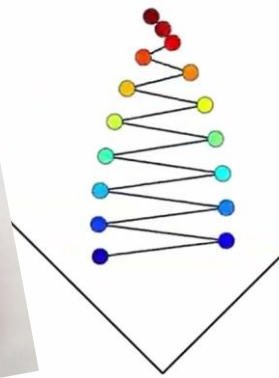


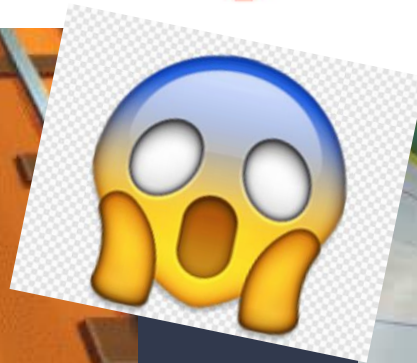
Compute Governance

Day 4.

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.



Compute governance !!!!!



What is compute ?

“Compute” = computing operations

The research and policy subfield focused on governing AI chips is often called **“compute governance”**

Fundamental properties of compute

- **Excludability**
 - Can't be used simultaneously by many actors
 - Access can be restricted (export control, cloud computing)
- **Quantifiability**
 - Physical (not like algorithms or data) : requires physical space, energy usage
 - Can keep track of number of operations computed
- **Necessity**
 - Required for AI development & deployment

Motivation

If governments could regulate the large-scale use of “AI chips,” that would likely enable them to govern frontier AI development

- “AI chips” to refer to cutting-edge, AI-specialized computer chips (like NVIDIA’s A100 and H100 or Google’s TPUv4)
- Frontier AI models like GPT-4 are already trained using tens of thousands of AI chips, and trends suggest that more advanced AI will require even more computing power

Motivation

Governments can likely regulate the large-scale use of AI chips

- Hard to enforce regulations on data or algorithms, because it is very easy to copy, transmit, or store them
- In contrast, AI chips are physical in nature, so governments can more easily track and restrict access to AI chips -> Regulation of AI chips is relatively feasible
- Extremely complex, global supply chain, in which a small number of companies and countries dominate key steps -> Coalitions of only a few actors could require other actors to meet safety standards in order to import AI chips. Any state would likely find it extremely difficult to manufacture AI chips on their own
- Since AI chips are so specialized, they can be regulated without regulating the large majority of computer chips

Main approaches

- **Monitoring compute usage:** By leveraging the quantifiable nature of compute, we can track usage and identify potentially high-risk systems.
 - Verification, arms control, transparency, assurances
 - Index for policies (audit on modes with compute > X)
- **Restricting compute usage:** The excludable property of compute allows for the possibility of denying access to certain resources (ex US's export restrictions on AI chips)
 - Emergency brakes, assurances, avoiding diffusion, limiting number of actors, regulating concerning training runs
- **Promoting compute usage:** In contrast to restriction, this strategy involves providing subsidized access to resources.
 - Sponsoring compute to promote safety research
 - Overall boosting safety conscious actors

Potential limitations

Governance of large numbers of AI chips faces large potential limitations

- Due to hardware progress and algorithmic advances, the number and sophistication of chips needed to train an AI model of a given capability decreases every year. -> Regulation of AI chips can only temporarily regulate the development of potentially dangerous AI models (but even temporary measures may buy time that is crucial for safety)
- Well-resourced actors may be able to support frontier AI development without relying on cutting-edge AI chips, by spending substantially more -> Strategies that rely on excluding these states from frontier AI development may be infeasible
- Compute supply chain could become more distributed in the future (e.g. due to new hardware paradigms) -> difficult to enforce restrictions

Compute governance for regulation

- **Governments have already taken major actions to regulate AI chips** : US has restricted the export of AI chips (and equipment needed to produce them) to China
- **Regulation of AI chips could also promote international cooperation** on AI safety, such as by enabling verification of international agreements on AI development
- Aside from regulating AI chips, it may be feasible and effective to **regulate other specialized hardware** used for frontier AI development, such as networking equipment in data centers -> very little research on this

Why compute ?

