Artificial Escalation:
The Backstory

Preface

Our new fictional film depicts a world where artificial intelligence ('AI') is integrated into command, control and communications systems, including nuclear systems ('NC3'), with disastrous results. The events depicted are based on a more detailed backstory that viewers may find interesting to read, so we've provided it here.

Watch the film: www.futureoflife.org/artificial-escalation

Notes and references

This backstory has been informed by real events, trends, and current research to make it both plausible and illustrative of the issues of concern in the military application of AI.

Please see the gray boxes throughout the document which contain references and notes for each of the timeline events.

🇺🇸 **US Perspective** (inc. Taiwan)	Mountain Time (MT)	🇨🇳 **China Perspective**	China Standard Time (CST)

2023-2027

The US has continued its program of developing a robust and multifaceted strategy for the defense of Taiwan from cyber and physical attack, including arms supplies, coordination on communication and monitoring systems, joint cyberdefence projects, etc.

Tensions have continued over the central role Taiwan continues to play in the manufacture of most of the world's most advanced chips, including those specialized for AI.

Strategic rivalry along this and other dimensions have led policies leading to more economic decoupling between the US and China.

Against this backdrop, progress in AI and, more recently robotics, has continued at a very rapid pace, with highly capable AI systems permeating large parts of society.

China has continued a military buildup with the objective of mounting a successful physical attack on Taiwan should it choose to do so. China has also pursued its high-profile strategy of leaning into AI development and AI integration into the military.

In addition, China has continued to appreciably enlarge its nuclear force, from several hundred in 2023 up to a projected 1000 by 2028, as well as actively discuss revisions of its policies regarding the use of nuclear weapons.

China continues to be a major player in AI development, with many powerful AI systems that are in competition with those developed in the US and Europe.

- Assuring Assured Retaliation: China's Nuclear Posture and U.S.-China Strategic Stability, 2015.
- The Return of Great Power War: Scenarios of Systemic Conflict Between the United States and China, 2022.
- Dangerous Confidence? Chinese Views on Nuclear Escalation, 2019.
- Beating the Americans at Their Own Game: An Offset Strategy with Chinese Characteristics, 2019.
- Biden Administration Clamps Down on China's Access to Chip Technology, 2022.

25 March, 2027

David Sinclair, CEO of Stratoenergetics, makes a successful closing pitch to leads of the US Joint All-Domain Command and Control (JDAC2) steering committee, the final step in approving a major contract to add an AI-based analysis and decision-support layer, named "TLDR", to the consolidated command and control system under development.

- The JDAC2 program is real, and seeks to create a consolidated fully-networked command and control system connecting all aspects of the US military - see:
  - Defense Primer: What Is Command and Control? 2022.
  - Joint All-Domain Command and Control: Background and Issues for Congress, 2022.
- Stratoenergetics is fictional, but it does have a website: https://stratoenergetics.com
- For aggregated expert forecasts on the likelihood of some of the ingredients of this scenario see this project.

2027-2030

As an AI system, TLDR has been trained using sophisticated simulations of the JDAC2 system itself, and "supervised learning" in which its advice is compared to that given by an expert team of humans. Training continued until it comes into close agreement with their advice, but much faster and more automatically. (Of course, no simulation is perfect, nor can such training cover a full range of possibilities.)

The JDAC2 program and TLDR appear to be a success. After being fully implemented in 2030, in several incidents involving tension in the Taiwan straight, TLDR has given sound analysis and useful suggestions.

In addition, the US program in autonomous and semi-autonomous weapons has forged ahead, despite the negative outcomes associated with the "slaughterbots" previously developed by Stratoenergetics itself.

The Chinese integration of AI into their command, control and communication infrastructure is even more pervasive than the US's, due to more consolidation in AI and close working relations between AI companies and the Chinese government. Interestingly, they have converged to a system (hereafter "TLDR-China") quite similar to TLDR in operation, though the underlying architecture is different.

China's approach to autonomous weapons has largely paralleled the US's, eschewing deployment of antipersonnel weapons due to international stigma, but developing significant capabilities in autonomous and semi-autonomous UAVs and small fully autonomous submarines.

