GOVERNING AND REGULATING ARTIFICIAL INTELLIGENCE

RITWIK GUPTA
UC Berkeley

Content Warning

The topics discussed here today are controversial, political, and uncomfortable. They are presented as subjectively as possible.

This is a seminar — debate matters more than finishing the slides. Engage freely under Chatham House Rules.

DAMAGE ASSESSMENT UTILIZING R-CNN (GIRSCHIK, DONAHUE, DARRELL, MALIK)   UC BERKELEY & MICROSOFT

‘Lavender’: The AI machine directing Israel’s bombing spree in Gaza

The Israeli army has marked tens of thousands of Gazans as suspects for assassination, using an AI targeting system with little human oversight and a permissive policy for casualties, +972 and Local Call reveal.

Smoke rises after Israeli airstrikes in Beit Lahia, in the northern Gaza Strip, December 28, 2023. (Yonatan Sindel/Flash90)

The lecture presentation slides, also obtained by +972 and Local Call, contain illustrations of how the machine works: it is fed data about existing Hamas operatives, it learns to notice their features, and then it rates other Palestinians based on how similar they are to the militants.

“We rank the results and determine the threshold [at which to attack a target],” “Col. Yoav” said in the lecture, emphasizing that “eventually, people of flesh and blood take the decisions. In the defense realm, ethically speaking, we put a lot of emphasis on this. These tools are meant to help [intelligence officers] break their barriers.”
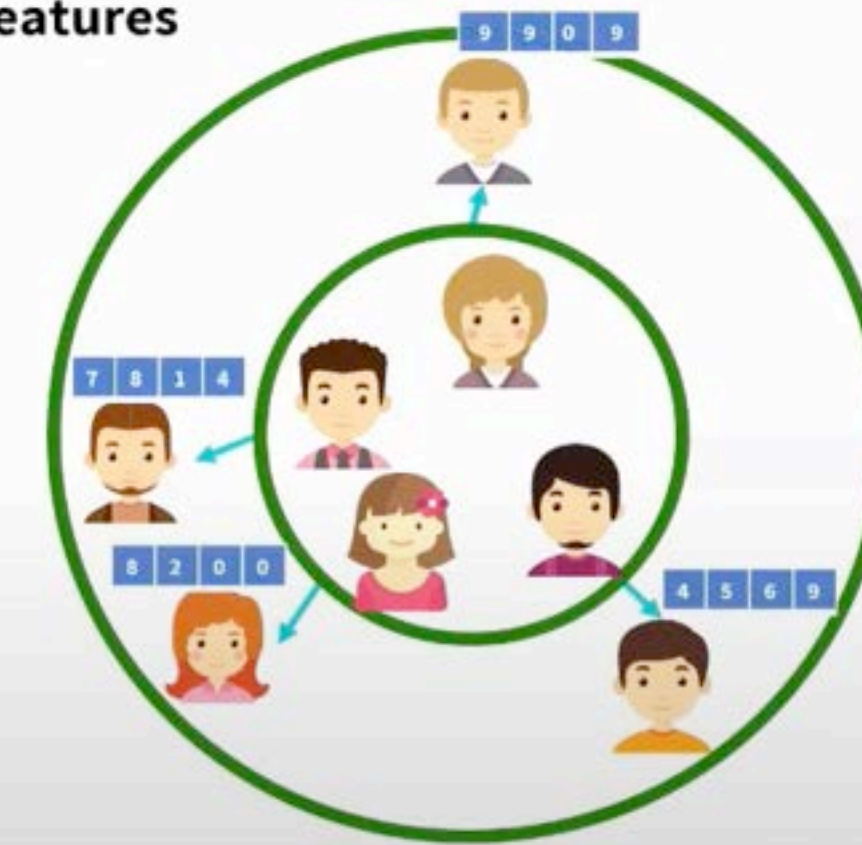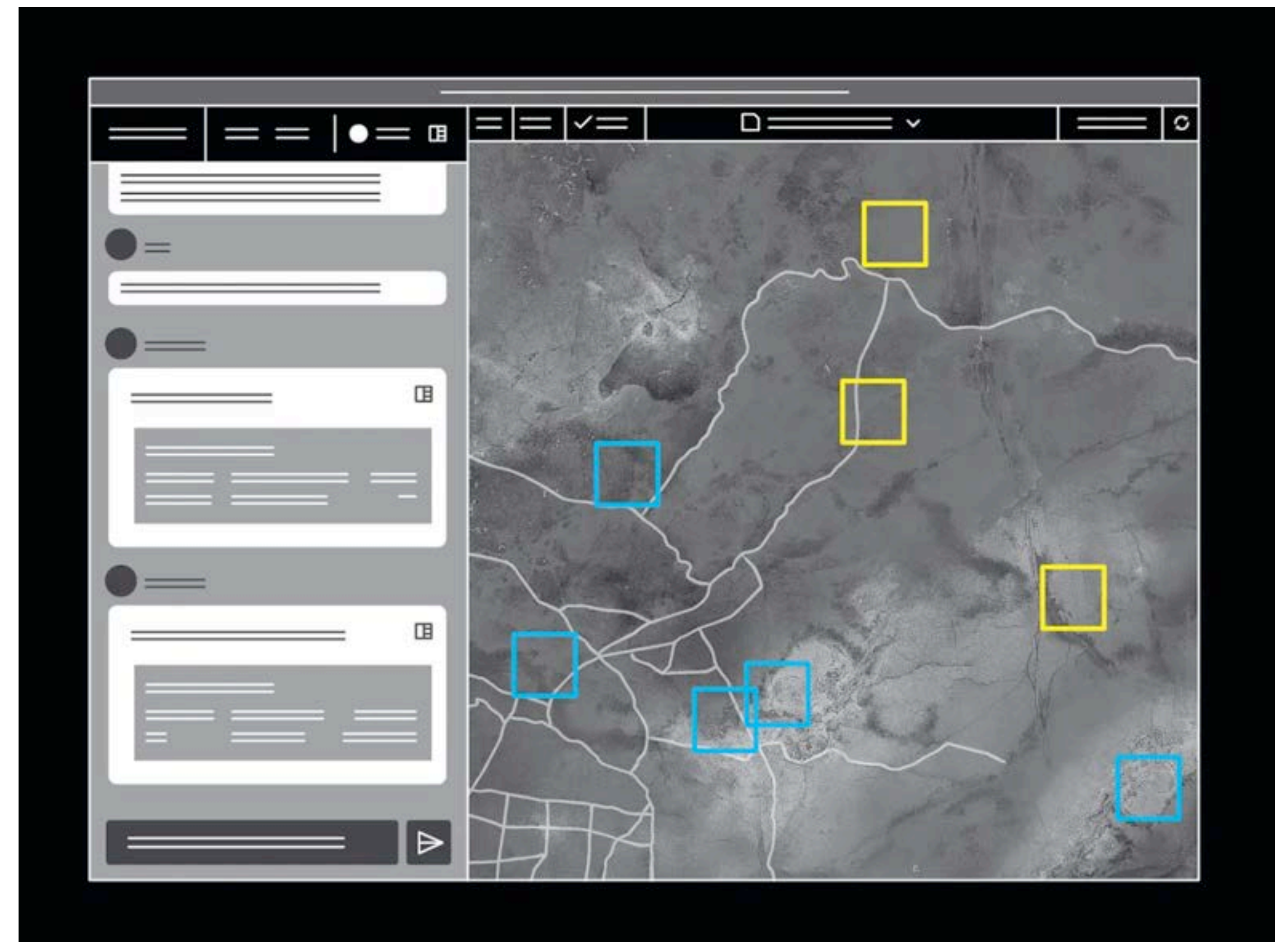
AI has a lot of potential for good and for bad

