Christine Ye

OSPBER 77 – Final Paper

November 30, 2024

On War: Clausewitz, Great Power Competition, and the Superintelligence Arms Race

I. Introduction

In 1997, the computer program Deep Blue defeated world chess champion Garry Kasparov. In reaching superhuman performance at a cognitive task, Deep Blue signaled a new era for artificial intelligence (AI), in which machines can increasingly outperform humans. In 2012, AlexNet popularized today's deep learning paradigm: massive, billion-parameter neural networks trained on Internet-scale datasets. Since 2020, models like OpenAI's GPT have not only brought highly capable AI systems to mass markets, but also achieved better-than-human performance on tests of mathematical reasoning, language understanding, and graduate-level knowledge. Today's most capable models can predict protein structures [AlphaFold], solve complex software engineering tasks [SWE-Bench], complete end-to-end science experiments [Sakana's AI Scientist], and pass for humans in online conversations, satisfying the Turing Test.

The development of Artificial General Intelligence (AGI), AI that matches human performance on cognitive tasks, and Artificial SuperIntelligence (ASI), AI that greatly surpasses human cognitive abilities, appears increasingly plausible. In a 2023 survey of thousands of AI researchers, 50% predicted AGI would be achieved before 2047, and that full automation of human labor could be possible before 2116 (Grace et al., 2024). Industry leaders predict even shorter timelines: Google DeepMind CEO Shane Legg gives AGI a 50% chance by 2028,

Anthropic CEO Dario Amodei expects AGI as early as 2026, and OpenAI CEO Sam Altman thinks we may have superintelligence "in a few thousand days" (Amodei, 2024; Altman, 2024).

Artificial intelligence is governed by *scaling laws* – input data and computing power can mathematically predict the capability of the output model – meaning that future models will require massive scale-ups in investment and energy. From 2019 to 2023, the resources used to train OpenAI's GPT models increased by 3,000-10,000x, and compute investment is continuing to grow at 3-5x per year; Microsoft has already floated plans for a $100B computer cluster to launch in 2028 ("Notable AI Models," 2024). Following this trend, achieving AGI/ASI may demand the resources of entire corporations and even entire nation-states. If so, developing AI could become a new axis for Great Power competition.

If AGI/ASI systems can complete software engineering tasks, a capable-enough model could autonomously research and advance its own capabilities, kicking off *recursive self-improvement* and thus exponential growth (Bostrom, 2014). In this "intelligence explosion" scenario, even small advantages matter, and being the first party to reach a "critical" threshold takes on utmost importance: exponential growth means an early edge can become a massive multiple down the line (Aschenbrenner, 2024). Even if there are unforeseen bottlenecks, such as societal friction, that lead to a slower-than-exponential "takeoff" (Chollet, 2024), automating human-level cognitive labor will still accelerate the automation of manual labor, enabling structural change in both the digital and physical world.

And if AGI/ASI does come to fruition, it will reshape human economics, politics, and society, including the balance of global power. State governments are already keenly aware of AI's potential, and are increasingly important to the AI landscape; frontier AI companies already work with governments to shape regulation, security, and defense (Gress, 2024). The next

generation of infrastructure, investment, and regulations to build AGI/ASI will necessitate increasingly intense action from states.

II. The Political Motive: Kicking Off an AI Arms Race

It is likely that states will become the primary actors in the buildout of AI, equaling or surpassing corporations; understanding state-level *interaction* will thus be crucial to understanding the trajectory of superhuman AI. This paper will apply Clausewitz's theory of war to argue that an "AI arms race" is likely to develop, and to study the potential dynamics of escalation and de-escalation. Similar to the Cold War nuclear arms race, in the AI arms race, states iteratively engage in internal balancing through AI development, responding to the security dilemma posed by other states' AI capabilities; this could simultaneously include research efforts, bargaining, sanctions, and creation/deployment of AI weapons arsenals. As a real-world comparison, I will focus on the United States and China, which together host ~all of the world's top technology corporations.

