic missiles come in two variants; hypersonic cruise missiles and hypersonic glide vehicles.

What is a hypersonic cruise missile? This type of missile reaches its target with the help of a high-speed jet engine that allows it to travel at extreme speeds, in excess of Mach-5. It is non-ballistic – the opposite of a traditional Intercontinental Ballistic Missiles (ICBM) which utilizes gravitational forces to reach its target.

[12] Similarly, Sherman (2022) states:

"Hypersonic glide vehicles exploit blind spots in today's radar networks… The Pentagon can detect the launch—but the hypersonic glide vehicle then slips out of view until late in the weapon's flight because of ground radar's line-of-sight limitations. As a result, defensive systems have little, if any, time left to halt an incoming weapon."

What is a hypersonic glide vehicle? This type of hypersonic missile utilizes re-entry vehicles. Initially, the missile is launched into space on an arching trajectory, where the warheads are released and fall towards the atmosphere at hypersonic speeds. Rather than leaving the payload at the mercy of gravitational forces – as is the case for traditional ICBMs – the warheads are attached to a glide vehicle which re-enters the atmosphere, and through its aerodynamic shape it can ride the shockwaves generated by its own lift as it breaches the speed of sound, giving it enough speed to overcome existing missile defense systems. The glide vehicle surfs on the atmosphere between 40-100km in altitude and reaches its destination by leveraging aerodynamic forces."

What developments might occur?

What developments seem *in theory possible* and notable

- Increased range
- Increased stealth
- Increased accuracy
- Increased yield
- Reduced cost or time required for production

Statements from analysts

Chen (2022) writes:

> "These countries [China, the US and Russia] have successfully test-fired hypersonic missiles and a few [of these hypersonic missiles] have already entered military service.
>
> France, Japan, North Korea, South Korea, Australia and India have also launched hypersonic weapon programmes and some test flights have been conducted."

Fit for Purpose? The U.S. Strategic Posture in 2030 and Beyond states:

> "By 2030, the United States, Russia, and China are likely to have deployed hypersonic strike weapons, potentially in significant numbers. While a few of these may be intercontinental in reach, the majority are likely to be medium- and intermediate-range and designed for theater-strategic functions. These new theater systems raise important questions about offense/defense and conventional/nuclear integration; accordingly, their fit with the regional deterrence architectures of the United States and its allies and with extended U.S. nuclear deterrence remains an open question. They also raise difficult new questions about how to protect strategic stability as competition intensifies."

A 2020 Atlantic Council report recommends:

> "To keep pace with Russian developments in hypersonics, the United States and its allies should continue to develop their own offensive hypersonic capabilities, in the form of both HGV and cruise missiles. The United States could also invest in counter-hypersonic and cruise-missile defenses as part of a deterrence-by-denial strategy against Russia's new hypersonic and cruise missiles. [...]

> Some analysts may be concerned that such measures may cause an arms race or threaten strategic stability. Moreover, adding new nuclear systems on top of existing modernization plans may stretch the existing national nuclear enterprise too thin. Certainly, a strategy of simply mirroring Russia's nuclear behavior does not make sense, and a wide range of effective actions is available outside the nuclear domain. But, allowing a revisionist state to achieve meaningful military advantages could be highly destabilizing to regional and global security."

Why might this increase nuclear risk?

Hypersonic missiles / HGVs could perhaps allow for a "splendid" first strike, undermining deterrence.

- On the other hand, at least one analyst argues that new missile detection technology might prevent that:
  - "Hypersonic Weapons Can't Hide from New Eyes in Space: Tracking the missiles is like picking out one light bulb against a background of light bulbs, but new technology aims to see them more clearly" (Sherman, 2022).
- Another paragraph in Sherman (2022) provides some support for the claim that *unless and until* that better missile detection technology arises, hypersonic missiles/glide vehicles really do pose a threat to deterrence stability: "But even as senior U.S. military officials publicly fretted about **missiles that are, for the moment at least, effectively invincible**, the Pentagon was quietly making strides on an entirely novel way to help shoot down these weapons." (emphasis added)

Fear of false negatives: AI and China's nuclear posture states:

> "If China enhances its development of cruise missiles and hypersonic glide platforms by applying AI and autonomy, close-range encounters off the coast of Taiwan and in the East and South China Seas could grow even more complicated. [...] Moreover, China has hedged on what kind of payload will be carried by hypersonic glide platforms such as the DF-ZF, which are designed to break through missile defenses. With the release of the 2018 Nuclear Posture Review and Vladimir Putin's subsequent declaration that Russia has developed new nuclear weapons, the United States and Russia have engaged in a game of tit-for-tat. If China follows suit, a new set of destabilizing variables could be introduced into a region that is already tense and crowded, with freedom-of-navigation operations carried out among competing territorial claims."

See also What's the big deal about hypersonic missiles?

Additional notes

NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in hypersonic weapons.

See AI-assisted maneuverability and precision guidance for missiles/vehicles for some notes on the possibility and consequences of China "applying AI and autonomy" to its "hypersonic glide platforms".

It seems plausible that the combination of (a) "nuclear entanglement" and (b) advances in missile defense systems, (conventional) hypersonic missiles, and/or long-range conventional prompt strike missiles will increase the chance of nuclear conflict.

- Specifically, we'd guess that some current or plausible missile defense systems would be able to defend against either nuclear or non-nuclear missiles, and so could be attacked to gain a conventional advantage but with this increasing the chance of nuclear retaliation.
- And we'd guess the chance of this could be increased by development and deployment of more numerous and capable conventional missiles, such as hypersonic missiles or conventional long-range prompt strike missiles.
- But we haven't looked into these possibilities at all.

More accurate nuclear weapons

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium	Medium/Low

Preamble

- Relative to most other possible technological developments discussed in this post, this development seems unusually likely to also or primarily have risk-*reducing* effects. **In fact, Michael's current low-confidence guess would be that this development would overall reduce risk and that we should actively consider trying to *increase* the odds/extent to which this development occurs.** As such, we discuss some potential risk-reducing effects here, even though that's not the focus of this post.
- Developments in this area would presumably in most or all cases also allow for more accurate *non-nuclear* weapons, since the same delivery systems can often be used for multiple warhead types. One implication is that developments in this area, and their effects, may be influenced by developments in long-range conventional strike capabilities (see the next section) and vice versa.

Why might this increase nuclear risk?

More accurate nuclear weapons could perhaps reduce the stability of the existing deterrence situation.

- This is because, if a state thinks their weapons might be accurate enough to completely (or close to completely) wipe out an adversary's arsenal in a first strike, and if the state thinks their adversary has similar capabilities, then there may be an incentive to be the

first mover (i.e., to launch a preemptive decapitation strike against the adversary, so that they can't do this to you).

- In line with this, the article China's Military Calls for Putting Its Nuclear Forces on Alert: states: "These and other statements suggest that a domestic conversation about raising the alert level of China's nuclear forces is taking place. The debate is driven in part by concerns about accurate U.S. nuclear weapons, high-precision conventional weapons, and missile defense—all of which are perceived as compromising China's current posture."
    - However, no clear justification is provided for the claim in that second sentence, so we aren't confident it's accurate.

Why might this increase *or decrease* nuclear risk?

A 2020 CSIS Briefs report considers developments in nuclear weapon accuracy/precision, and how this might impact nuclear risk:

> "Advocates of enhanced accuracy—such as Stephen M. Younger, the former director of the Sandia National Laboratories and former associate director of Los Alamos National Laboratory's nuclear weapons research and development division—have argued that 'a sizable factor governing the explosive force required to defeat a target of given hardness is the precision with which weapons can be delivered.' **A more precise weapon can lower the number of warheads and/or the explosive yield required per target and may reduce collateral damage from a limited nuclear strike. Opponents of enhanced precision argue that coupling enhanced accuracy and precision with reduced collateral damage could make nuclear options more palatable or encourage the development of first-strike counterforce options.** Complicating the ability to assess new or enhanced warhead capabilities are the tradeoffs between yield and precision that together determine the warhead's overall capability. For example, is a warhead that increases precision and decreases yield more capable? What about the reverse?" (emphasis added)

Why might this *decrease* nuclear risk?

Younger (2000) writes:

> "Some targets require the energy of a nuclear weapon for their destruction. However, **precision targeting can greatly reduce the nuclear yield required to destroy such targets**. Only a relatively few targets require high nuclear yields. Advantages of lower yields include reduced collateral damage, arms control advantages to the United States, and the possibility that such weapons could be maintained with higher confidence and at lower cost than our current nuclear arsenal." (emphasis added)

Gentzel (2021) writes:

> "Massive yield weapons aren't just big explosions: they produce more radioactive nuclear fallout that can spread over a thousand miles and kill people for years, larger (though less efficient) electromagnetic pulses that can disable electronic grids and large

electronic devices over millions of square miles, and mushroom clouds that can reach the stratosphere, contributing to nuclear winter.

With the proliferation of sensors, precision weapons, and fusion of information by narrow artificial intelligence, giant weapon yields may no longer [be] necessary to assure deterrence. In my view, this presents the opportunity [to] reduce the risk of nuclear winter, but how would you achieve that?

[...] Overall, my argument is that nuclear winter is uncertain with current arsenals and targeting plans, but that it could be made extremely unlikely without getting rid of nukes, or even shifting to very small numbers of nukes. Countries with lots of weapons are more likely to employ counterforce targeting to limit damage in the event of nuclear war (aiming at military forces, command and control, etc.) while if you have fewer weapons you are more likely to do counter value targeting for pure deterrence value: neglecting precision, increasing weapon yield, and targeting cities (how you produce nuclear winter). Low-yield weapons are still orders of magnitude worse than conventional weapons and can provide plenty of deterring power. If you thought current nuclear deterrence was insufficient, would your solution be to replace all warheads with Tsar Bombs? Probably not, you could increase the yield of some weapons selectively if you have intelligence problems, or instead get better intelligence, increase precision, and get more low-yield weapons.