AI has a lot of potential

for good and for bad

(I posit that the same models can be used for both)

> Collection

> Search

> Sort by

> Status Date

> Jurisdiction

> Status

> Category

> Sector

Apply Filters   Clear All

4161 results:

Artificial Intelligence and Intellectual Property: Part II-Identity in the Age of AI

Collection: Congressional Hearings
Status date: Feb. 2, 2024
Status: Issued
Source: House of Representatives
Category:
Societal Impact
Data Governance
System Integrity (see reasoning)
Sector:
Politics and Elections
Government Agencies and Public Services
Judicial system (see reasoning)
Keywords (occurrence): artificial intelligence (11) machine learning (1) deepfake (4) show keywords in context

The Use of Artificial Intelligence at the Library of Congress, Government Publishing Office, and Smithsonian Institution

Collection: Congressional Hearings
Status date: Jan. 24, 2024
Status: Issued
Source: Senate
Category:
Societal Impact
Data Governance
System Integrity (see reasoning)
Sector:
Politics and Elections
Government Agencies and Public Services
Academic and Research Institutions (see reasoning)
Keywords (occurrence): artificial intelligence (12) machine learning (2) automated (1) deepfake (1) large language model (1)
show keywords in context

Artificial Intelligence and Intellectual Property: Part III-- IP Protection for Ai Assisted Inventions and Creative Works

Collection: Congressional Hearings
Status date: April 10, 2024
Status: Issued
Source: House of Representatives