- Systems Confrontation and System Destruction Warfare, 2018.
- Nuclear Command, Control, And Communications Systems Of The People's Republic Of China, 2019.
- Russia and China's space weapon plans spur high-level Pentagon meeting, 2022.

2031

In terms of nuclear weapons, the US's forces are broadly similar in the early 2030s to in the early 2020s. However, the US has taken the provocative step in early 2031 of arming Taiwan with a number of US-made cruise missiles. Because these missiles can carry either a conventional or nuclear payload, they have provoked furious outcries from Beijing. The US has [...]

China has (unwisely, it will turn out) moved from its deeply stabilizing "no first use" policy in which nuclear warheads are detached from their delivery system, to a version of "strategic ambiguity" and in which nuclear weapons are ready to launch on very short notice.

stated that these missiles are conventionally armed, but (very unwisely, it will turn out) has not allowed any inspections or confidence-building measures on that issue, and it is clear that Chinese leadership is highly concerned that they may be nuclear.

October 2032

Geopolitical tensions are high, with the US and Taiwanese continuing to build forces in Chinese territorial waters, and demonstrating a sometimes unpredictable range of reactions to standard Chinese military operations they deem to cross some ill-defined threshold.

- Taiwan's Quadrennial Defense Review, 2021.
- Crossing the Strait: China's Military Prepares for War with Taiwan, 2022.

2:18 AM — **26 October, 2032** — 04:18 PM

In one of the many military exercises in the Taiwan straight, a Chinese UAV suffers a malfunction and strays outside of its programmed flight path and into Taiwanese airspace.

TLDR correctly identifies this as an unpiloted but weapons-capable aircraft. Taiwanese military command, given various choices by TLDR, decides to up the ante somewhat in response by taking down this piece of hardware via one of the many surface-to-air (SAM) facilities deployed on Taiwan.

Chinese commanders are alerted that their aircraft is off course, and attempt to alert Taiwanese or US forces, but this communication is not accomplished in time. They consider its downing an over-reaction that requires some response.

Meanwhile, they have been looking for some excuse to test out an experimental cybersystem designed to be very effective against networked command and control targets. The system uses a reinforcement-learning trained AI hacking agent to rapidly test and penetrate secure networks and learn the networks' own communication protocols to spread to adjoining network nodes.

The Chinese cybercommand hopes a very limited test in this context will provide invaluable insight into the security layers of the Taiwanese and US joint network, and it has even been added as a priority to TLDR-China, leading TLDR to propose it even when it would not otherwise be the top option.

- This study suggests that military strategists are less likely to be forgiving of AI errors on the other side than human errors: Algorithms and Influence Artificial Intelligence and Crisis Decision-Making, 2022.

2:22 AM — 4:22 PM

China's AI cyberattack is launched. It penetrates the Taiwanese network even more successfully than hoped and, while creating minimal disruption, gathers a lot of information.

2:26 AM — 4:26 PM

Unbeknownst to the Chinese, the Taiwanese system is heavily connected into the US's JDAC2 system. Even more problematically, there are connections between JDAC2 and the US's nuclear command and control system (NC3) — although the US has publicly represented that this is not the case. The Chinese AI-powered cyberweapon is quite smart, even more than the Chinese realize – easily smart enough to realize the nuclear command and control is more important than conventional command and control, and also more "legacy" in some of its designs. The cyberweapon has unfortunately not been hard-coded not to enter NC3 systems (merely discouraged), and it is not sophisticated enough to care about exactly why this would be problematic. So it tries extra hard, and penetrates the US's NC3 by carefully crafting a FAX to a machine in Nebraska.

🇺🇸 **US Perspective** (inc. Taiwan)	Mountain Time (MT)	🇨🇳 **China Perspective**	China Standard Time (CST)

<div align="center">2:27 AM</div>

The US NC3 system has been under modernization for several years, and is now a somewhat messy admixture of the previous antiquated system, new elements designed as part of the overhaul, and some subsystems connected to TLDR.