[...] Bringing this all together, **I think a good path for nuclear modernization would be to generally reduce nuclear weapon yields while increasing precision: this makes the weapons more credible that you will use them, and enhances deterrence in that manner while decreasing the odds of global nuclear winter if something ever goes wrong somehow.** For states with small numbers of nuclear weapons, this creates a far more credible threat of counterforce nuclear targeting: disincentivizing proliferation, while states that can deploy many nuclear weapons would become far more difficult to target despite the proliferation of long-range precision conventional weapons and sensors. I don't think smaller weapons do any less good of a job at deterring decision makers: at point blank range these weapons produce absurdly high overpressure that will crush any bunker, removing the need for extreme yield weapons to take out hardened bunkers while missing by hundreds of meters. I think it is better if deterrence shifts to deterring decision makers and militaries rather than inherently threatening entire societies. I don't hold these ideas with high certainty, but they should at least be debated, and if wrong, thoroughly debunked." (emphasis added)

Additional notes

Geist and Lohn (2018) write:

"Even with perfect knowledge of the target location, mobile targets can move between the time a weapon is launched and the time it arrives. Weapons for targeting mobile systems might be able to fly faster and adjust course better, but weapons would still need extremely sophisticated terminal guidance capabilities to substantially reduce the

amount of ordnance required. As a result, even with advances in image processing and target recognition, many large weapons would be needed, or smaller ones would need to be launched from close range. The figure on page 17 shows the number of warheads of various types that would be required to destroy a mobile target with a weapon radius of effect between 0 and 5 kilometers. Despite their huge "kill radius" measuring kilometers in diameter, multiple thermonuclear warheads delivered by ballistic missiles would be required to have a high assurance of destroying a missile launcher. For instance, three 475-kT W88 warheads delivered by Trident II missiles with a ten minute flight time would be required to cover one target, while five 100-kT W76 warheads would be necessary to cover it. The analysis finds, however, that accurate cruise missiles (CMs) launched from a position close to the targets (30-kT CM and 200-kT CM in the figure) could cover the mobile missile launchers with only one or two warheads. Fired from very close distances (i.e., flight times of a few minutes), even conventional munitions could become viable options, thereby significantly increasing the credibility of preemptive counterforce strikes."

See also The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence.

Long-range conventional strike capabilities

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/Low	Medium/High	Medium/Low

Preamble

- As with more accurate nuclear weapons, this technological development seems decently likely to partly or primarily *reduce* nuclear risk, so we include some discussion of the possible risk-reduction effects even though that's not the focus of this post.
- Developments in long-range conventional strike capabilities are related to both hypersonic missiles/glide vehicles and more accurate nuclear weapons, discussed in the previous two sections. As such, some of the information in those sections is relevant here as well (but we won't repeat it here).
- In this section, most of the information on what this topic is and what developments might occur is embedded in the subsections on why it might increase or decrease nuclear risk.

What this is

Wikipedia:

> "Prompt Global Strike (PGS) is a United States military effort to develop a system that can deliver a precision-guided conventional weapon airstrike anywhere in the world within one hour, **in a similar manner to a nuclear ICBM**."

Baum (2015a):

> "Conventional prompt global strike is a new technology that pairs high-speed, high-accuracy missiles with conventional bombs. The aim is to be able to strike anywhere in the world within one hour. Historically, only nuclear weapons could achieve this, but improvements in missile accuracy are making it possible to do the same with much smaller weapons. This new technology has already gained some consideration as a deterrent, including in recent remarks by Joseph Biden and Vladimir Putin. Because the conventional bombs are smaller, they are more suitable for threatening select military infrastructure, political leadership, or other small highvalue targets, especially in time-sensitive situations such as the transfer of weapons of mass destruction. They thus may only be capable of deterring smaller aggressions and can likewise only be one component of a broader deterrence regime."

Why might this increase nuclear risk?

Advances in this capability *might* increase nuclear risk, due to one or more of the following effects:

1. raising the risks of inadvertent nuclear war (via a conventional strike being misperceived as a nuclear strike)
2. undermining deterrence
3. undermining arms control/reduction efforts
4. increasing geopolitical tensions
5. motivating an increase in numbers of tactical nuclear weapons, which *might* make nuclear conflict more likely (though also might make it smaller in expectation)
6. motivating an increase in arsenal sizes more broadly

These potential effects are discussed in the following quotes.

Baum (2015a):

> "A major downside of conventional prompt global strike is its resemblance to nuclear missiles. **A conventional strike could look the same as a nuclear strike to another country's radar systems, prompting that country to believe it is under nuclear attack and respond in kind. This raises the risk of inadvertent nuclear war from false alarms mistaken as real.** One proposed solution is to reserve ballistic missile trajectories for nuclear missiles and to only use guided trajectories for conventional prompt global strike. Another is to disclose the situations in which each type of missile would be used. Such solutions can help, but the risk of inadvertent nuclear war would still be at least somewhat higher than it would be without any conventional prompt global

strike. That said, the entire issue is avoided if nuclear weapons are no longer in use." (emphasis added)

Baum (2015a):

> "**Conventional prompt global strike is also worrisome in its capability as a first-strike weapon, potentially capable of disabling the other side's deterrents, resulting in a destabilizing first-mover advantage.** Their first-strike potential at the heart of China's and Russia's concern about the American system under development: It someday may be able to knock out most of Russia's nuclear arsenal, perhaps with the rest of it being 'mopped up' by a missile defense system, thereby negating China's and Russia's deterrent and giving NATO a first-strike advantage, all without the harms and stigmas associated with the use of nuclear weapons. While such concerns are inconsistent with the present American/NATO plans, a large arsenal conventional prompt global strike arsenal—perhaps thousands of weapons—may someday be able to achieve this. This type of concern is important to address as conventional prompt global strike systems start to go live. **Indeed, imbalances in conventional prompt global strike capabilities between countries could even be an impediment to nuclear disarmament.**" (emphasis added)

As noted in the previous section, the article China's Military Calls for Putting Its Nuclear Forces on Alert states: "These and other statements suggest that a domestic conversation about raising the alert level of China's nuclear forces is taking place. The debate is driven in part by concerns about accurate U.S. nuclear weapons, high-precision conventional weapons, and missile defense—all of which are perceived as compromising China's current posture."

- However, no clear justification is provided for the claim in that second sentence, so we aren't confident it's accurate.

The article Fear of false negatives: AI and China's nuclear posture states:

> "**Officials in Beijing (link in Chinese) and Moscow have responded strongly and negatively to the 2018 US Nuclear Posture Review.** But a number of the changes specified in Washington's 2018 review actually grew out of the 2010 version. For example, **the 2010 review diminished the role of nuclear weapons in US national security strategy, but it placed a strong emphasis on ballistic missile defense and conventional prompt global strike. This emphasis has been uniformly condemned in both Beijing and Moscow**—two capitals that were linked in the document. While Chinese analysts may have lauded Washington's shift away from nuclear weapons in 2010, **they still viewed the review's focus on ballistic missile defense and conventional prompt global strike as preemptive and destabilizing**. **These US systems were seen as negating the strategic leverage of militaries whose conventional arsenals are weaker.**
>
> [...] For Beijing, which has been expanding its nuclear arsenal at a relatively modest pace, the prospect of the United States resuming a forward-deployed, tactical nuclear posture exacerbates its sense of encirclement. Such a posture also amplifies **China's**

> **perceived and real vulnerability to US ambitions to field kinetic and surveillance platforms such as prompt global strike, X-variant space planes, the surveillance aircraft Global Hawk, and so on."** (emphasis added)

Kristensen & Korda (2021) seem to think that current "expansion of US long-range conventional strike capabilities" **could trigger Chinese interest in nonstrategic nuclear weapons, and thus could perhaps *increase* the probability of nuclear war** (and also perhaps increase the severity of war if war does occur). They write:

> "The [US] pursuit of a new nuclear [submarine-launched cruise missile] to 'provide a needed nonstrategic regional presence' in Europe and Asia could increase Russia's reliance on nonstrategic nuclear weapons and could potentially even trigger Chinese interest in such a capability as well—**especially when combined with the parallel expansion of US long-range conventional strike capabilities including development of new conventional INF-range missiles**." (emphasis added)

Why might this *reduce* nuclear risk?

In contrast, Baum et al. (2018) implies that part of the rationale for conventional prompt global strike is that **this could reduce the role of nuclear weapons, and that it's at least plausible that this capability would *reduce* the probability of nuclear war**:

> "How aggressively should states pursue alternative weapons such as conventional prompt global strike? Some nuclear-armed states are developing new weapons technologies in order to reduce the role of nuclear weapons. One example is the **United States' development of conventional prompt global strike, in which high-precision conventional weapons hit distant targets that previously only nuclear weapons could hit**. These new weapons programs can be expensive. To the extent that these weapons reduce the probability of nuclear war, the more likely nuclear war is to occur without these new weapons, the more worthwhile investments in them are." (emphasis added)

Relatedly, Baum (2015b) states:

> "US nuclear doctrine states that the US 'would only consider the use of nuclear weapons in extreme circumstances to defend the vital interests of the United States or its allies and partners' (US DoD 2010:ix). The US has been steadily shifting military capabilities from nuclear weapons towards advanced high precision conventional weapons ("prompt global strike") and missile defense systems, which permit the US to reduce the role of nuclear weapons. Other countries with nuclear weapons may follow in this direction but lag in these technologies. With an eye to the future, these changes in weapon systems could render nuclear weapons irrelevant, permitting radical shifts in nuclear doctrine. Such shifts may even already be possible. Noting the military preference for precision weapons with minimal collateral damage, Wilson (2012:27) writes that 'Increasingly,

> nuclear weapons look like dinosaurs: really large and frightening creatures that were destined to die out because they could not adapt.'"

Additional notes

NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in "conventional prompt strike".

The Geist and Lohn (2018) quote in the previous section is also relevant here.

The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence states:

> "If nuclear forces are becoming increasingly vulnerable to counterforce, then states need to improve their retaliatory arsenals just to maintain the same level of deterrence. Given that nuclear delivery systems are expensive and must last for decades, the challenge for force planners is extraordinary: deploy weapon systems that will remain survivable for multiple generations, even as technology improves at an ever-increasing pace. Second, **the growing threat to nuclear arsenals (from** nuclear strikes, **conventional attacks,** missile defenses, ASW [anti-satellite weapons], and cyber operations) **raises major questions about the wisdom of cutting the size of nuclear arsenals.** In the past, many arms control advocates believed that arms cuts reduced the incentives for disarming strikes; whether right or wrong in the past, that assumption is increasingly dubious as a recipe for deterrence stability today."

Younger (2000) writes:

> "Advances in conventional weapons technology suggest that, by 2020, precision long-range conventional weapons may be capable of performing some of the missions currently assigned to nuclear weapons. Today, uncertainty in the location of road mobile missiles carrying weapons of mass destruction might require a nuclear weapon for assured destruction. Future real-time imagery and battle management, combined with precision strike long-range missiles, may mean that a conventional weapon could effectively destroy such targets."

It seems plausible that the combination of (a) "nuclear entanglement" and (b) advances in missile defense systems, (conventional) hypersonic missiles, and/or long-range conventional prompt strike missiles will increase the chance of nuclear conflict.

- Specifically, we'd guess that some current or plausible missile defense systems would be able to defend against either nuclear or non-nuclear missiles, and so could be attacked to gain a conventional advantage but with this increasing the chance of nuclear retaliation.
- And we'd guess the chance of this could be increased by development and deployment of more numerous and capable conventional missiles, such as hypersonic missiles or conventional long-range prompt strike missiles.
- But we haven't looked into these possibilities at all.

Detection and defense

Better detection of nuclear warhead platforms, launchers, and/or delivery vehicles

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/High	Medium/High	Medium/Low

Preamble

- This topic is closely related to possible nuclear-risk-relevant advances in AI, which are also discussed separately later in this post.
- The riskiness of various developments in this area would depend on (among other things) advances in stealth technologies and changes in arsenal sizes and compositions.