BIDEN EXECUTIVE ORDER ON AI

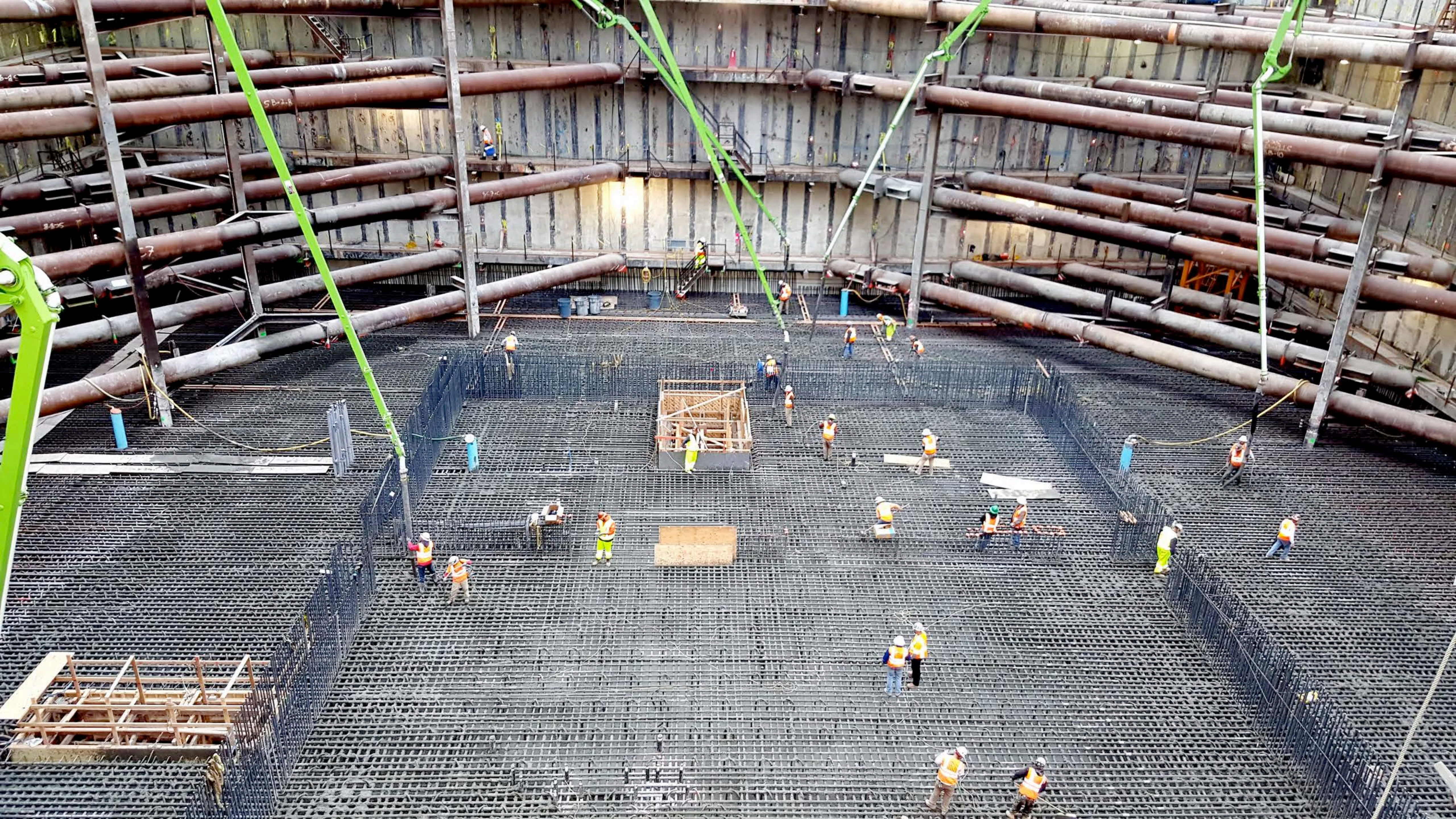The White House

DATA-CENTRIC AI GOVERNANCE

UC Berkeley

COMPUTE GOVERNANCE

Harvard

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

(k) The term "dual-use foundation model" means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

(k) The term "dual-use foundation model" means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by:

(i) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons;

(ii) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or

(iii) permitting the evasion of human control or oversight through means of deception or obfuscation.

	Terms	SSL	Large data	FLOPs	Params.
Terms					
Bommasani et al. (2022)	Foundation	✓	✓	–	–
Anderljung et al. (2023)	Foundation, Frontier	✓	✓	$> 10^{26}$	–
Alstott (2023)	Frontier	–	–	$> 10^{26}$	–
Governance					
The White House (2023)	Foundation, "Dual-Use"[2]	✓	✓	$> 10^{26}$	> 10B
Romney et al. (2024)	Frontier, General Purpose	–	–	$> 10^{26}$	–
European Union (2024)	General Purpose	✓	✓	$> 10^{25}$	> 1B
Wiener et al. (2024)	Frontier	–	–	$> 10^{25} / 10^{26}$	–

Table 1: Variance in model definitions across policy documents.

TSMC and Samsung Dominate Chip Foundry Business

Market share of semiconductor foundries* by revenue (in %)

Legend:
- TSMC 🇹🇼
- Samsung Foundry** 🇰🇷
- UMC 🇹🇼
- GlobalFoundries 🇺🇸
- SMIC 🇨🇳
- Others

Values for Q1 2024:
- Others: 9
- SMIC: 6
- GlobalFoundries: 5
- UMC: 6
- Samsung Foundry: 13
- TSMC: 62

X-axis: Q2 2021 — Q1 2022 — Q1 2023 — Q1 2024
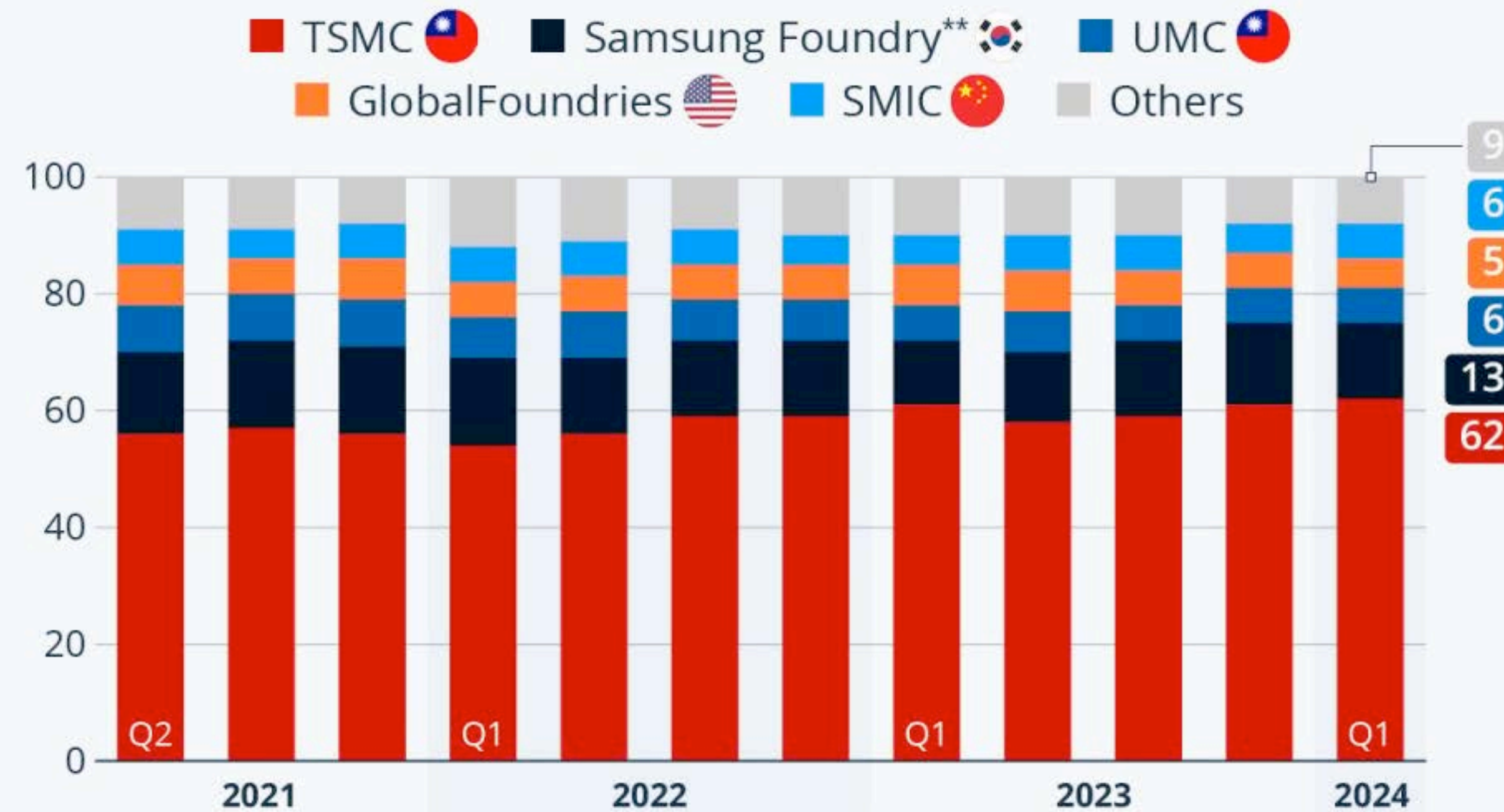
*   companies that only produce semiconductors for other companies who don't have the necessary facilities
** including internal foundry production