COMPUTE TRENDS ACROSS THREE ERAS OF MACHINE LEARNING

AI-specialized computer chips matter for AI policy because they are likely a crucial and governable ingredient of frontier AI development

- GPT-3 was trained with approximately $3e24$ computer operations.
- There are computer chips in everything but the chips used for cutting-edge AI development are highly specialized chips

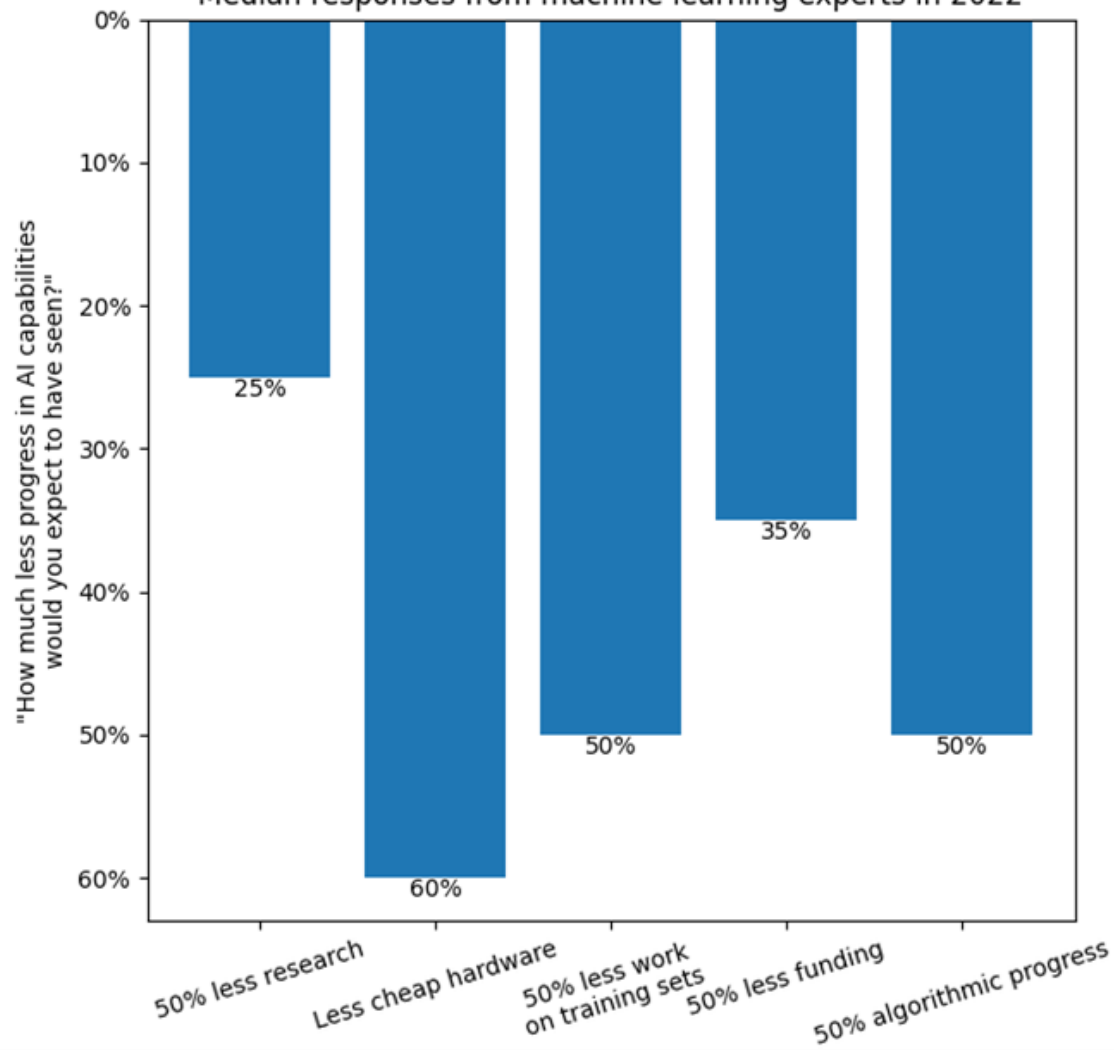
Why compute ?