What developments might occur, and why might this increase nuclear risk?

Sea

A potentially *very* destabilizing development would be better means of detecting ballistic missile submarines (SSBNs).

- One possibility would be advances in "remote sensing".
    - See Wren & May (1997)
- Brixey-Williams (2020) and NTI (2021) cover remote sensing plus other advances in submarine detection and monitoring.
- Regarding detection of SSBNs, Rodriguez (2019) writes:

> "**US submarine-launched ballistic missiles (SLBMs), housed on its strategic nuclear submarines (SSBNs) are considered the most survivable leg of the nuclear triad. This is because they are *extremely* hard to detect.** Historically, submarines could only be detected using radar and acoustic technologies, like sonar (Deutsche Welle, 2017). Satellites can't see SSBNs through the ocean, nor can they be heard using acoustic technologies if they're quiet enough. Besides reducing light and sound that travels through water, the salinity, temperature, and pressure of the ocean also *bend* sound waves, reducing the accuracy of sonar (Holmes, 2016, p. 228). Further, US submarine countermeasures are among the best in the world. Its submarines' surfaces are covered in plastic, which disperses radar signals instead of reflecting them. They are also quieter than any other country's subs (Clark, 2015, p. 1).

But technological developments may finally be catching up to stealth technologies. Crucially, the US relies on being able to successfully *conceal* its SSBNs, but increasingly, the survivability of its SLBMs is threatened by trends in technology that make it easier to track SSBNs down. **One area where things are evolving rapidly is remote sensing** (Lieber & Press, 2017a, p. 32).

New analytic techniques are making remote sensing increasingly accurate. For example, new sensors can now pick up fainter signals than before, and they can more successfully distinguish between signals from the target and ambient noise in the ocean (Lieber & Press 2017a). **Additionally, new acoustic techniques, improved signal processing techniques, and advancements in machine learning detection algorithms make it easier to detect a submarine by enabling a comparison between the expected ocean noise (for example waves and marine life) to the 'noise fields' that are actually being measured (Clark 2015).** Better computer modeling can enhance information about a target, not unlike the techniques used today to enhance photographic images (Lieber & Press 2017a; Clark 2015). **New detection techniques that are non-acoustic (such as lasers and light emitting diodes (LEDs)) could potentially provide more accurate intelligence and are expected to become more prominent in the coming years (Clark, 2015).**

Further, sensor platforms have become more diverse, and they can increasingly provide continuous intelligence and information about their targets (Lieber & Press, 2017a). This makes the SSBNs more vulnerable because their movement can be tracked over time, making it possible for an adversary to identify patterns in their routes. The older sensor platforms (satellites, submarines, and piloted aircraft) are supplemented by new, less vulnerable sensor platforms. For example, unmanned aircraft that are piloted from a distance and drones that go underwater decrease the survivability of SLBMs while posing minimal risk to the security 'seeker' (Lieber & Press 2017a).

Finally, intelligence data can be transmitted almost instantaneously, making it possible for the 'seeker' to make timely and informed decisions based on up-to-date information about the whereabouts of the SSBNs (Lieber & Press 2017a), making the SLBMs more vulnerable than ever before.

If Russia beat the US to incorporating these technologies into its intelligence gathering systems, it could gain a decisive advantage. According to Lieber & Press (2017a), the US would be hard-pressed to develop countermeasures quickly and effectively. Thus, the assumption that the sea-based leg of the triad is inherently invulnerable might not hold much longer.

But while the emerging technologies could eventually be game-changing, it has not yet become easy to track down submarines. Until that changes, I

> **expect basically all of the US's SSBNs and their nuclear cargo would survive a counterforce first strike.**" (emphasis added)

Land and air

Better means of detecting mobile launchers or mobile ICBMs (i.e., those that travel by road or rail) could be quite destabilizing.

- Pages 16-18 of Geist and Lohn (2018) discuss how AI could potentially undermine secure second strike by rendering mobile launchers trackable and targetable and thus not survivable.
- Ladish (2019) writes:
  - "Reliable launch detection capability and missile tracking is a good thing for global stability. Uncertainty in this domain creates more pressure to act on short timescales, which could be detrimental in a crisis.
  - [...] Advances in machine learning have the potential to both improve and weaken stealth tech. They will improve autonomous capabilities of unmanned stealth vehicles, but also improve signal processing and detection capabilities."

Additional notes

The article Fear of false negatives: AI and China's nuclear posture states:

> "For Beijing, which has been expanding its nuclear arsenal at a relatively modest pace, the prospect of the United States resuming a forward-deployed, tactical nuclear posture exacerbates its sense of encirclement. Such a posture also amplifies **China's perceived and real vulnerability to US ambitions to field kinetic and surveillance platforms such as prompt global strike, X-variant space planes, the surveillance aircraft Global Hawk, and so on.**"

See also The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence.

Some quotes from Geist and Lohn (2018):

> "The effect of AI on nuclear strategy depends as much or more on adversaries' perceptions of its capabilities as on what it can actually do. For instance, it is extremely technically challenging for a state to develop the ability to locate and target all enemy nuclear-weapon launchers, but such an ability also yields an immense strategic advantage. States therefore covet this capability and might pursue it irrespective of technical difficulties and the potential to alarm rivals and increase the likelihood of conflict. The case could be made on technical grounds that advanced AI would still struggle to overcome obstacles originating from data limitations and information-theoretic arguments, but the tracking and targeting system needs only to be *perceived* as capable to be destabilizing. A capability that is nearly effective might be even more dangerous than one that already works."

“The first discussion focused on the tracking and targeting problem and asked participants to consider how they would try to thwart an adversary seeking to render strategic forces vulnerable using AI. Participants suggested trying to neutralize this capability by attacking the associated sensors and communications network rather than the AI itself.”

“The increasingly multipolar strategic environment is also encouraging forms of competition that threaten stability. For instance, the United States is interested in developing the capability to track and target a minor nuclear power’s mobile missile launchers, but Russia and China fear that the same technology could mature into a threat to their more sophisticated retaliatory forces. In a crisis situation, the employment or availability of AI-enabled intelligence, surveillance, and reconnaissance (ISR) or weapon systems could stoke tensions and increase the chances of inadvertent escalation. Finally, the pursuit of advanced military capabilities is liable to cause arms race instability even if those technologies are nonviable, as in the historical case of missile defense.”

“Both Russia and China appear to believe that the United States is attempting to leverage AI to threaten the survivability of their strategic nuclear forces, stoking mutual distrust that could prove catastrophic in a crisis. As Paul Bracken observes, ongoing improvements in technology such as AI threaten to ‘undermine minimum deterrence strategies’ and ‘blur the line between conventional and nuclear war’ (Bracken, 2017).”

“The United States, meanwhile, explored the possibility that AI could be used to bolster its counterforce capability. One late 1980s research project, the Survivable Adaptive Planning Experiment (SAPE), sought to use the AI technology of the time to enable the United States to target the Soviet Union’s mobile ICBM launchers. The SAPE would not control nuclear weapons directly; rather, it would employ expert systems to translate reconnaissance data into nuclear targeting plans that would then be carried out by manned B-2 bombers. The SAPE was just one part of an envisioned suite of systems and capabilities that, if actualized, would have severely challenged the survivability of the Soviet Union’s nuclear arsenal (Roland and Shiman, 2002, p. 305; Long and Green, 2012).”

“AI technologies could help enable new breakthroughs in tracking and targeting and in antisubmarine warfare or make it easier for high-precision conventional munitions to destroy hardened ICBM silos (Holmes, 2016). Such capabilities would be especially destabilizing because decisionmakers could threaten to employ conventional weapons much more plausibly than any kind of nuclear attack. A conventional threat would place the adversary under enormous pressure during a crisis, which could force it to capitulate—but could also spiral into nuclear war. Such escalation could happen because the adversary felt the need to use its weapons before being disarmed, in retaliation for an unsuccessful disarming strike, or simply because the crisis triggered accidental use.”

> “A major challenge of nuclear strategy is that adversaries may interpret one nation’s secure retaliatory forces as a first-strike threat or a doomsday machine and react accordingly. [...] AI progress is also contributing to Russia’s doubling down on older types of systems with undesirable strategic properties. For example, with its RS-28 ‘Sarmat’ missile, Russia is reinvesting in large, silo-based ICBMs with multiple independently targetable reentry vehicle (MIRV) warheads, a category of weapon it once planned to abandon under the now-defunct Strategic Arms Reduction Talks II treaty. Western strategic theory generally considers large MIRVed ICBMs to be destabilizing because they are ideal for preemptive strikes and are vulnerable to preemption. At the dawn of the millennium, Moscow believed that it could ensure the survivability of its forces by emphasizing mobile ICBMs and scrapping large silo-based missiles inherited from the Soviet Union. However, Russian leaders’ anxieties about potential U.S. threats to the survivability of the mobile ICBMs seem to have changed this calculus and led them to try to ensure retaliation by launching during a U.S. attack instead of riding it out. This is tantamount to the adoption of a launch-under-attack posture that could place great pressure on Russian leaders to launch first in a crisis, increasing the chances of accidental escalation.”

> “In numerous convincing demonstrations, small amounts of adversarial effort toward subverting machine learning algorithms have shown outsized effect. Some researchers argue that this is a pervasive trait of machine learning and that they expect that it will persist for years to come. Where an effective AI for tracking and targeting might be destabilizing and lead to proliferation or worse, an adversary may regain trust in the survivability of its second-strike forces if it is confident in its ability to forestall detection using these adversarial methods, thereby reestablishing strategic stability. On the other hand, an actor may believe that it can subvert an AI’s ability to identify a preemptive first strike, making such a strike a viable option and therefore destabilizing”

Missile defense systems

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/Low	Medium/High	Medium/Low

What developments might occur, and why might this increase *or reduce* nuclear risk?

The article Fit for Purpose? The U.S. Strategic Posture in 2030 and Beyond states:

> “By 2030, the Ground-based Mid-Course Defense (GMD) system protecting the U.S. homeland against long-range missile attack is likely to have been improved incrementally, with the addition of the planned 20 new advanced interceptors to the existing fleet of 44 Ground-Based Interceptors (GBIs). It may also have been reinforced

> with an underlayer of shorter-range Aegis and THAAD systems. Some modest tailoring of the homeland defense to address the hypersonic threat with new sensors and interceptors is also likely. Some modest improvements to regional defenses may also occur, with possible new roles for directed energy weapons. By 2040, directed energy weapons and a significant move to space may have revolutionized the homeland defense.
>
> [...] The further development of U.S. missile defenses may be driven by a simple calculus of technology and money: that is, "we should have as much of the best available defensive technology as we can afford." A strategy-driven approach is more challenging to define. A central feature of strategy since 1999—to seek protection of the homeland from limited missile strikes—was recently set aside in law. The new push for the protection of the American homeland from larger scale strikes brings with it significant new questions about how much missile defense is enough. One is whether to compete with new developments in the missile postures of Russia and China. Another is whether to compete with rogue state forces if and as they gain the ability to conduct the larger-scale strikes that the United States has heretofore seen as beyond the scope of missile defense (because such strikes have been seen as deterrable by the threat of retaliation)."