Source: Counterpoint Research

statista ⎇

The Ambitious Dragon

Beijing's Calculus for Invading Taiwan by 2030

MAJ KYLE AMONSON, US ARMY
CAPT DANE EGLI, US COAST GUARD, RETIRED

Abstract

Chinese president Xi Jinping has a strategic window, in the 2030 timeframe, when favorable conditions exist to forcefully annex Taiwan if peaceful unification is not achieved before then. This hypothesis is based upon the fact that an emboldened China intends to fulfill its imperial—and geostrategic—objectives through expansionist behavior against Taiwan. The three main factors examined are (1) President Xi's "cult of personality" as a totalitarian leader to support the *why* of the invasion timeline, (2) the People's Liberation Army (PLA) defense modernization as an enabling planning factor, and (3) Chinese demographics against the backdrop of domestic election cycles and President Xi's life expectancy. These three factors offer a strategic harbinger that, if President Xi continues to pursue annexation of Taiwan, the PLA will be prepared by 2027, and he will likely take steps to realize these ambitions by 2030 as China's population ages, while pursuing unification to solidify his historic legacy in his lifetime. This article will begin with an overview of the current geopolitical tensions, provide an explanation for the fundamental factors contributing to President Xi's window of opportunity, and conclude by providing an integrated assessment of relevant global security.

***

*Taiwan is China's Taiwan. Resolving the Taiwan question is a matter for the Chinese, a matter that must be resolved by the Chinese. We will continue to strive for peaceful reunification with the greatest sincerity and the utmost effort, but **we will never promise to renounce the use of force, and we reserve the option of taking all measures necessary**...*(emphasis added)

—Xi Jinping

On 23 October 2022, the People's Republic of China (PRC) announced that Chinese President Xi Jinping would serve an unprecedented third term leading the Chinese Communist Party (CCP). It is noteworthy that during his speech to the 20th National Congress of the CCP, President Xi repeatedly reinforced the narrative that "complete reunification of our country must be realized, and it can, without a doubt, be realized."[1] He followed this bellicose statement by asserting that the PRC "reserves the option to take all measures necessary."[2] The international community must not only assess *if* President

Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification

A Rule by the Industry and Security Bureau on 10/13/2022

PUBLISHED DOCUMENT: 2022-21658 (87 FR 62186)

DOCUMENT HEADINGS

Department of Commerce
Bureau of Industry and Security
15 CFR Parts 734, 736, 740, 742, 744, 762, 772, and 774
[Docket No. 220930-0204]
RIN 0694-AI94

AGENCY:

Bureau of Industry and Security, Department of Commerce.

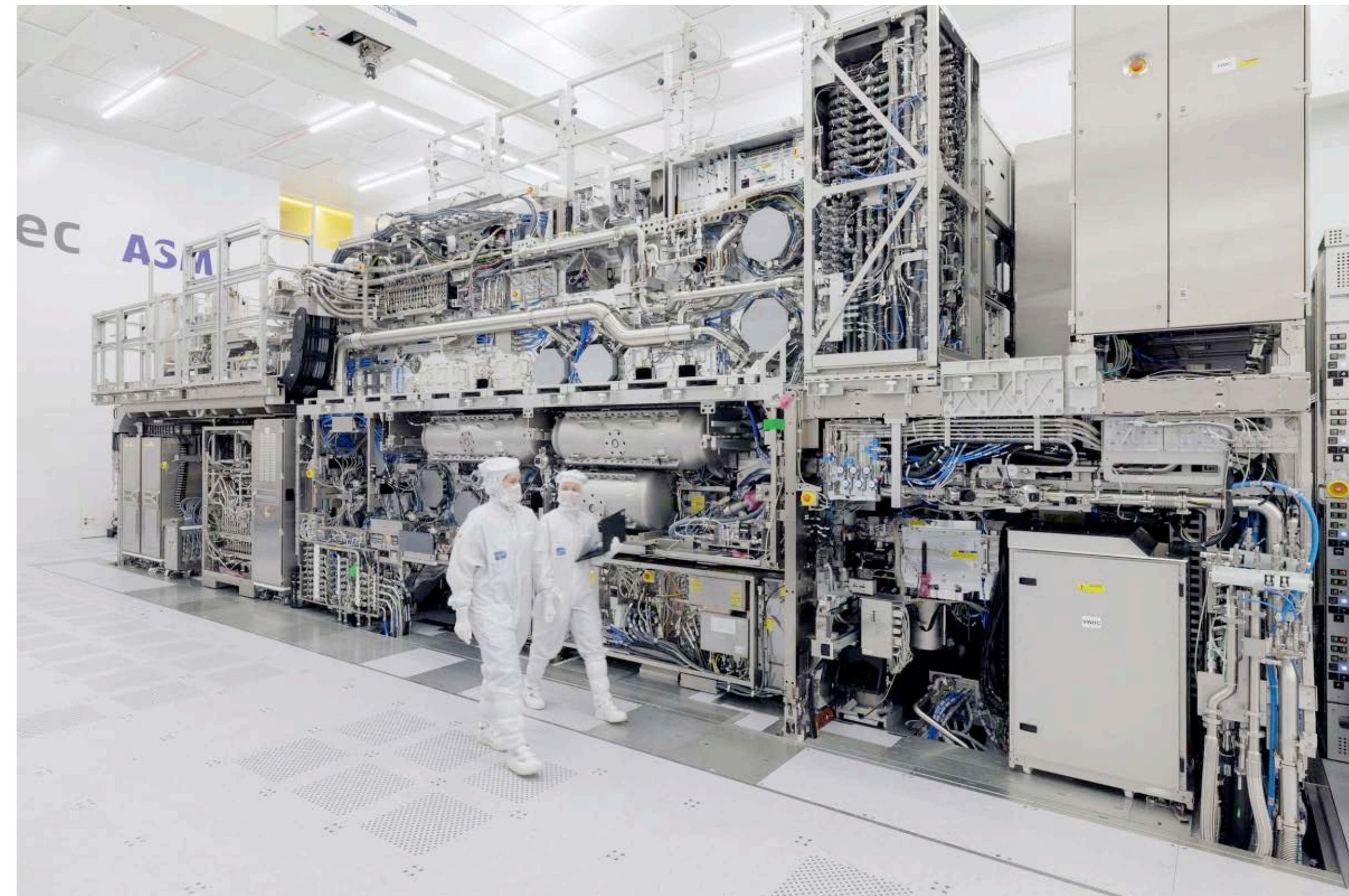**ACTION:**

Interim final rule; request for comments.