AI Chips: What They Are and Why They Matter

Some chips ("AI chips") are vastly more efficient than others

*"Training a leading AI algorithm can require **a month of computing time and cost \$100 million**. This enormous computational power is delivered by computer chips that not only pack the maximum number of transistors—basic computational devices that can be switched between on (1) and off (0) states—but also are tailor-made to efficiently perform specific calculations required by AI systems. Such leading-edge, specialized "AI chips" are essential for cost-effectively implementing AI at scale; **trying to deliver the same AI application using older AI chips or general-purpose chips can cost tens to thousands of times more.**"*

Median responses from machine learning experts in 2022



Why compute ?

Further context on AI chips :

- In 2023, some examples of cutting-edge AI chips are A100s (Nvidia), TPUs (Google), and H100s (Nvidia)
- Most cutting-edge AI chips are specialized versions of GPUs
- Computer chips have been rapidly becoming more efficient for decades. Researchers at Epoch AI estimate that, in recent decades, **the number of GPU operations that a dollar can buy has doubled every ~2.5 years.**

Compute : training vs deployment

The Inference Cost Of Search Disruption – Large Language Model Cost Analysis

Training an AI model from scratch requires much more compute than running one copy of it after it has been trained.

- Providing an AI service to millions of users requires running many copies of an AI model
-> the **computational costs of this can easily surpass those of training the model**
- Fine-tuning typically requires significant compute: much less than the initial training (pre-training), but much more than running one copy of it.

Compute trends

Training Compute-Optimal Large Language Models

Scaling Laws Literature Review

Scaling laws for single-agent reinforcement learning

When an AI model is trained with more compute, it tends to have more advanced capabilities.

- These trends are so consistent that they have been used to make verified predictions, and they can be described mathematically by what are known as "**scaling laws**."
- Unclear how far "scaling laws" will generalize
- Epoch AI estimates that, since 2010, the **amount of compute used to train milestone AI systems has grown by ~4x each year** -> increasingly specialized chips, in growing numbers, for longer periods of time
- Growing compute use has come with rising costs, which have led frontier AI development to **shift from being dominated by academia to being dominated by**

Scaling hypothesis

THE SCALING HYPOTHESIS

On the Opportunities and Risks of Foundation Models

Very advanced AI capabilities could not require major breakthroughs in algorithms; instead, the main bottleneck would be scaling compute

- Many qualitatively new AI capabilities have been achieved in recent years primarily by scaling up existing AI designs, rather than with profound research insights

Can we regulate chips ?

Inherent properties of chips facilitate regulation

- Hard to enforce regulations on things like data or algorithms (challenging to monitor their proliferation, as they can be easily distributed)
- AI chips are physical objects, rather than information -> more feasible for governments to monitor and regulate their distribution.