However, Ellsberg (2018) says:

> "**There never will be an anti-ballistic missile system that actually protects us from a large scale attack by Russian ICBMs**, accompanied by decoys and other evasive measures.
>
> [...] It is as infeasible as a highly effective anti-ballistic missile system." (emphasis added)

Baum (2015b) states:

> "Nuclear weapon systems can also influence the probability of intentional nuclear war, which occurs when both sides accurately understand the state of affairs and decide to launch nuclear weapons. In a crisis, a country whose nuclear weapons are vulnerable to attack may be more likely to launch its weapons first, a situation of "use it or lose it". For example, some recent scholarship has suggested that the US could take out Russia's and China's nuclear arsenals in a first strike attack (Lieber and Press 2006). **New US missile defense systems raise the additional concern that, even if a first strike would leave some weapons intact, retaliation would fail** (Steff 2013). Systems that protect nuclear weapons from being taken out, such as mobile missile launchers, nuclear submarines, and decoys to thwart missile defense, make it easier for countries to hold off on launch decisions during crises." (emphasis added)

But the same paper also notes:

> "US nuclear doctrine states that the US 'would only consider the use of nuclear weapons in extreme circumstances to defend the vital interests of the United States or its allies and partners' (US DoD 2010:ix). **The US has been steadily shifting military**

> **capabilities from nuclear weapons towards advanced high precision conventional weapons ('prompt global strike') and missile defense systems, which permit the US to reduce the role of nuclear weapons**. **Other countries with nuclear weapons may follow in this direction but lag in these technologies.** With an eye to the future, these changes in weapon systems could render nuclear weapons irrelevant, permitting radical shifts in nuclear doctrine. Such shifts may even already be possible. Noting the military preference for precision weapons with minimal collateral damage, Wilson (2012:27) writes that 'Increasingly, nuclear weapons look like dinosaurs: really large and frightening creatures that were destined to die out because they could not adapt.'"

Risks from the UK's planned increase in nuclear warheads (2021) states:

> "In March 2021 the [UK] Integrated review announced details of where this spending would be spent, including an increase of the number of nuclear warheads from an estimated 185, of which 120 are operational, to a cap of 260. The increase scraps a 2010 government plan to "reduce our total stockpile to no more than 180 by the mid 2020s". The reasons for the reversal are unclear but seems to reflect fears of growing Chinese and Russian military capabilities, with [the] **UK secretary of defence specifically mentioning Russian investments in ballistic nuclear missile defence**. The increase in both arms funding and nuclear warheads seem to align with the UK prime minister Boris Johnson's vision for a post-Brexit 'Global Britain' to be a major player on the world stage." (emphasis added)

As noted in an earlier section, the article China's Military Calls for Putting Its Nuclear Forces on Alert: states: "These and other statements suggest that a domestic conversation about raising the alert level of China's nuclear forces is taking place. The debate is driven in part by **concerns about** accurate U.S. nuclear weapons, high-precision conventional weapons, and **missile defense—all of which are perceived as compromising China's current posture**."

- However, no clear justification is provided for the claim in that second sentence, so we aren't confident it's accurate.

Similarly, as noted in an earlier section, the article Fear of false negatives: AI and China's nuclear posture states:

> "**Officials in Beijing (link in Chinese) and Moscow have responded strongly and negatively to the 2018 US Nuclear Posture Review.** But a number of the changes specified in Washington's 2018 review actually grew out of the 2010 version. For example, **the 2010 review diminished the role of nuclear weapons in US national security strategy, but it placed a strong emphasis on ballistic missile defense and conventional prompt global strike. This emphasis has been uniformly condemned in both Beijing and Moscow**—two capitals that were linked in the document. While Chinese analysts may have lauded Washington's shift away from nuclear weapons in 2010, **they still viewed the review's focus on ballistic missile defense and conventional prompt global strike as preemptive and destabilizing**. **These US systems were seen as negating the strategic leverage of militaries whose conventional arsenals are weaker**." (emphasis added)

Arnold and Toner (2021) give the following hypothetical example of how (autonomously activated) missile defense could lead to catastrophe:

> "Phantom missile launches: In missile defense, seconds of delay can spell the difference between an interception and a miss. U.S. Strategic Command's new missile defense system, Global Eye, eliminates delay by scanning gigabytes of real-time data every second. If the system's algorithms detect a missile launch with high certainty, the system can quickly and autonomously trigger an interceptor launch in order to shoot down the missile. One day, unusual atmospheric conditions over the Bering Strait create an unusual glare on the horizon. Global Eye's visual processing algorithms interpret the glare as a series of missile launches, and the system fires interceptors in response. As the interceptors reach the stratosphere, China's early-warning radar picks them up. Believing they are under attack, Chinese commanders order a retaliatory strike."

On the other hand, the article Smaller and Safer states:

> "Our modeling found that the United States and Russia could limit their strategic nuclear arsenals to a total level of 1,000 warheads each on no more than 500 deployed launchers without weakening their respective security. De-alerting these forces actually helped stabilize deterrence at these and lower levels. And **the modeling showed that fairly extensive missile defense deployments would not upset this stability**."[13]

Additional notes

NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in missile defense.

See also The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence.

It seems plausible that the combination of (a) "nuclear entanglement" and (b) advances in missile defense systems, (conventional) hypersonic missiles, and/or long-range conventional prompt strike missiles will increase the chance of nuclear conflict.

- Specifically, we'd guess that some current or plausible missile defense systems would be able to defend against either nuclear or non-nuclear missiles, and so could be attacked to gain a conventional advantage but with this increasing the chance of nuclear retaliation.
- And we'd guess the chance of this could be increased by development and deployment of more numerous and capable conventional missiles, such as hypersonic missiles or conventional long-range prompt strike missiles.
- But we haven't looked into these possibilities at all.

[13] The article's section "Partners in defense" is also relevant to this topic.

AI and cyber

Note: Although this post is intended to mostly focus on possible *new technological developments* as opposed to *proliferation of existing technologies* or changes in *how those technologies are deployed*, this "AI and cyber" category ended up including discussion of all three of those things, sometimes without the distinctions being flagged explicitly. In particular, "More integration of AI with NC3 systems" might actually be more about proliferation or deployment of existing technologies than about new developments.

Advances in AI capabilities

This is an especially broad and heterogenous area, and intersects with many other areas discussed in this post.

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/High	Medium/High	Medium

What developments might occur, and why might they increase nuclear risk?

Below, we list some areas where AI advances could be important.

AI-assisted detection of nuclear warhead platforms, launchers, and/or delivery vehicles

AI-assisted nuclear-weapons-related production/design

Advances in autonomous weapons

Advances in AI systems that are integrated with NC3 systems

Better cyberattack capabilities

AI-assisted maneuverability and precision guidance for missiles/vehicles

- We haven't looked into this beyond reading the following quote, and we're not confident that this possibility is worth highlighting or that it's worth highlighting as distinct from advances in autonomous weapons.
- Fear of false negatives: AI and China's nuclear posture states:

  > "As China further develops its concept of "rapid response" (快速反应), as cited in its 2015 Military Strategy [中国的军事战略] (link in Chinese), Beijing's integration of AI and autonomy into its military systems is likely to increase. Such integration could range anywhere from automation-enabled launch-on-warning for its missiles to **autonomy- and AI-enabled maneuverability and precision**

> **guidance for hypersonic glide platforms and space planes.**
>
> [...] **If China enhances its development of cruise missiles and hypersonic glide platforms by applying AI and autonomy, close-range encounters off the coast of Taiwan and in the East and South China Seas could grow even more complicated**. China's ground-launched DH-10 missile is believed to carry a conventional warhead, but indications have emerged that the air-launched CJ-10 may have both nuclear and conventional variants. Moreover, China has hedged on what kind of payload will be carried by hypersonic glide platforms such as the DF-ZF, which are designed to break through missile defenses. With the release of the 2018 Nuclear Posture Review and Vladimir Putin's subsequent declaration that Russia has developed new nuclear weapons, the United States and Russia have engaged in a game of tit-for-tat. If China follows suit, a new set of destabilizing variables could be introduced into a region that is already tense and crowded, with freedom-of-navigation operations carried out among competing territorial claims."

Advances in any of a wide range of AI systems that could accelerate economic growth and/or technological/scientific progress

- This could in turn lead to other technological developments discussed in this post.
    - See also Differential progress.
- Separately from that, it could also increase proliferation, increase expected arsenal sizes, and undermine stability, for similar reasons to why advances in methods for nuclear-weapons-related production and design could.
- Shulman (2020) notes:
    - "[...] so with 1950s proportional military expenditures, half going to nukes, the US and China could each produce 20,000+ ICBMs, each of which could be fitted with MIRVs and several warheads, building up to millions of warheads over a decade or so; the numbers could be higher for cheaper delivery systems
>
> [...] **if AI and robotics greatly increase economic growth the above numbers could be increased by orders of magnitude**" (emphasis added)

More advanced AI of the sort that could adversarially optimize against humans by using nuclear weapons

- This is less commonly discussed than most of the other items on the list and we haven't thought much about it.
- We're therefore unsure whether it warrants attention, and we're unsure whether any attention it receives should be from the perspective of / under the banner of "nuclear risk" as opposed to "AI risk".
- But it does seem plausibly important.
- Some further discussion can be found here. One of us (Michael) also has some additional rough notes he could potentially share with interested people on request.

Additional notes relevant to this section as a whole

Boulanin (2020) state:

> "**Advances in AI can have an impact on strategic stability relations among nuclear-armed states even before they are fully developed, much less deployed. For example, a state may perceive that an adversary's investment in AI, even non-nuclear-related, could give that adversary the ability to threaten the state's future second-strike capability.** This could be sufficient to generate insecurity and lead that state to adopt measures that could decrease strategic stability and increase the risk of a nuclear conflict.
>
> [...] In this light, states and international organizations should take a number of measures—sequentially or simultaneously—to deal pragmatically with the strategic challenges that AI raises. One measure would be to support awareness-raising measures that will help the relevant stakeholders—governmental practitioners, industry and civil society, among others—gain a realistic sense of the challenges posed by AI in the nuclear arena. Another measure would be to support transparency and confidence-building measures that can help to reduce misperception and misunderstanding among nuclear-armed states on AI-related issues. An additional measure would be to support collaborative resolution of the challenges posed by AI and the exploration of beneficial use of AI for arms control. A final possible measure would be to discuss and agree on concrete limits to the use of AI in nuclear forces."

Cyberattack (or defense) capabilities

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/High	Medium/High	Medium

What developments might occur, and why might they occur?