SUMMARY:

In this rule, the Bureau of Industry and Security (BIS) is amending the Export Administration Regulations (EAR) to implement necessary controls on advanced computing integrated circuits (ICs), computer commodities that contain such ICs, and certain semiconductor manufacturing items. In addition, BIS is expanding controls on transactions involving items for supercomputer and semiconductor manufacturing end uses, for example, this rule expands the scope of foreign-produced items subject to license requirements for twenty-eight existing entities on the Entity List that are located in China. BIS is also informing the public that specific activities of "U.S. persons" that 'support' the "development" or "production" of certain ICs in the PRC require a license. Lastly, to minimize short term impact on the semiconductor supply chain from this rule, BIS is establishing a Temporary General License to permit specific, limited manufacturing activities in China related to items destined for use outside China and is identifying a model certificate that may be used in compliance programs to assist, along with other measures, in conducting due diligence. `

DATES:

Integrated circuits are controlled under the new ECCN 3A090 if:

- Aggregate bidirectional transfer rates are 600 Gbyte/s or more, or
- Digital processors can execute >= 4,800 TOPS (trillion operations per second), or
- Digital primitive computational units can execute >= 4,800 TOPS, or
- Analog/multi-level computational units can execute >= 38,400 TOPS

Integrated circuits are controlled under the new ECCN 3A090 if:

- Aggregate bidirectional transfer rates are 600 Gbyte/s or more, or
- Digital processors can execute >= 4,800 TOPS (trillion operations per second), or
- Digital primitive computational units can execute >= 4,800 TOPS, or
- Analog/multi-level computational units can execute >= 38,400 TOPS
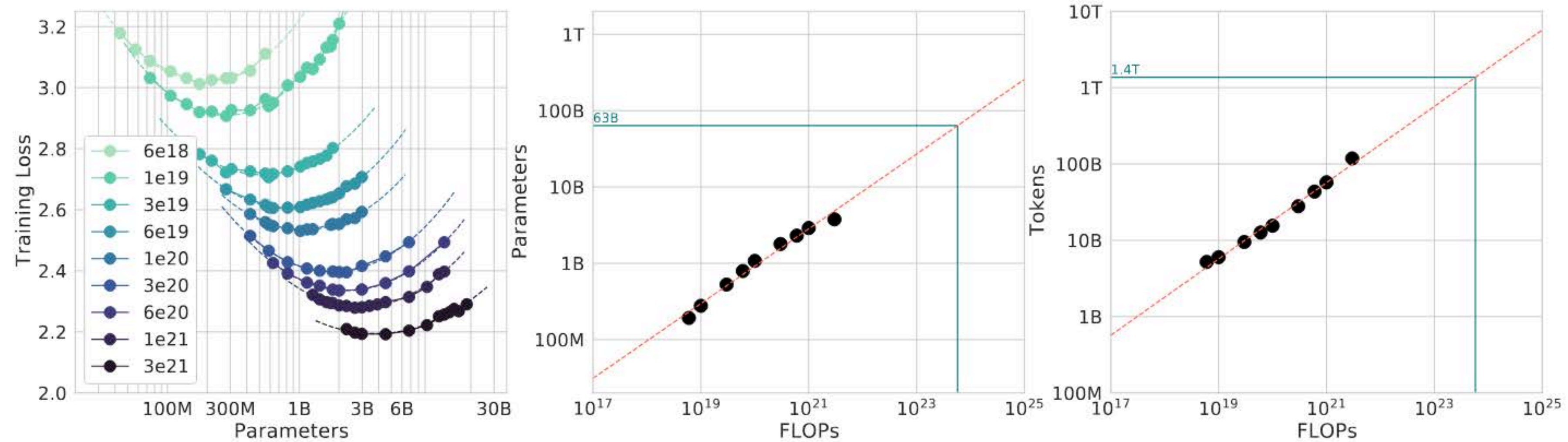
Figure 3 | **IsoFLOP curves.** For various model sizes, we choose the number of training tokens such that the final FLOPs is a constant. The cosine cycle length is set to match the target FLOP count. We find a clear valley in loss, meaning that for a given FLOP budget there is an optimal model to train (**left**). Using the location of these valleys, we project optimal model size and number of tokens for larger models (**center** and **right**). In green, we show the estimated number of parameters and tokens for an *optimal* model trained with the compute budget of *Gopher*.

Accelerating the Evolution of AI Export Controls

RITWIK GUPTA, ANDREW W. REDDIE / SEP 21, 2023

Ritwik Gupta is a computer vision Ph.D. student at the University of California, Berkeley, in the Berkeley AI Research Lab and a fellow at the Berkeley Risk and Security Lab and Berkeley Center for Security in Politics. He serves as the Deputy Technical Director for Autonomy at the Defense Innovation Unit. Andrew W. Reddie is an Associate Research Professor at the University of California, Berkeley's Goldman School of Public Policy, and Founder and Faculty Director of the Berkeley Risk and Security Lab.

Artificial intelligence is poised to revolutionize economies, scientific research, and the power dynamics between governments and industries. As AI research expands and its capabilities grow, a nation's prowess in AI will increasingly signify its overall strength.

Integrated circuits are controlled under the new ECCN 3A090 if:

- Digital processors can execute >= 4,800 TOPS (trillion operations per second), or

---

further China's goals of surpassing the military capability of the United States and its allies, a goal noted in the February 6, 2023 Annual Threat Assessment of the U.S. Intelligence Community. That same report indicated that "China is rapidly expanding and improving its artificial intelligence (AI) and big data analytics capabilities, which could expand beyond domestic use." These national security concerns were paramount in the issuance of this AC/S IFR.