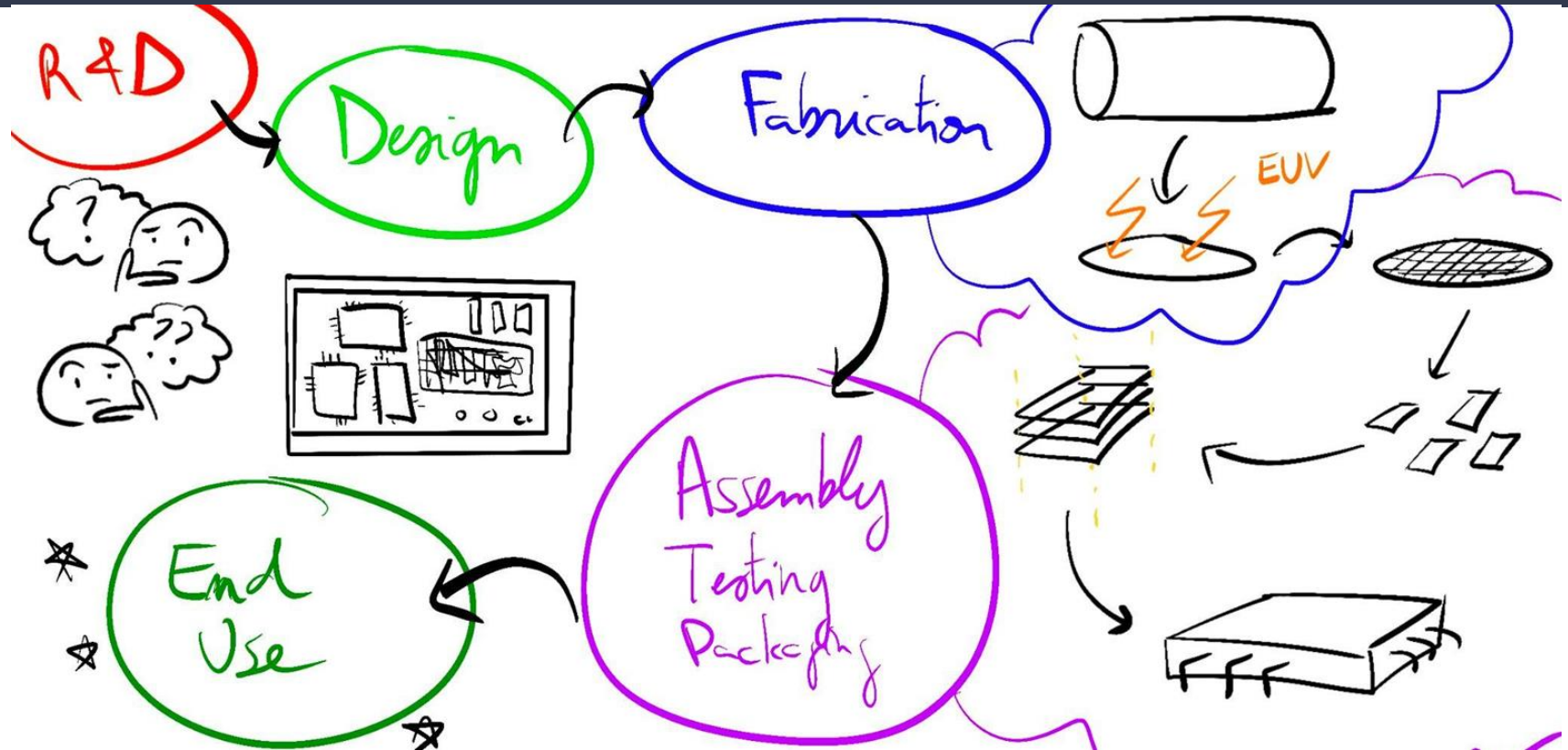
Can we regulate chips ?

Securing Semiconductor Supply Chains

The AI chip supply chain further facilitates regulation, because key steps in the global supply chain are dominated by a few countries