Which systems could be targeted?

There could be increases in cyberattack capabilities which are relevant to cyberattacks on nuclear weapons and delivery systems, detection and defense systems, autonomous weapons, NC3, and/or perhaps other nuclear-weapons-relevant systems.

What might lead to these capability increases?

These increases in cyberattack capabilities could occur either due to AI advances or via other means.

Furthermore, these capability increases could be deliberately targeted at nuclear-weapons-relevant systems or could be just as one effect of more general capability increases. In the former case, this could be incentivized partly by an adversary's actual or expected increase in reliance on AI, automation, or electronic systems in its nuclear-weapons-relevant systems, since that might increase how impactful a cyberattack could be. For example, if a state or nonstate actor expects an adversary to increase the integration of AI in its NC3 systems, that could motivate the actor to increase its investment in cyberattack capabilities in order to disable, disrupt, or manipulate those NC3 systems.

What actors might have these increased capabilities?

We'd guess that state actors, in particular states with nuclear weapons and/or large R&D budgets, will tend to develop much more advanced cyberattack capabilities than other states or nonstate actors.

However, we'd also guess there's a decent chance nonstate actors will sometimes either develop or otherwise gain access to cyberattack capabilities that are sufficiently advanced to pose nontrivial risks.[14]

Why might this increase nuclear risk?

Cyberattacks could in theory be used to:

- Disable or disrupt nuclear-weapons-relevant systems
    - E.g., detection systems, missile defense systems, NC3 systems necessary for giving launch orders
- Discover the location of nuclear-weapons-relevant systems (e.g., ballistic missile submarines)
- Launch nuclear weapons
- Cause false alarms
    - E.g., causing detection systems to signal high confidence that a large nuclear strike is incoming

These actions, or the mere expectation that they *could* be taken, could have a complicated variety of risk-increasing or risk-reducing effects, depending on exactly what is done and on other variables of the situation at the time. For example:

---

[14] One thing informing this view is that the costly 2017 WannaCry and NotPetya attacks used advanced hacking tools that had been stolen and leaked from the NSA by a group known as "The Shadow Brokers" (Greenberg, 2018; Schneier, 2017; White House, 2017; my thanks to Alex Lintz for these sources). This provides some evidence for the hypothesis that actors will sometimes be able to gain access to cyberattack capabilities beyond those capabilities that they could develop themselves. However, the identify of The Shadow Brokers is not (publicly) known, and it's possible they're tied to the Russian government, so this may not provide much evidence that *nonstate actors* could gain access to very advanced cyberattack capabilities without a state helping them.

- These actions, or the expectation of them, could **reduce confidence in a particular actor's second strike capabilities** (i.e., their ability to powerfully retaliate to a nuclear first strike against themselves), undermining deterrence and incentivizing both that actor and other actors to launch a pre-emptive nuclear strike.
- Or these actions, or the expectation of them, could ***increase* confidence in a particular actor's second strike capabilities**, enhancing deterrence.
    - E.g., if the US is worried that cyberattacks could undermine US missile defense systems, that increases how much the US should fear each individual Russian or Chinese nuclear weapon, which increases the expected downside to the US from striking first and destroying *most* Russian or Chinese weapons.[15] If Russia and China believe the US is thinking this way, that should make them less worried about a pre-emptive US first strike and hence less motivated to launch their own pre-emptive first strikes.
- Or these actions could **create a false alarm that causes humans to launch a nuclear strike**.[16]
    - This could be very harmful if the strike that's launched involves many detonations on cities or if it triggers retaliation that immediately or iteratively results in many detonations occurring on cities.
- Or (in future) **a cyberattack could directly trigger a nuclear strike** (e.g., via controlling an NC3 system or creating a false alarm that triggers an automated retaliatory strike).

What about cyberdefense?

Cyberdefense capabilities are of course also likely to improve in the coming years and decades. It's plausible that the overall contribution of cyberattack risk to nuclear risk will therefore decrease rather than increase in future.[17] The reason we focus more on cyberdefense than cyberattacks in this post is simply that the scope of this post is technological developments that could *increase* risks from nuclear weapons.

That said, **it's also possible that in particular situations, or even in general, increases in cyber*defense* capabilities would *increase* nuclear risk**. This is essentially for the same reasons noted above for why increases in cyberattack capabilities might *decrease* nuclear risk and why better detection and missile defense systems might *increase* nuclear risk. For example, better US cyberdefense could reduce US worries that cyberattacks could undermine US missile

---

[15] This is just an example, and we're not sure if it's actually an important point, in part because we're not sure if US missile defense systems are really reliable enough to make a big difference to deterrence anyway.

[16] Relatedly, NTI (2018) writes: "Magnifying risks of a nuclear mistake are cyber threats to warning and command and control systems. Issues surrounding decision time become more acute in a world of increased cyber risks and little communication or cooperation between political and military leaders. Malicious hackers today may insert the same message that panicked Hawaiians in January 2018—"Ballistic missile threat inbound … seek immediate shelter. This is not a drill"—into national warning and alert systems. How would the leaders of Pakistan, India, North Korea, the United States, Russia, China, Britain, France, or Israel respond?" (NTI, 2018)

[17] See also "How does the offense-defense balance scale?" (Garfinkel & Dafoe, 2019).

defense systems, which in theory could increase US and hence Russian and Chinese incentives to launch a pre-emptive nuclear strike, for the reasons outlined above.[18]

Additional notes

Views and statements from NTI:

- Cybersecurity is a focus area for the NTI.
- NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in cyberspace.
- In a call with SoGive, senior NTI staff seemed to indicate they thought that cyber advances could increase nuclear risks to a worrying extent, that technical solutions wouldn't work, and that it would therefore be better to use policy solutions - primarily those which increase how long people have to make decisions in crisis situations. One example would be having warheads in a secure facility rather than loaded on delivery systems most of the time. SoGive's paraphrasing/summary of part of the call:

> "Cyber is a new area that they [NTI] are focused on, particularly with regard to the intersection between cyber and nuclear. Having looked into this, they believe the bottom line is that there is no technical solution to the cyber vulnerabilities of the nuclear weapon system. There's no way you're ever going to have complete confidence that the nuclear system is not vulnerable in some way given the large number of digital components. So you really need to think about policy solutions. For example, if you're worried that a country like Russia is going to get into your command and control systems (through a backdoor) and maybe confuse you to make it look like you're under attack, one solution is to remove warheads from delivery systems in both countries to increase leadership decision time. We have missiles with multiple warheads on them ready to launch at a moment's notice. As an alternative, you could put the warheads in a secure facility somewhere, and buy yourself time, so that you could test the info you're receiving from command and control systems.
>
> Question: is it fair to say that everyone already wants this sort of protection against the risk that the enemy could take control of your systems?
>
> Ans: Not necessarily. The military of course wants to reduce the risk that an adversary can spoof its command and control systems, but it is largely focused on developing technical solutions, and it remains unconvinced that policy solutions, such as taking vulnerable nuclear forces off 'prompt launch,' are a good idea. The military is trained to be 'ready' – i.e. ready to launch, even if the price of that readiness is a higher risk of blundering into nuclear war by firing a

[18] Again, this is just an example of something that *in theory could* happen, and might not actually be important in practice.

weapon erroneously, under severe time pressure, based on false warning information from a possible cyber event."

- Stoutland and Pitts-Kiefer (2018) makes the following policy recommendations (among others):
    - "Secure and diversify critical systems. [...] Possible measures could include maintaining and enhancing reliance on nondigital systems, reducing complexity, hardening satellite and other communications systems, securing and diversifying the supply chain, and increasing diagnostic testing of components. Additional measures could include using dynamic solutions that increase resilience of critical communications systems."
    - "Bilateral and multilateral dialogues should consider norms and rules of the road—for example, agreement to refrain from using cyberattacks against nuclear weapons systems. Those dialogues also should consider unilateral or reciprocal actions to reduce the risk of nuclear weapons use that could result from cyberattacks. As an example, the United States should seek ways to cooperate internationally to improve early warning systems— including through military-to-military cooperation—to further reduce the possibility of a cyber-induced false warning. The United States also should work independently and with other states to explore and develop improved verification tools that could be used to enhance confidence in future cyber arms control or confidence-building agreements and measures."

Rodriguez (2019) writes: "But the expanded use of technology in nuclear weapons systems may introduce *new* risks. Unal and Lewis warn that, because nuclear weapons now rely more heavily on digital technology, they've become more vulnerable to cyber attack (Unal & Lewis, 2018)."[19]

Baum (2015a) writes: "A third limitation is the possibility that cyber weapons could attack command and control systems. If they can, then they could be used as first strike weapons, to the detriment of strategic stability. On the other hand, countries are very likely to attempt protecting their command and control systems from cyber attacks, to the extent that they can."

One thing that may make developments in cyberattack capabilities more worrying is "entanglement" of command and control systems.

- Acton (2019) writes that "Entanglement describes how militaries' nuclear and non-nuclear capabilities are becoming dangerously intertwined", increasing the odds of nuclear conflict.

[19] For context, the preceding paragraph was: "It's tempting to imagine that the risk of accidental nuclear war has decreased as the technologies that have failed in the past, like early warning radar and fail-safes have likely improved. I found *some* evidence of this, though not as much as I expected. For example, the US's early warning systems that detect incoming missiles (which have been partially responsible for near-miss events in the past) have been upgraded with technology that is better at classifying and tracking projectiles (Owens, 2017). This likely reduces the risk that the US or Russia would mis-identify meteorological events or non-nuclear projectiles as an impending nuclear attack."

- For example, many command-and-control systems could be used for both nuclear and non-nuclear operations. Similarly, US early-warning warning satellites can detect (and trigger ballistic missile defenses against) either nuclear or non-nuclear attacks. Thus, an adversary may launch an attack (including a cyberattack) on such command-and-control systems or satellites to gain advantage in conventional conflict. But this may be perceived as an attempt to gain advantage in a nuclear conflict, or may otherwise cross an escalation threshold, causing the attacked state to launch a nuclear strike (Acton, 2019; Downman & Messmer, 2019).
- For some additional notes on entanglement, see Aird (2022)

The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence states: "the growing threat to nuclear arsenals (from nuclear strikes, conventional attacks, missile defenses, ASW, and **cyber operations**) raises major questions about the wisdom of cutting the size of nuclear arsenals. In the past, many arms control advocates believed that arms cuts reduced the incentives for disarming strikes; whether right or wrong in the past, that assumption is increasingly dubious as a recipe for deterrence stability today." (emphasis added)

Advances in autonomous weapons

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium/High	Medium

What developments might occur and why might this increase risks?