Consistent with the national security and foreign policy concerns described in the October 7 IFR, BIS is updating the EAR to enhance effectiveness of the controls in addressing these ongoing concerns. Following the implementation of the controls last year, BIS continued to study and assess their effectiveness. This rule strengthens and improves those controls by addressing the national security considerations that have come to light through open-source reporting, public comments, and the intelligence community. Through this process, BIS learned that certain additional ICs could provide nearly comparable AI model training capability as those controlled in the October 7 IFR BIS also seeks to further impair diversion channels through third countries, particularly those with AI commercial and research ties to the PRC.

In addition, credible open source reporting identified PRC companies using foreign subsidiaries to purchase chips subject to EAR controls, and accessing and operating datacenters located outside of the PRC with the ICs subject to EAR controls. Moreover, BIS is also concerned about certain additional ICs, which in turn can be used to train frontier AI models that have the most significant potential for advanced warfare applications, including unmanned intelligent combat systems, enhanced battlefield situational awareness and decision making, multidomain operations, automatic target recognition, autopiloting, missile fusion, precise guidance for hypersonic platforms, and cyber attacks. Accordingly, to address these issues, BIS is making several changes to the rule.

Reuters

World | US Election | Business | Markets | Sustainability | Legal | Breakingviews | Technology | More

China's military and government acquire Nvidia chips despite US ban

By Eduardo Baptista

January 15, 2024 12:01 AM PST · Updated 10 months ago

THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Audio | Latest Headlines | More

Latest World Business U.S. Politics Economy Tech Markets & Finance Opinion Arts Lifestyle Real Estate Pers

TECHNOLOGY | ARTIFICIAL INTELLIGENCE   Follow

China's AI Engineers Are Secretly Accessing Banned Nvidia Chips

Brokers are making overseas computing power available and offering a high degree of anonymity

tom's HARDWARE

US Edition

Home | Best Picks | Raspberry Pi | CPUs | GPUs | 3D Printers | News | Coupons | More | Forums

TRENDING   Black Friday SSD Deals Live   AMD Ryzen 7 9800X3D   Black Friday Monitor Deals Live   Core Ultra

Tech Industry

Chinese AI Firm Stockpiled 18 Months of Nvidia GPUs Before Export Ban

News By Anton Shilov published November 10, 2023

There are loads of unused Nvidia AI and HPC GPUs in China, it seems.

Comments (1)

BRSL

AI Frontiers Initiati

Whack-a-Chip: The Futility of Hardware-Centric Export Controls

Ritwik Gupta[1,2,*] Leah Walker[1] and Andrew W. Reddie[1]

[1]Berkeley Risk and Security Lab, University of California, Berkeley and [2]Berkeley AI Research Lab, University of California, Berkeley
*Corresponding author: ritwikgupta@berkeley.edu

U.S. export controls on semiconductors are widely known to be permeable, with the People's Republic of China (PRC) steadily creating state-of-the-art artificial intelligence (AI) models with exfiltrated chips. This paper presents the first concrete, public evidence of how leading PRC AI labs evade and circumvent U.S. export controls. We examine how Chinese companies, notably Tencent, are not only using chips that are restricted under U.S. export controls but are also finding ways to circumvent these regulations by using software and modeling techniques that maximize less capable hardware. Specifically, we argue that Tencent's ability to power its Hunyuan-Large model with non-export controlled NVIDIA H20s exemplifies broader gains in efficiency in machine learning that have eroded the moat that the United States initially built via its existing export controls. Finally, we examine the implications of this finding for the future of the United States' export control strategy.

arXiv:2411.14425v1 [cs.CY] 21 Nov 2024

Introduction

In an effort to maintain a technical edge, the United States has sought to limit the People's Republic of China's (PRC) artificial intelligence (AI) research and development by pursuing export controls on hardware (notably chips). As part of this strategy, the United States has created a series of limits on AI hardware procurement including not only chips but also the manufacturing equipment needed to create them [2]. This strategy reflects the assumption that limiting a country's access to cutting-edge computing hardware limits the ability to create cutting-edge "AI" and other advancements. In reality, we argue that both the strategy and its underlying assumption have proven unreliable.

First, major Chinese AI firms—despite the export controls that they face—are still managing to access high-end export-controlled chips. Second, Chinese AI labs have leveraged advancements in machine learning (ML) training tools to successfully train state-of-the-art (SOTA) models on lower quality, non-export controlled chips (including NVIDIA's H20 GPUs), demonstrating that an export strategy based on hardware thresholds can be overcome through better software. The latter reality reflects an evolution in the field of machine learning at large and illustrates the importance of keeping abreast of the cutting edge of machine learning and compute efficiency as part of forecasting competitor capability gains.

The strategy of overcoming limited hardware resources with better software is not unique to the PRC—American academic labs are bellwethers for this phenomenon. Both are resorting to clever optimizations [15] due to their inability to source high-quality chips against better-resourced organizations. Interestingly, these trends in academic computing can also offer a useful proxy to understand ways in which the PRC might achieve its AI goals in spite of an unreliable stream of advanced hardware.

In light of these challenges, this paper examines recent AI models released by Chinese companies and examines evidence of Chinese firms using export-controlled chips and cases of efficiency gains with non-export-controlled chips. Specifically, this paper focuses on case studies and examples from Tencent, the Chinese video game juggernaut and Chinese leader in AI research and development. This work represents the first public analysis demonstrating how Chinese companies are successfully using NVIDIA's H20 GPUs, explicitly allowed under U.S.

export controls, to train state-of-the-art AI, including Tencent's recently released Hunyuan-Large Model. We highlight Tencent's previous use of cutting-edge, export-controlled chips like the NVIDIA A100 for its research and show how codebase signals can reveal which chips are used for training.

While there has been previous reporting on the porous nature of the U.S. export regime [41, 19, 28], this paper takes a deeper dive, focusing on the methods and techniques enabling Chinese firms to overcome its reliance on export-controlled chips and the acquisition and use of restricted microelectronics. Finally, we analyze the broader implications of Beijing's evolving evasion strategies.