Why would nations go to war over developing AI? In *On War,* Clausewitz conjectured that "war is a mere continuation of policy by other means" and a "real political instrument" (Clausewitz, 1832/1976). One motive is to secure the state's economic prosperity and global power. If AGI/ASI truly can automate large portions of the current economy, nations that control AI will profit massively. Forecasts for the economic impact of AI in 2030 range from ~hundreds of billions USD to over 100% year-over-year GDP growth (Lovely, 2024). Advanced AI is also expected to enable countless novel technologies, from medicine to energy and defense. Nations with control over AI will gain power not just from creating and selling these breakthroughs, but also from directly weaponizing technology against adversaries. If AGI/ASI enables massive progress in defense technology, possession of the AI gives a decisive military advantage

(Aschenbrenner, 2024). A second, closely linked reason is to propagate the state's political and belief system. Today, computers and the Internet enable people to interact, share information, and make decisions. Regimes regulate the Internet to shape citizens' lives and reality itself (Arendt, 1967); transmission cables and computer protocols are an extension of a state's political power. Control of AI will similarly shape states' control over political systems: AI could enable a regime to engage in surveillance and mass manipulation, or in better democratic discourse (Amodei, 2024). Finally, tied with emotions of national pride and desire for prestige, there might be an *existential* political motive: the relative strength of AI's haves over its have-nots could be great enough to threaten state self-determination.

Although these motives rest on theoretical projections about AI's capabilities, they have already manifested real-world rhetoric pointing towards an AI arms race. Currently, Chinese models lag just months to single-digit years behind the best American models: Hangzhou-based DeepSeek reproduced key parts of OpenAI's newest model just two months after release (Ng, 2024). China's State Council set the goal of making China "the world's primary AI innovation center" by 2030 (Lovely, 2024). U.S. AI policy has long pushed similarly competitive rhetoric, and the 2024 annual report from the US-China Economic and Security Review Commission (USCC) made a direct comparison to the nuclear arms race, calling on Congress to "establish and fund a Manhattan Project-like program dedicated to racing to and acquiring an Artificial General Intelligence" (Cleveland, 2024). Anthropic CEO Dario Amodei called for an "entente strategy" to ensure the "triumph of liberal democracy", where a U.S.-led coalition "seeks to gain a clear advantage on powerful AI" and thus "achieve robust military superiority" (Amodei, 2024). This rhetoric is also reflected in years of increasingly aggressive action: ever-growing investments, export controls on chip technology, or international infrastructure projects such as the Digital

Silk Road. Between the United States and China, an AI arms race is not just well-motivated in theory but also already under way.

III. Absolute War and the Dynamics of AI Acceleration

The specifics of AI's development and deployment will uniquely shape the dynamics and behaviors of AI acceleration and the arms race. In *On War*, Clausewitz's notion of *absolute war* follows from the idea that actions by either party "compels its opponent to follow suit", leading to "a reciprocal action… which must lead, in theory, to extremes" (Clausewitz, 1832/1976). As a result, all wars are reduced to their abstract extremes, and "the greatest effort must be exerted". However, in Clausewitz's view, *absolute war* never happens in the real world; *real war*, the dramatic and drawn-out process, occurs in practice. Clausewitz argues that this occurs for three reasons: *war is never an isolated act*, as it "never breaks out wholly unexpectedly" and cannot "be spread instantaneously"; *war does not consist of a single short blow*, because "the very nature of war impedes the simultaneous concentration of all forces"; and *in war the result is never final*, reducing "the vigour of the effort".

AGI/ASI, and the extremely powerful technology it enables, breaks all three assumptions. While internal balancing is somewhat predictable and escalates over time, sufficiently advanced technology could make launching attacks nearly instantaneous. This is particularly true if, as some have suggested, wartime decisions are increasingly made by superhuman AI models (Harper, 2024); this would remove even the emotional friction of applying force. Likewise, AI could enable the construction of novel arsenals of mass destruction, potentially allowing wars to be reduced to *single short blows*. If AI enables such weapons, it also promises to make the results of such a war final; the level of destruction, and the immense technological superiority of the victor, would permanently cripple challengers. These arguments are particularly true in an

intelligence explosion scenario, where the exponential growth of one party's technological strength allows it to swiftly overpower opposition; any sufficiently dominant technology could enable instantaneous and annihilation force. Thus in an intelligence explosion, Clausewitz's mechanisms for *real war* fail, and a war of theoretical extremes is plausible.

If an intelligence explosion occurs, we should expect the leader to gain solid technological dominance over time. However, another plausible scenario is one where multiple parties have relatively well-matched strength, and where an intelligence explosion is possible but not yet triggered. In early stages, the security dilemma will encourage both sides to steadily ramp up investment and prepare to match an adversary's capabilities, creating the arms race. The delicate balance of relative power, dictated by both sides' ever-growing capabilities, would force leaders to focus on reason and probability, rather than emotion, in playing out the future. Decisions from either side could then trigger further acceleration dynamics predicted by rationalism, particularly *brinkmanship*, *preventive war*, and *preemptive war*.