Unmanned underwater vehicles (UUVs)

- E.g., Russia's Poseidon system (also discussed above)
- Ladish (2019) writes:

> "While UUVs won't completely replace nuclear submarines anytime soon, I expect the undersea arena to be increasingly autonomous as UUVs become more capable. Stealth and lurk capabilities are the obvious advantages. A dormant UUV can sit on the bottom and won't get tired or bored.
>
> Advances in machine learning have the potential to both improve and weaken stealth tech. They will improve autonomous capabilities of unmanned stealth vehicles, but also improve signal processing and detection capabilities."

- Geist and Lohn (2018) write:

> “The difficulty of communicating underwater would require a degree of autonomous capability on the part of the drone that has become possible only recently as a result of progress in AI.”

- Perhaps AI could enable swarm systems that can effectively wage anti-submarine warfare, and that could in turn undermine deterrence?
    - We’re very unsure whether that’s a reasonable speculation, and are basing it solely on drawing inferences from the following passages in the article Fear of false negatives: AI and China's nuclear posture:

> > “When the topic turns to leveraging new means of warfare, **Chinese writings discuss the use of swarm systems (link in Chinese) for a number of purposes, with battlefield applications focusing on anti-submarine warfare and countering integrated air defense.**
> >
> > AI and autonomy provide China an opportunity to exploit a new pocket of excellence, but they are hardly ends in themselves. This is one of myriad reasons that China has been reluctant to engage in arms control efforts to constrain the deployment of autonomous systems. **Moreover, the amount of Chinese research already being conducted in this arena, particularly at the university level, is substantial. Research is unlikely to diminish any time soon**. (Programs on AI and autonomy receive ample government support through such funds as the Laboratory of National Defense Technology for Underwater Vehicles, Project for National Key Laboratory of Underwater Information Processing and Control, National Key Basic Research and Development Program, China Aviation Science Foundation, National Science and Technology Major Project, National 973 Project, National Key Laboratory Fund, National “863” High-tech Research and Development Program, and Ministry of Communications Applied Basic Research Project, among a number of others.)
> >
> > **Expansive programs to turn AI and autonomy into a weaponized reality, even in challenging or illusory domains such as underwater swarms, indicate the emphasis this research receives within the hierarchy of Chinese defense planning.** Whether or not China is able to achieve all of these capabilities, the vast resources and manpower allocated to these endeavors merit great attention by the United States. **The direct implications of aerial and underwater swarms for larger, more lumbering US nuclear and conventional platforms remain to be seen. However, if the US Congress provides funding for the low-yield submarine-launched ballistic and cruise missiles proposed under the 2018 Nuclear Posture Review, China could deploy swarms to track and potentially intercept US dual-capable platforms.** In short,

whether intentionally or unintentionally, an escalatory scenario could develop." (emphasis added)

Unmanned aerial vehicles (UAVs)

Autonomous weapons in general

- Increases in the (perceived) cheapness, ease of development, versatility, or capability of autonomous weapons could lead to them being deployed, relied on, and worried about in a wider range of situations and to a greater extent.
    - This *might* have effects such as increasing the chance of conflict between nuclear-armed states or increasing the extent to which such conflicts escalate. If so, that would probably (though not *definitely*) increase nuclear risk.
    - Further discussion of that topic can be found in the pieces linked to from Autonomous weapon - EA Forum and the pieces that they in turn link to.
- Shulman (2020) notes:
    - "all-out discharge of strategic nuclear arsenals is also much more likely to be accompanied by simultaneous deployment of other WMD, including pandemic bioweapons (which the Soviets pursued as a strategic weapon for such circumstances) **and drone swarms (which might kill survivors in bunkers)**; the combined effects of future versions of all of these WMD at once may synergistically cause extinction" (emphasis added)

Additional notes

- See "How does the offense-defense balance scale?" (Garfinkel & Dafoe, 2019)
- Fear of false negatives: AI and China's nuclear posture states:
    - "When it comes to platforms, this author's preliminary review of Chinese technical writings on AI and autonomy reveals that Beijing's greatest emphasis, at least where the most flexible systems are concerned, is on unmanned aerial and underwater vehicles. In China's view, these systems can be leveraged for a range of activities, including enhanced accuracy in: battlefield reconnaissance, surveillance, patrolling, electronic reconnaissance, communications, electronic interference, combat assessment, radar deception, projectile firearms, laser guidance, target indication, precision bombing, interception and launch of tactical missiles and cruise missiles, and anti-armor, anti-radiation, and anti–naval vessel capabilities; as well as nuclear, chemical, and biological detection and operations.

      [...] Chinese discussions about keeping "a human in the loop" (人在环路) with meaningful human control (link in Chinese) are limited to nonexistent. This indicates a gap in the current discourse—a neglect of AI and autonomy's potential adverse impact on military command and control."

More integration of AI with NC3 systems

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium	Medium	Medium

Background

The US's 2018 Nuclear Posture Review (NPR) states:

> "The United States must have an NC3 [nuclear command, control, and communications] system that ensures command and control of U.S. nuclear forces at all times, even under the enormous stress of a nuclear attack. NC3 capabilities must assure the integrity of transmitted information and possess the resiliency and survivability necessary to reliably overcome the effects of adversary nuclear attack. The NC3 architecture is essential for deterrence and enables a response if deterrence fails.
>
> During peacetime and crisis, **the NC3 system performs five crucial functions: detection, warning, and attack characterization; nuclear planning; decision-making conferencing; receiving Presidential orders; and enabling the management and direction of forces**.
>
> Today's NC3 system is a legacy of the Cold War, last comprehensively updated almost three decades ago. **It includes interconnected elements composed of warning satellites and radars; communications satellites, aircraft, and ground stations; fixed and mobile command posts; and the control centers for nuclear systems.**
>
> › Warning systems include fixed, terrestrial phased array warning radars; the Defense Support Program (DSP) system and its replacement, the Space Based Infrared System (SBIRS); and the U.S. Nuclear Detonation Detection System (USNDS).
>
> › Communications systems include the Military Strategic and Tactical Relay (MILSTAR) satellites and its replacement, the Advanced Extremely High Frequency (AEHF) satellites; a wide variety of ground-based transmission systems across the radio frequency spectrum; and Take Charge and Move Out (TACAMO) relay aircraft.
>
> › The fixed command posts include the National Military Command Center (NMCC) and the U.S. Strategic Command Global Operations Center. Fixed command posts also include linkages to U.S. forward-deployed forces in USEUCOM and elsewhere. Mobile command posts include the E4B National Airborne Operations Center (NAOC), the E6B Airborne Command Post (ABNCP), and ground mobile systems.

› Control centers for nuclear systems are in ICBM Launch Control Centers, on SSBNs, and aboard bomber aircraft.

While once state-of-the-art, the NC3 system is now subject to challenges from both aging system components and new, growing 21st century threats. **Of particular concern are expanding threats in space and cyber space, adversary strategies of limited nuclear escalation, and the broad diffusion within DoD of authority and responsibility for governance of the NC3 system, a function which, by its nature, must be integrated**. [emphasis added]

What developments might occur and why might this increase risks?

Two aspects of NC3 systems where it seems plausible AI will become more integrated and that could be important are:

- **Early warning systems**
- **Launch decision-support systems**

Something like an AI-powered Dead Hand/Perimeter would fall into both of these categories.

Fear of false negatives: AI and China's nuclear posture states that:

> "Chinese discussions about keeping "a human in the loop" (人在环路) with meaningful human control (link in Chinese) are limited to nonexistent. This indicates a gap in the current discourse—a neglect of AI and autonomy's potential adverse impact on military command and control."

Why might this increase risk?

See Geist and Lohn (2018). Some relevant quotes are below:

> "There is a range of different approaches to subverting AI systems, and it appears that subversion is likely to be an effective option for a long time to come. We will briefly discuss hacking, training data attacks, and input manipulation as illustrations of the types of concerns that exist."

> "It is also possible that an adversary could become convinced that it is able to subvert an AI and avoid retaliation, leading it to pursue paths that would otherwise be escalatory in nature, up to and including preemptive first strike. For example, the adversary might be convinced that it has discovered a pattern of launches and trajectories that would lead the AI to view the data and conclude that such a pattern is safe even as missiles are en route to targets."

> "AI presents an array of new vulnerabilities that are difficult to detect in real time. Yet it will almost certainly—eventually or gradually—be given more prominence in road-to-war, escalation, and even launch decisions. Any system with those responsibilities should have to go through rigorous testing that would include adversarial approaches. The

simulation of adversaries in testing is fully effective only if the tester can envision the full range of attacks an adversary might create. This impossibly tall order is nonetheless faced for all military systems that are deployed."

"Workshop participants agreed that the riskiest periods will occur immediately after AI enables a new capability, such as tracking and targeting or decision support about escalation. During this break-in period, errors and misunderstandings are relatively likely. With time and increased technological progress, those risks would be expected to diminish. If the main enabling capabilities are developed during peacetime, then it may be reasonable to expect progress to continue beyond the point at which they could be initially fielded, allowing time for them to increase in reliability or for their limitations to become well understood. Eventually, the AI system would develop capabilities that, while fallible, would be less error-prone than their human alternatives and therefore be stabilizing in the long term"

"In numerous convincing demonstrations, small amounts of adversarial effort toward subverting machine learning algorithms have shown outsized effect. Some researchers argue that this is a pervasive trait of machine learning and that they expect that it will persist for years to come. Where an effective AI for tracking and targeting might be destabilizing and lead to proliferation or worse, an adversary may regain trust in the survivability of its second-strike forces if it is confident in its ability to forestall detection using these adversarial methods, thereby reestablishing strategic stability. On the other hand, an actor may believe that it can subvert an AI's ability to identify a preemptive first strike, making such a strike a viable option and therefore destabilizing"

"Some workshop participants were convinced that humans would be unwilling to let the computer influence decisions about nuclear war, while others could easily envision growing comfortable with the idea. Anecdotally, the difference in perspective was generational, suggesting that those who will have inherited the reins by 2040 will be more comfortable with abdicating some degree of control, especially as AI continues to prove itself in increasingly complex and day-to-day tasks over the coming decades. It is already common for Americans to rely on AI to make routing decisions when driving, facilitate scheduling tasks, and respond to simple e-mails. Perilously, these successes may build confidence that is unwarranted considering the chasm between routine decisions and nuclear war."

- See also the section on radical transparency (pp. 21-22)

Hruby and Miller (2021) states:

"Vulnerabilities within Russia's Perimeter system could lead to high-consequence mistakes. For instance, risk researcher and RAND contributor Anthony Barrett suggests that Perimeter could misconstrue and misattribute data caused by a meteorite strike to detonations from a U.S. nuclear attack or that the system could launch nuclear weapons at the United States if communication lines were severed during a terrorist attack."

On the other hand, AI and International Stability: Risks and Confidence-Building Measures states:

> "The assurance of automated retaliation could be valuable as a deterrent and/or to reduce the incentives for a nation's leaders to launch a strike under ambiguous warning, if they felt confident that a second strike was assured. An agreement to rule out the use of automated 'dead hand' systems might increase the risk of first strike instability, because nations could have a larger incentive to strike first—or perhaps launch in response to a false alarm—before being decapitated."