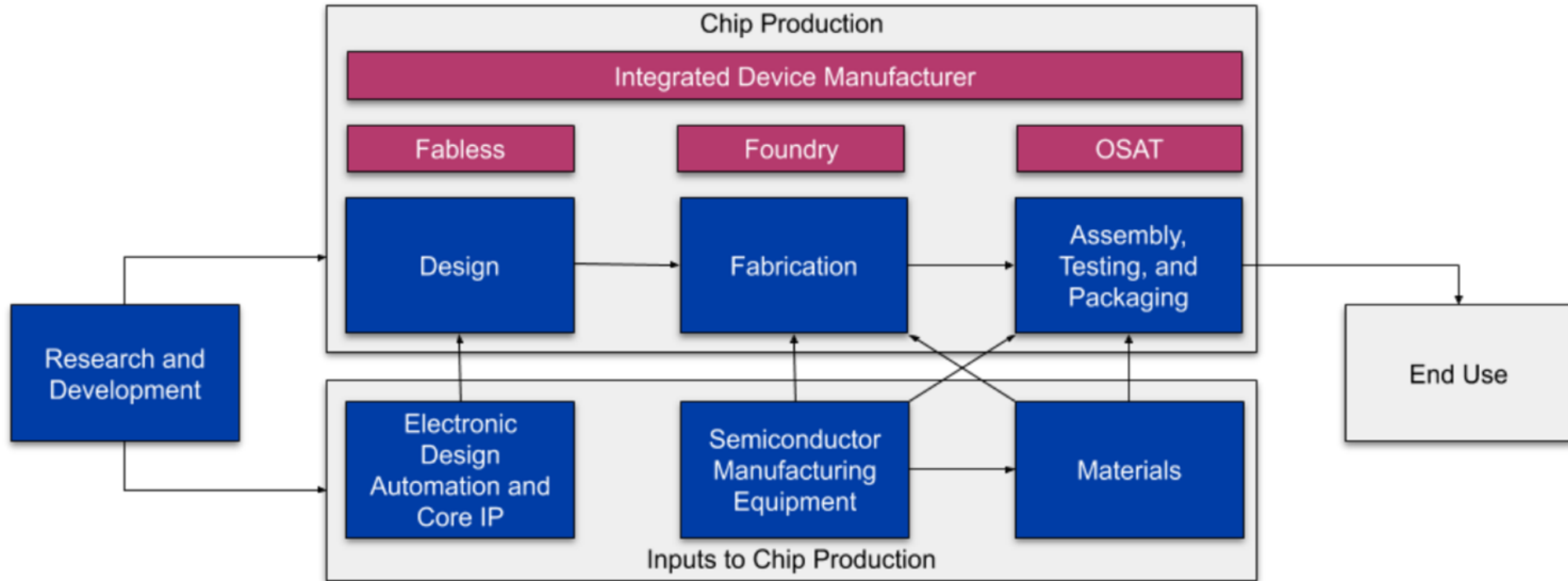
- No actor could easily stop relying on this global supply chain
- CSET report -> it would be **“incredibly difficult and costly for any country” to manufacture AI chips on its own**—even the United States or China.
- Some countries have outsized leverage over how AI is developed globally -> they can set conditions for technology exports.

Semiconductors supply chain (overview)