Rationalist theories of war predict preventive war when there is the expectation of changing relative power, so the presently-dominant nation wages war to reduce the adversary's strength and secure a better outcome (Frieden et al., 2016). A weaker state that acquires more powerful AI, will eventually overpower a stronger adversary; thus in an AI arms race, the stronger party has incentive to escalate early and weaken its adversary. Concretely, this is already reflected in Biden's chip war against China. Actions to weaken China's AI innovation by cutting off crucial supply chains aim to prevent China from catching up to American capabilities (Hsu, 2024). This dynamic would be particularly strong if actors expect a future intelligence explosion; for either nation, preventive action against the adversary's AI development could mean the difference between exponentially growing dominance or weakness. As a result, some researchers

predict state-actor attacks will become increasingly common in the next decade of innovation (Aschenbrenner, 2024). And while preventive strikes can occur even against today's limited AI capabilities, in the future, the first-strike advantage could also trigger preemptive war, where decisive benefits to offensive action motivate both parties to attack before the other (Frieden et al., 2016). This could move the world from an arms race to all-out war.

Even if preventive and preemptive war are avoided, the brinkmanship dynamic may still escalate states towards war. Although the state of AI research itself is still relatively transparent – most research is either published directly in academic papers or otherwise can be reproduced by those in the field – an AI arms race still poses a *credibility* problem. Building and weaponizing AGI/ASI is predicted to cost trillions USD, along with significant social and political costs associated with the rollout (Aschenbrenner, 2024). Although real military mobilizations are currently unlikely, both the United States and China are already investing on the order of hundreds of billions USD, making costly commitments to competing over AI. Aggressive rhetoric by American China hawks, and President Trump's purported desire to be "tough on China", has also upped the audience cost of competing. If this trajectory continues, the AI arms race will provoke increasingly costly and aggressive actions between states.

IV. Engineering A Technological Peace

Clausewitz doubted the efficacy of "kind-hearted" approaches to conflict, and his own writing does not focus on how peace might be maintained. Indeed, preventing escalation in an AI arms race seems very difficult: there are essentially no mechanisms to enforce any existing global agreements to limit AI development or weaponry, and strong incentives for individual states to break them. However, rationalist theory, which further develops Clausewitz's notion of reason as a governing principle for war, suggests negotiation is still possible. To reach a

settlement, though, states must overcome the challenges of *incomplete information*, *commitment problems*, and *indivisible goods* (Frieden et al., 2016). This kind of peace appears most achievable when a single state has a decisive technological advantage, effectively creating a hegemonic peace. A sufficiently strong state – especially in an intelligence explosion scenario, where this state's technological advantage increases exponentially over time – could clearly communicate its dominance and enforce negotiated peace agreements. As long as both parties are willing to compromise on a new world order, this technological hegemonic peace is possible, and the primacy of AI over other technology means it would be potentially permanent.

Acceleration towards an AI war could be also averted by normative shifts or sentiments related to *AI safety*, or the practice of minimizing potential risks of powerful AI, as has occurred with nuclear safety. Clausewitz and rationalist calculations would have assigned high likelihood to all-out nuclear war. After all, nuclear weapons have annihilative and nearly friction-free capabilities, and the "fog of war" increases the probability of mistakes, human emotions, and chance in triggering irreversible acts. However, nuclear war has been averted thus far, in part thanks to concern around nuclear existential risk. Many researchers believe that AGI/ASI could pose its own existential risk, by making inter-human conflict increasingly destructive or by models themselves manipulating humans for their own emergent goals (Brown, 2023). An accelerating AI arms race would vastly increase the probability of dangerous oversights and the level of AI existential risk; global efforts to slow down AI development and ensure safety might thus temper these competitive geopolitical dynamics. While this would require difficult-to-enforce peace treaties or global agreements, it could be feasible in a hegemonic world order, or if new technology enables better third-party supervision. Thus emotional and normative factors might act as a brake on acceleration dynamics.

V. Discussion

In this paper, I have used Clausewitz's theory of war to understand the motivation for a global AI arms race, the dynamics of escalating competition, and possible conditions for a "technological peace". However, my work omits several considerations which should be studied in greater depth. First, the view of states as the sole important actors is reductive; although states and inter-state competition will undoubtedly be important in AI's development, future research and thought will also be shaped by corporations, cultural dynamics, and individual decisions. In particular, if open-source AI remains competitive with private development, multipolar competition could become secondary to rogue actors or a decentralized system.