(But note that elsewhere the same article raises concerns about increasing the integration of AI with NC3.)

What could be done about this?

- **Minimize integration of AI with NC3 systems**
  - Get countries to themselves refrain from this
  - Promote a norm of not doing this
  - See the forecasts collected below
  - See Intersections between nuclear risk and AI for quotes from some high-profile reports (including NSCAI, 2021) arguing in favor of this

(As with most other parts of this post, this is unlikely to be a comprehensive list of all things that might be worth mentioning.)

Relevant forecasts

(These are all from Metaculus. Emphasis added.)

- By 2024, will **the next [US] Nuclear Posture Review** explicitly affirm that decisions to authorize nuclear weapons employment must only be made by humans?
- By 2024, will **Russia** clearly and publicly affirm that decisions to authorize nuclear weapons employment must only be made by humans?
- By 2024, will **China** clearly and publicly affirm that decisions to authorize nuclear weapons employment must only be made by humans?
- By 2024, will **a nuclear-armed state other than the US, Russia, or China** clearly and publicly affirm that decisions to authorize nuclear weapons employment must only be made by humans?

Additional notes

- See Intersections between nuclear risk and AI for some relevant quotes and sources
- NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in AI.
- Arnold and Toner (2021) give the following hypothetical example of how integration of AI with command and control of *missile defense* could lead to catastrophe:

  > "Phantom missile launches: In missile defense, seconds of delay can spell the difference between an interception and a miss. U.S. Strategic Command's new

missile defense system, Global Eye, eliminates delay by scanning gigabytes of real-time data every second. If the system's algorithms detect a missile launch with high certainty, the system can quickly and autonomously trigger an interceptor launch in order to shoot down the missile. One day, unusual atmospheric conditions over the Bering Strait create an unusual glare on the horizon. Global Eye's visual processing algorithms interpret the glare as a series of missile launches, and the system fires interceptors in response. As the interceptors reach the stratosphere, China's early-warning radar picks them up. Believing they are under attack, Chinese commanders order a retaliatory strike."

- Roberts (2020) states:

  "Another key question is whether AI will strengthen or erode NC2's deterrence effectiveness. AI could help buy decision time for national leadership and improve situational awareness, thereby reducing pressures to launch under attack. But it might prove brittle and thereby fail to gain the trust of decision makers."

Additional notes relevant to this category as a whole

AI and International Stability: Risks and Confidence-Building Measures states:

"Some U.S. military leaders and official DoD documents have expressed skepticism about integrating uninhabited vehicles into plans surrounding nuclear weapons. The Air Force's 2013 Remotely Piloted Aircraft (RPA) Vector report proposed that nuclear strike 'may not be technically feasible unless safeguards are developed and even then may not be considered for [unmanned aircraft systems] operations.' U.S. Air Force general officers have been publicly skeptical about having uninhabited vehicles armed with nuclear weapons. General Robin Rand stated in 2016, during his time as head of Air Force Global Strike Command, that: 'We're planning on [the B-21] being manned. … I like the man in the loop … very much, particularly as we do the dual-capable mission with nuclear weapons.'

Other U.S. military leaders have publicly expressed support for limits on the integration of AI into nuclear command-and-control. In September 2019, Lieutenant General Jack Shanahan, head of the DoD Joint AI Center, said, 'You will find no stronger proponent of the integration of AI capabilities writ large into the Department of Defense, but there is one area where I pause, and it has to do with nuclear command and control.' In reaction to the concept of the United States adopting a 'dead hand' system to automate nuclear retaliation if national leadership were wiped out, Shanahan said, 'My immediate answer is "No. We do not." … This is the ultimate human decision that needs to be made which is in the area of nuclear command and control.'

While the motivation for these statements about limits on the use of autonomy may or may not be strategic stability—bureaucratic factors could also be at play—they are examples of the kinds of limits that nuclear powers could agree to set, unilaterally or collectively, on the integration of AI, autonomy, and automation into their nuclear

operations.

Nuclear states have a range of options for how to engage with these kinds of risks. On one end of the spectrum are arms control treaties with some degree of verification or transparency measures to ensure mutual trust in adherence to the agreements. On the other end of the spectrum are unilateral transparency measures, which could have varying degrees of concreteness ranging from informal statements from military or civilian leaders along the lines of Shanahan's and Rand's statements, all the way to formal declaratory policies. In between are options such as mutual transparency measures, statements of principles, or non-legally binding codes of conduct or other agreements between nuclear states to ensure human control over nuclear weapons and nuclear launch decisions. Even if states that desired these restraints found themselves in a position where others were unwilling to adopt more binding commitments, there may be value in unilateral transparency measures both to reduce the fears of other states and to promulgate norms of responsible state behavior. As with other areas, it is important to consider incentives for defection from an agreement and the extent to which one state's voluntary limitations depend on verifying others' compliance with an agreement. If some states, such as the United States, desire strict positive human control over their nuclear weapons and nuclear launch authority for their own reasons, then verifying others' behavior, while desirable, may not be a necessary precondition to those states adopting their own limits on the use of AI, autonomy, or automation in nuclear operations.

Two possible CBMs for AI applications in the nuclear arena involve nuclear weapons states agreeing to strict human control over nuclear launch decisions and ensuring any recoverable delivery vehicles are human-inhabited, to ensure positive human control."

The article then discusses those two possible CBMs.

Fear of false negatives: AI and China's nuclear posture states:

"Research within these [Chinese] organizations focuses on integrating AI and autonomy in order to facilitate functions such as fault detection and diagnosis, embedded training, and simulation and modeling, as well as data accumulation and processing for remote sensing and situational awareness. None of these activities is necessarily destabilizing, even when one factors in the cross-over between China's civilian and military research and development. In some respects, improvements in China's reconnaissance capabilities and in the reliability of a range of its platforms could even be a stabilizing measure. If China gains greater situational awareness and ensures its retaliatory capabilities in the nuclear sphere, some of its insecurities about a "bolt-out-of-the-blue" strike may be mitigated. Yet Chinese insecurities are not simply a question of technology. They are also rooted in a set of concerns about false negatives—that is, failures by China's early warning systems to detect a real attack—and assumptions about US intent.

> In the United States, military analysts are often preoccupied with the concern that alarms or early warning systems, accidentally or even intentionally triggered, could produce false positives. Chinese analysts, in contrast, are much more concerned with false negatives. In other words, they are preoccupied with the inability of their systems to identify, much less to counter, a stealthy and prompt precision strike. China's assumptions about its own deficiencies in early warning (link in Chinese), combined with its concerns over US advances in high-precision, high-speed systems ranging from conventional prompt global strike to space planes, imply that technologies such as AI and autonomy could indeed take on destabilizing qualities. As China further develops its concept of "rapid response" (快速反应), as cited in its 2015 Military Strategy [中国的军事战略] (link in Chinese), **Beijing's integration of AI and autonomy into its military systems is likely to increase. Such integration could range anywhere from automation-enabled launch-on-warning for its missiles to autonomy- and AI-enabled maneuverability and precision guidance for hypersonic glide platforms and space planes.**
>
> [...] New pockets of excellence. In its relations with Russia and the United States, China has long contended with nuclear asymmetry. AI and autonomy, in contrast, offer Beijing the long-term potential to disrupt Washington's traditional strengths. They open the door for **swarm and other technologies that could overwhelm conventional and nuclear platforms that are larger, more cumbersome, and less agile**. While **China may be concerned about potential adversaries tracking its own nuclear platforms and systems, Beijing is just as likely to avail itself of these relatively inexpensive methods of disrupting US activities. Also, Chinese publications indicate that Beijing is building autonomy into its own "bolt-out-of-the-blue" systems, for example in hypersonic glide vehicles such as the DF-ZF. As China debates integration of automation via launch-on-warning, doing so with a greater range of AI and autonomy in its tool kit could lead to destabilizing trends.** Again, the most sensational advances in these enabling technologies do not necessarily carry the greatest implications for China's military and nuclear force structure. Instead, what counts is the level of AI and autonomy introduced into Beijing's command and control structure."

Perhaps increasing perceptions that automation can be unpredictable, unreliable, brittle, etc. would be a useful intervention for reducing risks?

- This could be useful via reducing the chances of risk-increasing uses of automation, and/or via making people behave more cautiously in other respects.
- This might work whether the perceptions are warranted or not (though of course one should typically be wary of saying misleading or false things).
- This was inspired by the following passage from Horowitz, Scharre, and Velez-Green (2019):

> "How automation is perceived also matters. If policymakers view automation as unpredictable and unreliable and therefore its use induces caution, then countries might be less willing to engage in brinkmanship. On the other hand, if nations

> viewed automation as more safe and reliable than it actually was, then it could lead policymakers to underestimate the chances of accidents or miscalculation and take risks they do not understand."

Horowitz, Scharre, and Velez-Green (2019) conclude:

> "Technology is not destiny. The rapid progress of AI and automation in the commercial sector opens up opportunities for militaries, but militaries have a choice about how they incorporate automation into their forces. Some forms of automation could increase reliability and surety in nuclear operations, strengthening stability, while other forms could increase accident risk or create perverse incentives, undermining stability. As in other aspects of nuclear stability, second and third-order consequences must be understood. Actions that appear beneficial can sometimes have counterintuitive consequences, especially when accounting for an adversary's decision calculus. When modernizing nuclear arsenals, policymakers should aim to use automation to decrease the risk of accidents and false alarms and increase human control over nuclear operations."

Earlier, the same paper discusses "trust gaps" and "automation bias":

> "Questions about adopting autonomous systems require potential adopters to grapple with how to balance the risk that humans will not trust machines to operate effectively against the risk that humans will trust machines too much. Trust gaps occur when people do not trust machines to do the work of people, even if the machine outperforms humans in benchmark tasks. This can lead to an unwillingness to deploy or properly use systems.
>
> [...] Automation bias, whereby humans effectively surrender judgment to machines, therefore represents one important risk from automation. For example, Army investigators found that automation bias was a factor in the 2003 Patriot fratricides, in which Army Patriot air and missile defense operators shot down two friendly aircraft during the opening stages of the Iraq War. In both instances, humans were "in the loop" and retained final decision authority for firing, but operators nevertheless trusted the (incorrect) signals they were receiving from their automated radar systems."

Non-nuclear warmaking advancements (other than AI and cyber)

Anti-satellite weapons (ASAT)

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/Low	Medium	Medium/Low

What this is

Wikipedia states:

> "**Anti-satellite weapons (ASAT) are space weapons designed to incapacitate or destroy satellites for strategic or tactical purposes. Several nations possess operational ASAT systems**. Although no ASAT system has yet been utilised in warfare, **a few countries (China, India, Russia, and the United States) have successfully shot down their own satellites to demonstrate their ASAT capabilities in a show of force.** ASATs have also been used to remove decommissioned satellites.
>
> **ASAT roles include: defensive measures against an adversary's space-based and nuclear weapons, a force multiplier for a nuclear first strike, a countermeasure against an adversary's anti-ballistic missile defense (ABM)**, an asymmetric counter to a technologically superior adversary, and a counter-value weapon." (emphasis added)

What developments might occur?