The Semiconductor Supply Chain

Assessing National Competitiveness

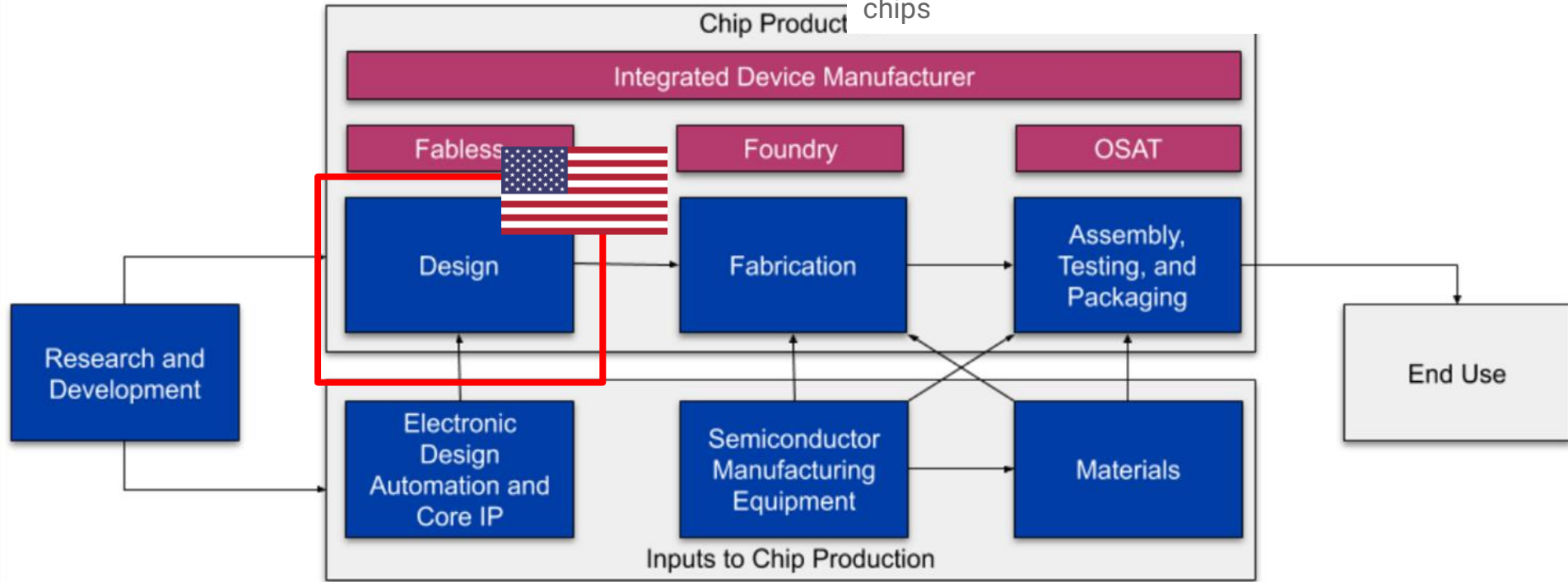


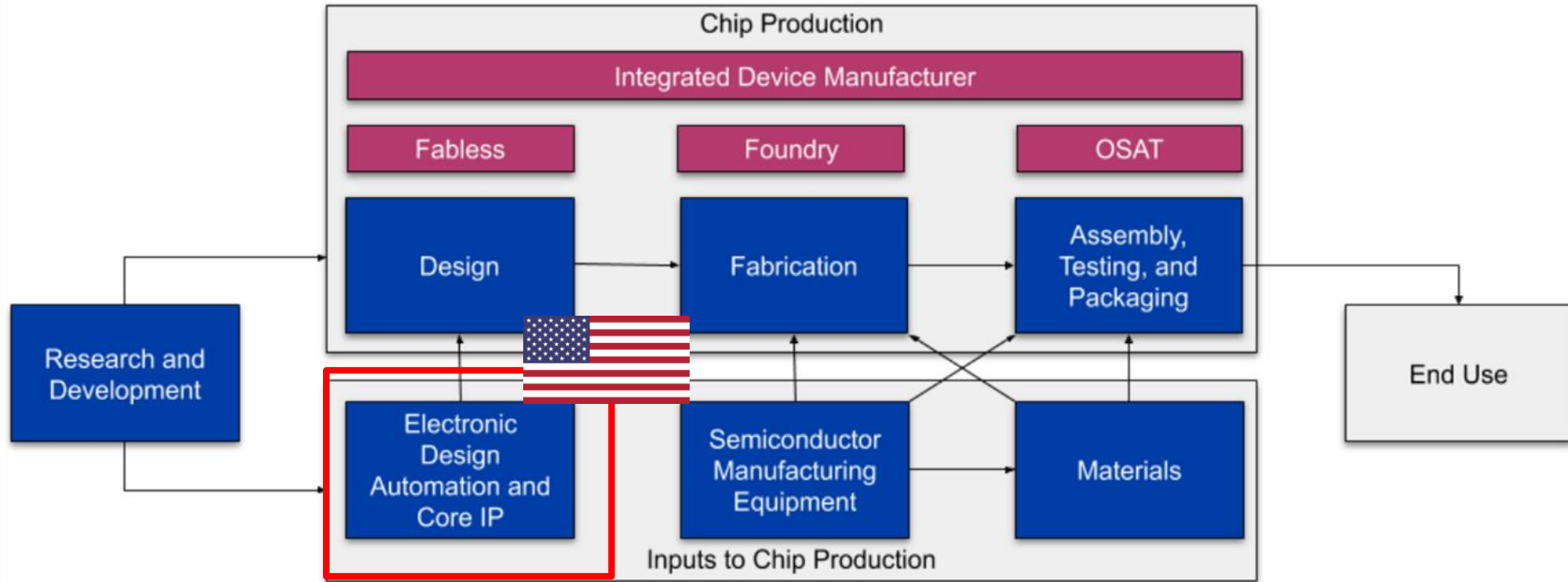
Note: Blue: Supply chain segment; Purple: Business model for production