Second, there are reasons to believe the "AI arms race" itself may only be a perceived inevitability. On one hand, American politicians and industry leaders may have overestimated China's desire or capability to develop AGI/ASI (Lovely, 2024). For example, some have suggested that China's strict censorship of AI models, in line with its social harmony policies, could handicap its models' capabilities (Sheehan, 2023). On the other hand, some China experts argue that China possesses a "quiet confidence" with respect to its technological growth, and believes that while Chinese research "may not fully catch up", Western companies "just can't compete with" Chinese manufacturing (Hsu, 2024). Especially in a slow takeoff scenario, the rollout infrastructure may be what determines the balance of power; building AGI in isolation is then geopolitically worthless, reducing the motivation to race on research (Hsu, 2024). And if China is indeed less capable or motivated than perceived, an aggressive internal balancing approach would be an irrational and dangerous choice. American rhetoric of onshoring supply chains, reducing economic interdependence with China, and cutting off scientific exchange could be the *cause*, not the *response*, of the arms race.

In *On War,* Clausewitz – arguably the first philosopher to connect war to its political intentions, and to systematize the dynamics of conflict – put forward ideas that are extremely relevant to understanding the future of AI and international competition. While realistic consideration of acceleration dynamics is important for global security, overreliance on rational-play assumptions could make the AI arms race a dangerous self-fulling prophecy. Continued analysis that integrates Clausewitz, other theories of international relations, and reality of novel technology, will be necessary to safely steer state and global decisions.

Works Cited

Altman, S. (2024, September 23). *The Intelligence Age*. https://ia.samaltman.com/

Amodei, D. (2024, October). *Machines of Loving Grace*. https://darioamodei.com/machines-of-loving-grace

Arendt, H. (1967, February 25). Truth and Politics. *The New Yorker*.

Aschenbrenner, L. (2024). *Situational Awareness*. https://situational-awareness.ai/leopold-aschenbrenner/

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brown, S. (2023, May 23). Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI. *Ideas Made to Matter -- MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/why-neural-net-pioneer-geoffrey-hinton-sounding-alarm-ai

Chollet, F. (2017, November 27). The Implausibility of Intelligence Explosion. *Medium*. https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec

Cleveland, R., & Price, R. (2024, November). *2024 Annual Report to Congress*. U.S.-China Economic and Security Review Commission. https://www.uscc.gov/annual-report/2024-annual-report-congress

Echevarria II, A. J. (2007). *Clausewitz and Contemporary War*. Oxford University Press.

Frieden, J., Lake, D., & Schultz, K. (2016). Why Are There Wars? In *World Politics: Interests, Interactions, Institutions* (3rd ed.). W. W. Norton.

Grace, K., Stewart, H., Fabeinne Sandkuhler, J., Thomas, S., Weinstein-Raum, B., & Brauner, J. (2024). *Thousands of AI Authors on the Future of AI* (AI Impacts).

https://aiimpacts.org/wp-content/uploads/2023/04/Thousands_of_AI_authors_on_the_future_of_AI.pdf

Gress, M. (2024, November 7). Anthropic and Palantir Partner to Bring Claude AI Models to AWS for U.S. Government Intelligence and Defense Operations. *Palantir Investor Relations*. https://investors.palantir.com/news-details/2024/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations/

Harper, E. (2024, September 26). Will AI fundamentally alter how wars are initiated, fought and concluded? *International Council of the Red Cross*. https://blogs.icrc.org/law-and-policy/2024/09/26/will-ai-fundamentally-alter-how-wars-are-initiated-fought-and-concluded/

Hsu, S. (2024, November 21). Letter from Shanghai: Reflections on China in 2024—#73. *Manifold1*. https://www.manifold1.com/episodes/letter-from-shanghai-reflections-on-china-in-2024-73/transcript

Lovely, G. (2024, November 20). China Hawks are Manufacturing an AI Arms Race. *LessWrong*. https://www.lesswrong.com/posts/KPBPc7RayDPxqxdqY/china-hawks-are-manufacturing-an-ai-arms-race

Ng, A. (2024, November 27). Reasoning Revealed: DeepSeek-R1, a transparent challenger to OpenAI o1. *The Batch*. https://www.deeplearning.ai/the-batch/deepseek-r1-a-transparent-challenger-to-openai-o1/

Notable AI Models. (2024, June 19). *Epoch AI*. https://epoch.ai/data/notable-ai-models?view=table#explore-the-data

Sheehan, M. (2023). *China's AI Regulations and How They Get Made*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

von Clausewitz, C. (1832/1976). *On War* (M. Howard & P. Paret, Trans.). Oxford University Press.