ASATs could potentially become cheaper/easier to produce, more accurate, or more effective in other ways. Each of those developments of ASATs would presumably increase the chances of various countries (e.g., Israel) gaining ASATs, the number of ASATs each country has, and the extent to which countries do and are expected to rely on ASATs in their nuclear strategy. That is, **there could be increased proliferation, numbers, effectiveness of, and/or reliance on ASATs**.

However, we're not sure how much ASATs would really affect nuclear strategy or conflicts. For example, Wikipedia notes (a) some limitations to the efficacy of ASAT and (b) some limitations the importance of satellites for US intelligence, surveillance, and reconnaissance in the first place:

> "While it has been suggested that a country intercepting the satellites of another country in a conflict, namely between China and the United States, could seriously hinder the latter's military operations, the ease of shooting down orbiting satellites and their effects on operations has been questioned. Although satellites have been successfully intercepted at low orbiting altitudes, the tracking of military satellites for a length of time could be complicated by defensive measures like inclination changes. Depending on the level of tracking capabilities, the interceptor would have to pre-determine the point of impact while compensating for the satellite's lateral movement and the time for the interceptor to climb and move.
>
> U.S. intelligence, surveillance and reconnaissance (ISR) satellites orbit at about 800 km (500 mi) high and move at 7.5 km/s (4.7 mi/s), so a Chinese Intermediate-range ballistic missile would need to compensate for 1350 km (840 mi) of movement in the three minutes it takes to boost to that altitude. Even if an ISR satellite is knocked out, the U.S. possesses an extensive array of manned and unmanned ISR aircraft that could perform missions at standoff ranges from Chinese land-based air defenses, making them somewhat higher priority targets that would consume fewer resources to better engage."

Why might this increase (or decrease) nuclear risk?

- Increased proliferation, numbers, effectiveness of, and/or reliance on ASAT **could perhaps undermine deterrence**.
    - In particular, more numerous or effective ASATs could perhaps significantly increase the expected effectiveness of a counterforce nuclear first strike, such as by impairing an adversary's detection, missile defense, and NC3 systems. Knowledge of this could increase both parties' incentives to strike first.
- These developments could also **increase risk of accidental, authorized, or inadvertent nuclear conflict**.
    - If satellites are taken out, this could cut off communications between higher and lower levels of the chain of command, which could increase the chances that lower-level commanders launch nuclear weapons on their own initiative (e.g., because they fear Moscow or DC has already been destroyed). That could increase the risk of nuclear war starting, of retaliation, and/or of escalation.
    - Ellsberg (2018) indicated that he was worried about this risk from ASAT and that he saw this as one of many reasons why a nuclear exchange would be very unlikely to remain limited.[20]

> "Making [nuclear threats] feasible opened the possibility that deliberately or not, they would be carried out by someone or other, if not the Prime Minister or Premiere or General Secretary, by someone else who had access to those weapons. It was true on the U.S. side, this is another revelation in the book, that in order to make it impossible to paralyze our response, our retaliation, by a single weapon on Washington on our command post, a few weapons on command post, **the authority to initiate or to use the weapons, the U.S. weapons, had been delegated by President Eisenhower to a number of high-level commanders who had in turn delegated it for the same reason to their lower commanders.**
>
> If the lower ones were out of communication, which happened every day in those days for technical reasons, that was before we had the system of satellites, and Washington was out of communication with our headquarters in Oahu part of everyday. That's not true now. It does depend on satellites, by the way, which **both sides are working on anti-satellite weapons to clear the air, the space, of those connecting nodes so that really very early in a war, the weapons may well, or could even say probably, will be out of contact with central headquarters. And it will depend, then, on human responses and decisions what to do. And in an environment in which nuclear**

[20] But note that one of us - Michael - is fairly confident some of Ellsberg's claims in that interview and in his book *The Doomsday Machine* overstate some aspects of nuclear respect, especially in the sense of making overly confident or extreme claims about nuclear winter. As such, Michael expects some of Ellsberg's other claims - which we lack the knowledge required to easily verify - also overstate risks.

> **weapons are going off and war is on.** So the idea of controlling that and limiting that to a small exchange is very low and it doesn't take much to destroy this society." (emphasis added)

- On the other hand, these developments **could also perhaps *strengthen* deterrence**.
    - For example, more numerous or effective ASATs could perhaps undermine missile defense systems, which could perhaps strengthen deterrence given that missile defense systems can under some circumstances undermine deterrence.
- Finally, these developments **could also perhaps reduce reliance on nuclear weapons**.
    - This is because ASAT *might* provide an alternative option for a countervalue weapon that can be effectively used for deterrence.
        - But we haven't looked into this possibility at all; this is just our speculation.

Additional notes

The following passage from Rose (2020) is relevant to what developments might occur and how best to respond to this (though it doesn't explicitly address how this affects nuclear risk):

> "Over the past decade, the threats to U.S. space systems from countries like Russia and China have continued to grow. Former Director of National Intelligence Dan Coats testified to Congress in 2019 that "Russia and China are training and equipping their military space forces and fielding new antisatellite weapons to hold US and allied space services at risk." Indeed, potential adversaries understand how dependent the U.S. military is on outer space, so these trends are likely to continue for the foreseeable future.
>
> Consistent with actions taken during the Obama administration, the Trump administration has pursued steps to increase the resiliency of U.S. space systems and enhance deterrence in outer space, including through the establishment of the U.S. Space Force, and the re-establishment of U.S. Space Command. Some progressive Democrats have called on the Biden administration to eliminate the Space Force. However, given the growing threat to U.S. and allied space systems, such a move would be unwise. Instead, the Biden administration should work to ensure that the Space Force improves the integration of space across the Joint Force; encourages further integration of U.S. allies and partners into space operations; and enhances the resiliency of our national security space systems. In addition to these military-related steps, the Biden administration should take steps to revitalize America's space security diplomacy, which has largely been an afterthought during the Trump administration. A Biden administration should consider ways to expand space security consultations with allies and partners, and promote norms of behavior that can advance the security and sustainability of the outer space environment."

One thing that may make developments in ASATs more worrying is "entanglement" of command and control systems.

- Acton (2019) writes that "Entanglement describes how militaries' nuclear and non-nuclear capabilities are becoming dangerously intertwined", increasing the odds of nuclear conflict.
- For example, many US early-warning warning satellites can detect (and trigger ballistic missile defenses against) either nuclear or non-nuclear attacks. Thus, an adversary may launch an attack on such satellites to gain advantage in conventional conflict. But this may be perceived as an attempt to gain advantage in a nuclear conflict, or may otherwise cross an escalation threshold, causing the attacked state to launch a nuclear strike (Acton, 2019; Downman & Messmer, 2019).
- For some additional notes on entanglement, see Aird (2022)

"Space planes" and other (non-ASAT) space capabilities

Tentative bottom-line views about this development

Importance	Likelihood / Closeness	Steerability
Medium/Low	Medium	Medium/Low

(We've looked into this possible development especially little, so this section is especially rough and preliminary.)

- NTI (2020) mentions as one possible cause for concern recent/ongoing/possible advances in "space capabilities".
- The article Fear of false negatives: AI and China's nuclear posture states that "Beijing's integration of AI and autonomy into its military systems is likely to increase" and that this could include (among other things) "autonomy- and AI-enabled maneuverability and precision guidance" for "space planes".
- The same article also states: "For Beijing, which has been expanding its nuclear arsenal at a relatively modest pace, the prospect of the United States resuming a forward-deployed, tactical nuclear posture exacerbates its sense of encirclement. Such a posture also amplifies **China's perceived and real vulnerability to US ambitions to field kinetic and surveillance platforms such as** prompt global strike, **X-variant space planes**, the surveillance aircraft Global Hawk, and so on." (emphasis added)
- Roberts (2020) states:

> "By 2030, U.S. space capabilities will have become more strategic in potential effect; by 2040, they could become more decisively advantageous to the United States. This is in sharp contrast to the last decade, when long-standing U.S. advantages in space eroded as adversaries fielded counter-space capabilities. By 2030, the United States may have redressed the worst through augmentation strategies, hybrid architectures, and space control measures. By 2040, it can push further ahead of Russia and China—but not without some fundamental and revolutionary changes to the way it thinks about space as a warfighting domain and prepares for war in space and for space in war.

> [...] By 2040, directed energy weapons and a significant move to space may have revolutionized the homeland defense."

Potential developments we ran out of time to look into at all

We tentatively expect the following things matter less than most of the potential developments discussed above, but we're not confident about that. We've spent less than 30 minutes learning or thinking about each of these things.

Much smaller nuclear weapons

- See also Suitcase nuclear device - Wikipedia and W54 - Wikipedia

Advances in nuclear energy

- Could increase proliferation of nuclear energy to additional countries, number of nuclear energy facilities per country, and/or countries' unwillingness to abstain from nuclear energy programmes.
- This *might* in turn increase nuclear weapons proliferation, though that's not obvious.
- See also our rough notes on "How might nuclear energy stuff affect nuclear weapons risk?"

Directions for future work

See Research project idea: Technological developments that could increase risks from nuclear weapons.

Acknowledgements

Michael's work on this post was supported by Rethink Priorities (though he ended up pivoting to other topics before having time to get this up to RP's usual standards). Will helped with the research and editing in a personal capacity.

We're grateful to Ben Snodin, Damon Binder, Fin Moorhouse, and Jeffrey Ladish for feedback on an earlier draft. Mistakes are our/Michael's own.

Appendix/Tangential thought

[To be included as a comment on the post.]

An in-our-view interesting tangential point: **It might decently often be the case that a technological development initially increases risks but then later increases risk by a smaller margin or even overall reduces risks.**

- One reason this can happen is that developments may be especially risky in the period before states or other actors have had time to adjust their strategies, doctrine, procedures, etc. in light of the development.
    - (This seems in some ways reminiscent of the Collingridge dilemma or the "pacing problem".)
- Another possible reason is that a technology may be riskiest in the period when it is just useful enough to be deployed but not yet very reliable.
- Geist and Lohn (2018) suggest this might happen, for the above two reasons, with respect to AI developments and nuclear risk:

    > "Workshop participants agreed that the riskiest periods will occur immediately after AI enables a new capability, such as tracking and targeting or decision support about escalation. During this break-in period, errors and misunderstandings are relatively likely. With time and increased technological progress, those risks would be expected to diminish. If the main enabling capabilities are developed during peacetime, then it may be reasonable to expect progress to continue beyond the point at which they could be initially fielded, allowing time for them to increase in reliability or for their limitations to become well understood. Eventually, the AI system would develop capabilities that, while fallible, would be less error-prone than their human alternatives and therefore be stabilizing in the long term"