Chinese Firms are Outpacing Hardware Restrictions

On November 4, 2024, the Chinese tech giant Tencent released "Hunyuan-Large," the largest, open-source, transformer-based mixture-of-experts model which achieves state-of-the-art performance on multiple downstream tasks [37].

Despite its state-of-the-art nature and size, a close examination of Hunyuan-Large reveals that it was not trained on state-of-the-art hardware. Rather Hunyuan-Large was trained on an export control-compliant chip, the NVIDIA H20, a feat made possible by a series of techniques used to maximize the value of throttled hardware.

Hunyuan-Large: A SOTA, Open-Source LLM

Hunyuan-Large represents Tencent's newest, open-source LLM. When compared to other state-of-the-art LLMs such as Meta's Llama 3.1 [20], Mixtral-8x22B [1], and DeepSeek-V2 [7], it achieves state-of-the-art performance on multiple downstream benchmarks such as MMLU [13] and CommonsenseQA [38].

On November 5, 2024, Hunyuan-Large's training codebase[1] and weights[2] were openly released by Tencent on GitHub and Hugging Face,

---

[1] https://github.com/RitwikGupta/Tencent-Hunyuan-Large. We link to a forked version of the repository to ensure that it is frozen in place, as analyzed.
[2] https://huggingface.co/tencent/Tencent-Hunyuan-Large

1

BRSL

Whack-a-Chip: The Futility of Hardware-Centric Export Controls

Ritwik Gupta[1,2,*] Leah Walker[1] and Andrew W. Reddie[1]

[1]Berkeley Risk and Security Lab, University of California, Berkeley and [2]Berkeley AI Research Lab, University of California, Berkeley
*Corresponding author: ritwikgupta@berkeley.edu

U.S. export controls on semiconductors are widely known to be permeable, with the People's Republic of China (PRC) steadily creating state-of-the-art artificial intelligence (AI) models with exfiltrated chips. This paper presents the first concrete, public evidence of how leading PRC AI labs evade and circumvent U.S. export controls. We examine how Chinese companies, notably Tencent, are not only using chips that are restricted under U.S. export controls but are also finding ways to circumvent these regulations by using software and modeling techniques that maximize less capable hardware. Specifically, we argue that Tencent's ability to power its Hunyuan-Large model with non-export controlled NVIDIA H20s exemplifies broader gains in efficiency in machine learning that have eroded the moat that the United States initially built via its existing export controls. Finally, we examine the implications of this finding for the future of the United States' export control strategy.

Introduction

In an effort to maintain a technical edge, the United States has sought to limit the People's Republic of China's (PRC) artificial intelligence (AI) research and development by pursuing export controls on hardware (notably chips). As part of this strategy, the United States has created a series of limits on AI hardware procurement including not only chips but also the manufacturing equipment needed to create them [2]. This strategy reflects the assumption that limiting a country's access to cutting-edge computing hardware limits the ability to create cutting-edge "AI" and other advancements. In reality, we argue that both the strategy and its underlying assumption have proven unreliable.

First, major Chinese AI firms—despite the export controls that they face—are still managing to access high-end export-controlled chips. Second, Chinese AI labs have leveraged advancements in machine learning (ML) training tools to successfully train state-of-the-art (SOTA) models on lower quality, non-export controlled chips (including NVIDIA's H20 GPUs), demonstrating that an export strategy based on hardware thresholds can be overcome through better software. The latter reality reflects an evolution in the field of machine learning at large and illustrates the importance of keeping abreast of the cutting edge of machine learning and compute efficiency as part of forecasting competitor capability gains.

The strategy of overcoming limited hardware resources with better software is not unique to the PRC—American academic labs are bellwethers for this phenomenon. Both are resorting to clever optimizations [15] due to their inability to source high-quality chips against better-resourced organizations. Interestingly, these trends in academic computing can also offer a useful proxy to understand ways in which the PRC might achieve its AI goals in spite of an unreliable stream of advanced hardware.

In light of these challenges, this paper examines recent AI models released by Chinese companies and examines evidence of Chinese firms using export-controlled chips and cases of efficiency gains with non-export-controlled chips. Specifically, this paper focuses on case studies and examples from Tencent, the Chinese video game juggernaut and Chinese leader in AI research and development. This work represents the first public analysis demonstrating how Chinese companies are successfully using NVIDIA's H20 GPUs, explicitly allowed under U.S.

export controls, to train state-of-the-art AI, including Tencent's recently released Hunyuan-Large Model. We highlight Tencent's previous use of cutting-edge, export-controlled chips like the NVIDIA A100 for its research and show how codebase signals can reveal which chips are used for training.

While there has been previous reporting on the porous nature of the U.S. export regime [41, 19, 28], this paper takes a deeper dive, focusing on the methods and techniques enabling Chinese firms to overcome its reliance on export-controlled chips and the acquisition and use of restricted microelectronics. Finally, we analyze the broader implications of Beijing's evolving evasion strategies.

Chinese Firms are Outpacing Hardware Restrictions

On November 4, 2024, the Chinese tech giant Tencent released "Hunyuan-Large," the largest, open-source, transformer-based mixture-of-experts model which achieves state-of-the-art performance on multiple downstream tasks [37].