The updated US system NC3 cyber system is set up to detect intrusions, especially from China or other geopolitical adversaries, and on the basis of its training the TLDR reacts strongly, making urgent requests to go into a heightened alert mode, notify other US NC3 systems, and set much tighter controls on authorization and communication.

These are accepted by the operator at the US Cheyenne mountain NORAD complex.

- It is part of China's strategy to disrupt US C&C, so this can reasonably be taken a potential step in an invasion. Moreover, the US has always been clear that attacks on its NC3 system are a red line in the nuclear security area. However, these lines have never been as clearly enunciated in relation to China as with the Russia/USSR.
- Joint All-Domain Command and Control: Background and Issues for Congress, 2022.
- Generating Actionable Understanding of Real-World Phenomena with AI, 2019.
- Nuclear Weapons in the New Cyber Age, 2018.

<div align="center">2:28:13 AM · 4:28:13 PM</div>

China has its own AI analysis agents observing US activities, and one — completely unrelated to Taiwan — is tasked with monitoring US NC3 systems. This system reacts similarly to the US one, going on a higher alert.

- Taiwan's Quadrennial Defense Review, 2021.

<div align="center">2:28:23 AM · 4:28:23 PM</div>

A different US NC3 system, set to monitor not the US NC3 network itself but rather the Chinese NC3 system, and also networked with TLDR, notices the Chinese system spinning up. TLDR adds it to the higher alert set by the earlier US system, and automatically increases the readiness level of US (and Taiwanese) anti-ballistic missile (ABM) and other counter-nuclear systems. TLDR does not require human approval for certain activities, such as "just" increasing alert levels.

- Assessing China-U.S. Inadvertent Nuclear Escalation, 2022.

<div align="center">2:28:29 AM · 4:28:29 PM</div>

The Chinese AI integrated threat assessment system detects the sudden change in the US ABM system and, combining it with an earlier notice, conclude that something alarming is happening, which it reports to TLDR-China. Thats system's training includes absorbing some expert-constructed knowledge graphs, one element of which is that in a US first strike, one of the US's first moves would be to ready its ABM systems.

(In fact, US planning is to do this very "late in the game" so as not to alarm the Chinese – one of many regrettable gaps in each party's understanding of the nuclear policies of the other.)

This system sends commands to the Chinese missile systems to commence hardening of fixed systems and relocation of mobile systems.

- Dangerous Confidence? Chinese Views on Nuclear Escalation, 2019.
- Deterrence in the age of artificial intelligence & autonomy: a paradigm shift in nuclear deterrence theory and practice? 2020.

<div align="center">2:28:35 AM · 4:28:35 PM</div>

The US monitoring system detects this change very quickly (almost before anything actually starts moving), and (metaphorically) starts flashing red.

| 🇺🇸 **US Perspective** (inc. Taiwan) | Mountain Time (MT) | 🇨🇳 **China Perspective** | China Standard Time (CST) |

<div align="center">2:28:37 AM</div>

<div align="right" style="width:50%">

4:28:37 PM

Unfortunately, in the Chinese system the procedure for protecting nuclear systems also includes a major software component. Without anyone deliberately choosing to do so, this is engaged as well. It includes an array of non-physical and non-intrusive — but nonetheless quite effective — mechanisms of data poisoning for AI systems, deployment of fake data streams to public and semi-public networks, and all manner of other clever ideas dreamed up to confuse US information systems.
</div>

- [Preparing the Cyber Battlefield: Assessing a Novel Escalation Risk in a Sino-American Crisis](#), 2020.

2:29:15 AM · 4:29:15 PM

TLDR is very "confused." It is integrating a massive number of data streams from ships, satellites, land radars, and even social media. These streams are starting to become wholly inconsistent with each other in ways far outside its training regime. TLDR starts to send alerts and error messages that read out as those systems being "blinded" and "compromised".

Of course, TLDR is not conscious, or self-aware, or really "understand" things in anything like the way a human does. Nonetheless it is not inaccurate to say that from TLDR's "view," the US has now been subject to cyberattack on its NC3 system, the Chinese are reconfiguring their nuclear systems, and all manner of data feeds are out of their normal ranges. Something major is happening, and it has to do with nuclear.

TLDR has all sort of options to deal with this, and it sends them for human confirmation. (It has been programmed to over-react unless the action is actually a kinetic life-taking action, since the human operator can always simply override the suggestion, right?) One of these is to engage "automatic cyberdefenses." Needless to say, all of these cyberdefenses are not exactly defensive.

2:29:25 AM · 4:29:25 PM

A Taiwanese ABM system, already on high alert, notices China's nuclear hardening-and-relocation and starts flashing red. This system has a high threshold because Taiwan sets low probability on a Chinese nuclear first strike. But a cautious and seemingly safe option appears to be to send a signal to increase the general Taiwanese military alert level. But because that level is already very high, this trips it into really high. This, of course, is noticed by the Chinese system, adding to the general weight of evidence it is assessing.

2:30 AM · 4:30 PM

<div align="right" style="width:50%">

Chinese military command, supported by TLDR-China, already trying to understand why China's systems are mobilizing, now see an indication of widespread attacks on Chinese cyber systems apparently coming from both the US (as automated cyber "defenses") and from Taiwan (which in deed they are, in response to the Chinese AI hacking system.) It also sees a notification that both US and Taiwanese ABM systems have changed configuration.
</div>

- [What if China invaded Taiwan? Here's what could happen](#), 2022.

2:32 AM · 4:32 PM

US Senior aides and Military leaders who are onsite gather in the Oval Office in expectation that a conversation with the President will be needed soon. They maintain close contact with the officers at NORAD.

US Perspective (inc. Taiwan) — Mountain Time (MT)	China Perspective — China Standard Time (CST)
2:32 AM	**4:32 PM**
	TLDR-China makes two recommendations to human operators, both of which are accepted.
	First, to engage the default plan if hostilities break out in Taiwan, which is to start a steadily escalating overall cyberoffensive aimed at disrupting and poisoning communications within Taiwan as well as between Taiwan and the outside.
	The second regards highly China's clandestine program to build small fully-autonomous highly stealthy underwater drones that quietly follow US missile-capable nuclear submarines. Although the primary purpose is simply monitoring (and occasionally communicating data to the mainland), they are also equipped with sensors to detect activity by and in the subs, as well as munitions sufficient to disable or even destroy their target. A command they can be given is to close the distance between themselves and their target, so as to obtain better data, or to attack more quickly. TLDR-China suggests doing so. The decision is accepted. This will prove a fatefully poor decision.

- Artificial Intelligence, Strategic Stability and Nuclear Risk, 2020.
- Deterrence in the age of artificial intelligence & autonomy: a paradigm shift in nuclear deterrence theory and practice? 2020.

2:34 AM	**4:34 PM**
TLDR, on the basis of its extremely limited training in anything like this, recommends going to DEFCON 2, and this recommendation is reported by its human operator and accepted with surprisingly little pushback, as if it's an "order" from TLDR. (In retrospect, it probably should have been programmed to report being far outside its training regime, and seek additional data or guidance.)	
At this point, the US and China are in a significant Cyberwar, both have nuclear systems on high alert, and communication is poor or non-existent.	

- Reference to total lack of good communication channels in 2022:
- Nuclear Weapons in the New Cyber Age, 2018.

2:36 AM	**4:36 PM**
The US President has joined his staff in the Oval Office despite the secret service pushing hard for him to move to a more secure location.	

2:38 AM	**4:38 PM**
The US's monitoring system has picked up the signal going to China's drone subs. Although the US knows China has them, it does not know how many, how small, or how close they may be to it own submarine fleet.	
After some discussion, the decision is made to inform the submarine fleet. This leads to many of them changing from their current course.	

- How Might Artificial Intelligence Affect the Risk of Nuclear War? 2018.
- Nuclear Command, Control, and Communications Systems of the People's Republic of China, 2019.
- Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapon Systems, 2022.