Despite its state-of-the-art nature and size, a close examination of Hunyuan-Large reveals that it was not trained on state-of-the-art hardware. Rather Hunyuan-Large was trained on an export control-compliant chip, the NVIDIA H20, a feat made possible by a series of techniques used to maximize the value of throttled hardware.
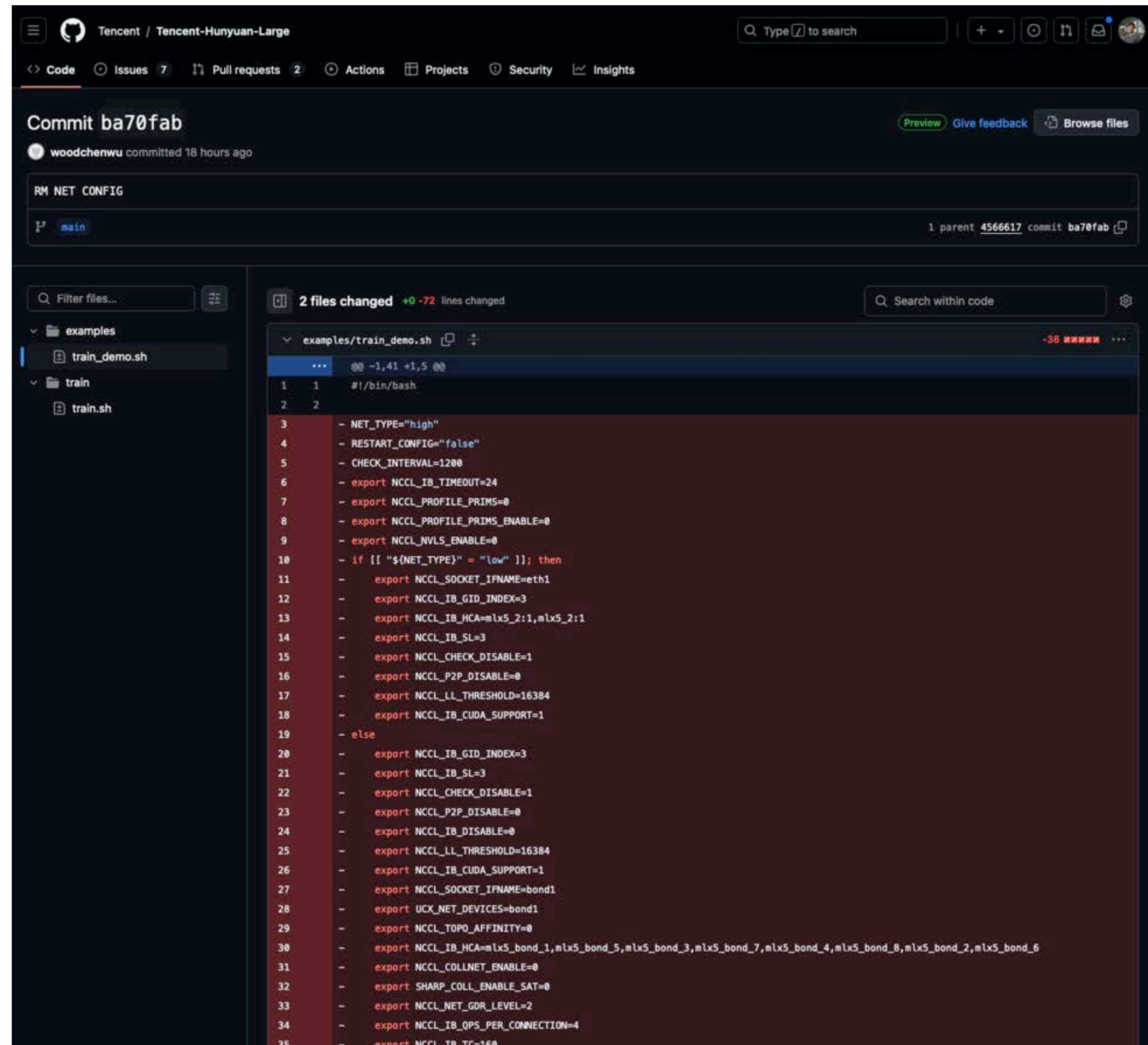
Hunyuan-Large: A SOTA, Open-Source LLM

Hunyuan-Large represents Tencent's newest, open-source LLM. When compared to other state-of-the-art LLMs such as Meta's Llama 3.1 [20], Mixtral-8x22B [1], and DeepSeek-V2 [7], it achieves state-of-the-art performance on multiple downstream benchmarks such as MMLU [13] and CommonsenseQA [38].

On November 5, 2024, Hunyuan-Large's training codebase[1] and weights[2] were openly released by Tencent on GitHub and Hugging Face,

---

[1] https://github.com/RitwikGupta/Tencent-Hunyuan-Large. We link to a forked version of the repository to ensure that it is frozen in place, as analyzed.

[2] https://huggingface.co/tencent/Tencent-Hunyuan-Large

Tencent / Tencent-Hunyuan-Large

Code | Issues 7 | Pull requests 2 | Actions | Projects | Security | Insights

Commit ba70fab

woodchenwu committed 18 hours ago

RM NET CONFIG

main — 1 parent 4566617 commit ba70fab

2 files changed +0 −72 lines changed

examples
  train_demo.sh
train
  train.sh

examples/train_demo.sh — −36

```bash
@@ -1,41 +1,5 @@
 #!/bin/bash

-NET_TYPE="high"
-RESTART_CONFIG="false"
-CHECK_INTERVAL=1200
-export NCCL_IB_TIMEOUT=24
-export NCCL_PROFILE_PRIMS=0
-export NCCL_PROFILE_PRIMS_ENABLE=0
-export NCCL_NVLS_ENABLE=0
-if [[ "${NET_TYPE}" = "low" ]]; then
-    export NCCL_SOCKET_IFNAME=eth1
-    export NCCL_IB_GID_INDEX=3
-    export NCCL_IB_HCA=mlx5_2:1,mlx5_2:1
-    export NCCL_IB_SL=3
-    export NCCL_CHECK_DISABLE=1
-    export NCCL_P2P_DISABLE=0
-    export NCCL_LL_THRESHOLD=16384
-    export NCCL_IB_CUDA_SUPPORT=1
-else
-    export NCCL_IB_GID_INDEX=3
-    export NCCL_IB_SL=3
-    export NCCL_CHECK_DISABLE=1
-    export NCCL_P2P_DISABLE=0
-    export NCCL_IB_DISABLE=0
-    export NCCL_LL_THRESHOLD=16384
-    export NCCL_IB_CUDA_SUPPORT=1
-    export NCCL_SOCKET_IFNAME=bond1
-    export UCX_NET_DEVICES=bond1
-    export NCCL_TOPO_AFFINITY=0
-    export NCCL_IB_HCA=mlx5_bond_1,mlx5_bond_5,mlx5_bond_3,mlx5_bond_7,mlx5_bond_4,mlx5_bond_8,mlx5_bond_2,mlx5_bond_6
-    export NCCL_COLLNET_ENABLE=0
-    export SHARP_COLL_ENABLE_SAT=0
-    export NCCL_NET_GDR_LEVEL=2
-    export NCCL_IB_QPS_PER_CONNECTION=4
-    export NCCL_IB_TC=160
```