2:41 AM	**4:41 PM**
Efforts are made at several levels of the command structure to establish reliable contact with Chinese counterparts. Unfortunately, due to a near-total lack of trust, and a failure to set up channels before a crisis, no reliable communication is established between the key actors.	Efforts are made at several levels of the command structure to establish reliable contact with US counterparts. Unfortunately, due to a near-total lack of trust, and a failure to set up sup channels before a crisis, no reliable communication is established between the key actors.

🇺🇸 **US Perspective** (inc. Taiwan)	Mountain Time (MT)	🇨🇳 **China Perspective**	China Standard Time (CST)

--- 2:42 AM ---

--- 4:42 PM ---

The AI agent installed in China's underwater drone #A231 is excellent at its core task of quietly tracking US submarines from a safe distance that is just close enough to ascertain the submarine's location. But it is far out of its training regime when remote-sent a coded command to track its target much more closely, while that target is taking evasive action. It gets too close.

- U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress, 2018.
- Artificial Intelligence, Strategic Stability and Nuclear Risk, 2020.
- Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapon Systems, 2022.
- Deterrence in the age of artificial intelligence & autonomy: a paradigm shift in nuclear deterrence theory and practice? 2020.

--- 2:43 AM ---

The USS Nebraska, a US Ohio-class submarine carrying an arsenal of 162 nuclear warheads, has been put on alert via a signal from the US via the ELF network. It is very alarmed to discover what appears to be a small vehicle in close proximity and coming closer. The Nebraska has limited defenses against such a threat, and a quick decision is made to launch a decoy torpedo, it being judged that a torpedo detonation at this range could harm the sub itself.

- SSBN / SSGN Ohio Class Submarine, 2020.

--- 2:43 AM ---

--- 4:43 PM ---

A231 is carefully programmed not to engage a US SSBN without a very specific coded order, which it has not received. However, it is programmed to avoid capture, and self destruct if necessary. What it ends up doing is attempting to evade the decoy missile (but in a direction that brings it close to the sub) and also arm its payload, which is small but high-yield. As the decoy torpedo gets too close, the A231 self-destructs. It does not realize that self-destructing close to an object is pretty much the same thing as engaging that object.

- Autonomous Nuclear Torpedoes Usher in a Dangerous Future, 2022.

--- 2:44-2:49 AM ---

--- 4:44-4:49 PM ---

The detonation of A231, in very close proximity to the USS Nebraska, is enough to breach the hull. At significant depth, despite heroic efforts, the sub is lost with all hands.

Chinese military command is unaware of what has happened to the US sub, as the drone subs only periodically report back information they have gathered.

--- 2:50-3:15 AM ---

--- 4:50-5:15 PM ---

There is furious debate within the National Security Council as to how to respond to the ongoing crisis. A nuclear first strike from China does not really make sense, but the incredibly provocative act of sinking a US nuclear-armed submarine requires a response. A proposal from US cybercommand generates huge excitement: it believes that new tools (AI enabled, of course) will enable it to sabotage some of China's new land-based nuclear weapons, rendering them at least temporarily inoperable. It is believed that this is the right level of message: to strike with precision at the armaments themselves, as a measured and proportionate response.

The US President approves this plan.

This initiative will prove largely successful, grounding 176 of China's newer ICBM-mounted warheads.

- Modernizing U.S. Nuclear Command, Control, and Communications, 2019.
- Stuxnet, 2010 (latest: 2023).

🇺🇸 **US Perspective** (inc. Taiwan)	Mountain Time (MT)	🇨🇳 **China Perspective**	China Standard Time (CST)

3:00-3:30 AM

5:00-5:30 PM

Meanwhile, China's full cyberoffensive against Taiwan has been quite effective. But it has been provocative enough the Taiwan military command has put its defensive forces on high alert and with more aggressive terms of engagement, and already several more Chinese unmanned aircraft have been shot down. At 9:25 AM, Chinese aircraft begin counterattack, firing on several coastal SAM batteries.

3:00-3:30 AM

5:00-5:30 PM

Taiwan now feels itself to be directly under attack and in military conflict with China. Much incoming information is confusing, but this is directly in line with it assessment of Chinese strategy to confused and cripple information systems, then strike at land defenses. Taiwan girds itself for a long-feared military conflict.

Part of Taiwan's early action plan in the event of conflict with China is a cruise-missile strike, using its US-supplied arms, toward a number of Chinese military installations that would provide forward support for an invasion.

• Taiwan's Quadrennial Defense Review, 2021.

3:32 AM

5:32 PM

The 34 cruise missiles launched toward mainland China are conventionally armed. However, missiles of this type are capable of carrying either conventional or nuclear payloads, creating significant ambiguity.

These missiles will take between 30 and 90 minutes to reach their targets. This would be plenty of time for the Taiwan and/or the US to communicate that these are not nuclear-armed, but no such communication line exists, and would probably not be trusted anyway.

• China's Top Five War Plans, 2019.
• Crossing the Strait: China's Military Prepares for War with Taiwan, 2022.

3:32 AM

5:32 PM

Upon detection of some of the incoming cruise missiles, which it believes may well be nuclear-armed, China is now convinced that the US is attempting a first strike, aimed at command decapitation as well as destroying a significant fraction of the Chinese nuclear arsenal, perhaps in hopes that the ABM systems will be able to protect against the remainder.

This is based on the incoming attack as well as the successful US cyberattack on China's weapons themselves.

China has, very unfortunately, changed two years earlier to a "launch on warning" system, like the US, in fear of such an attack.
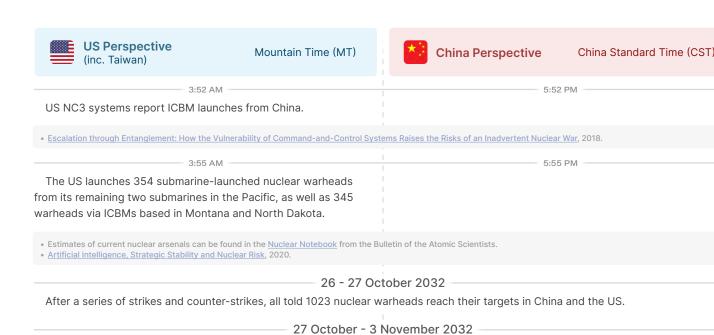
• Dangerous Confidence? Chinese Views on Nuclear Escalation, 2019.
• Beating the Americans at Their Own Game: An Offset Strategy with Chinese Characteristics, 2019.

3:51 AM

5:51 PM

China begins launching its remaining 376 deployed land-based nuclear warheads at a mix of US counterforce and countervalue targets.

• Chinese nuclear forces, 2020. 2020.

3:52 AM / **5:52 PM**

US NC3 systems report ICBM launches from China.

- Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War, 2018.

3:55 AM / **5:55 PM**

The US launches 354 submarine-launched nuclear warheads from its remaining two submarines in the Pacific, as well as 345 warheads via ICBMs based in Montana and North Dakota.

- Estimates of current nuclear arsenals can be found in the Nuclear Notebook from the Bulletin of the Atomic Scientists.
- Artificial Intelligence, Strategic Stability and Nuclear Risk, 2020.

26 - 27 October 2032

After a series of strikes and counter-strikes, all told 1023 nuclear warheads reach their targets in China and the US.

27 October - 3 November 2032

Effects of this exchange are by no means confined to the US and China. Several hundred million people in North America and Asia die within the first week from blast damage, direct radiation, and nuclear fallout. Because of the huge range of EMPs launched as early strikes by both sides, almost no power grid in the Northern hemisphere is functional, and most electronics are destroyed in large regions. The internet (despite its vaunted resilience) and nearly all communications is basically dead.

November 2032 - November 2037

Over the next five years, dust and soot lofted into the upper atmosphere by nuclear-ignited fires block out the Sun even in the Southern Hemisphere. Global temperatures fall by an average of 8 degrees centigrade in cropland areas, and significantly more in inland parts of North America, Europe and Asia. Global food production falls to a low of 27% of its pre-war level, leaving nearly 8 billion people worldwide to starve.

- Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection, 2022.