Today's frontier models are exclusively trained on chips designed by NVIDIA and Google, both US companies.

Fabless : chipmakers design but don't produce the chips

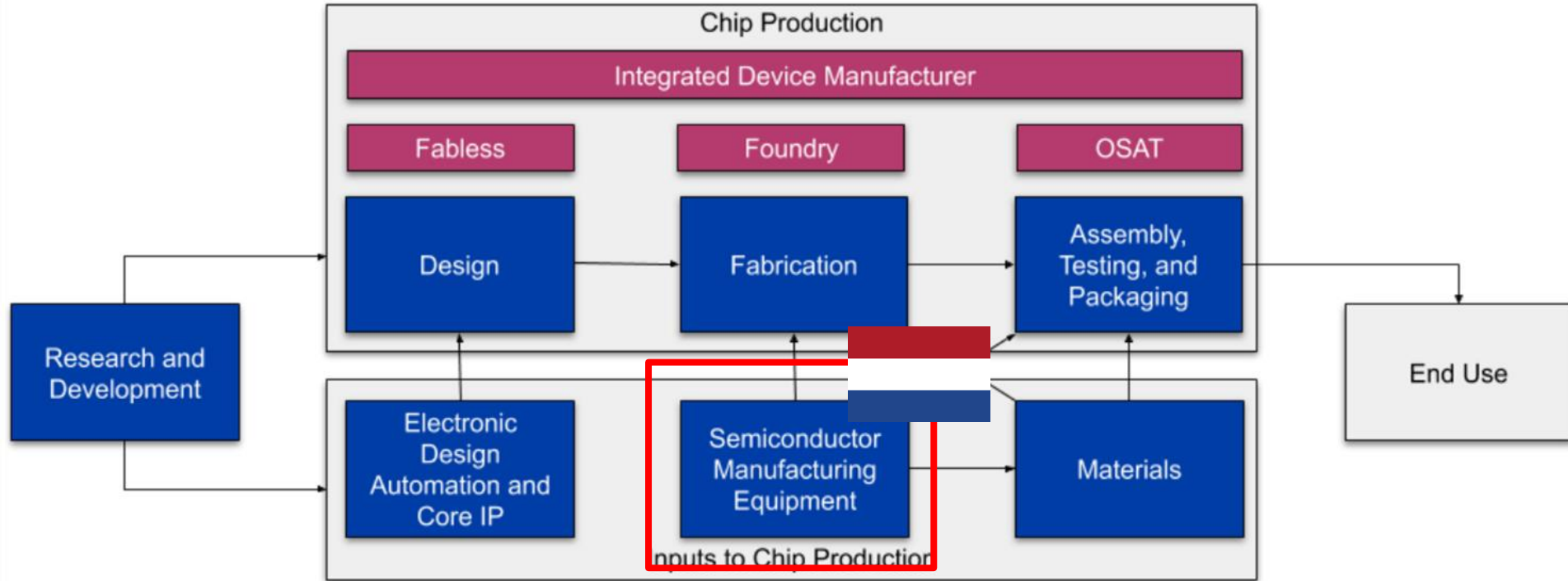


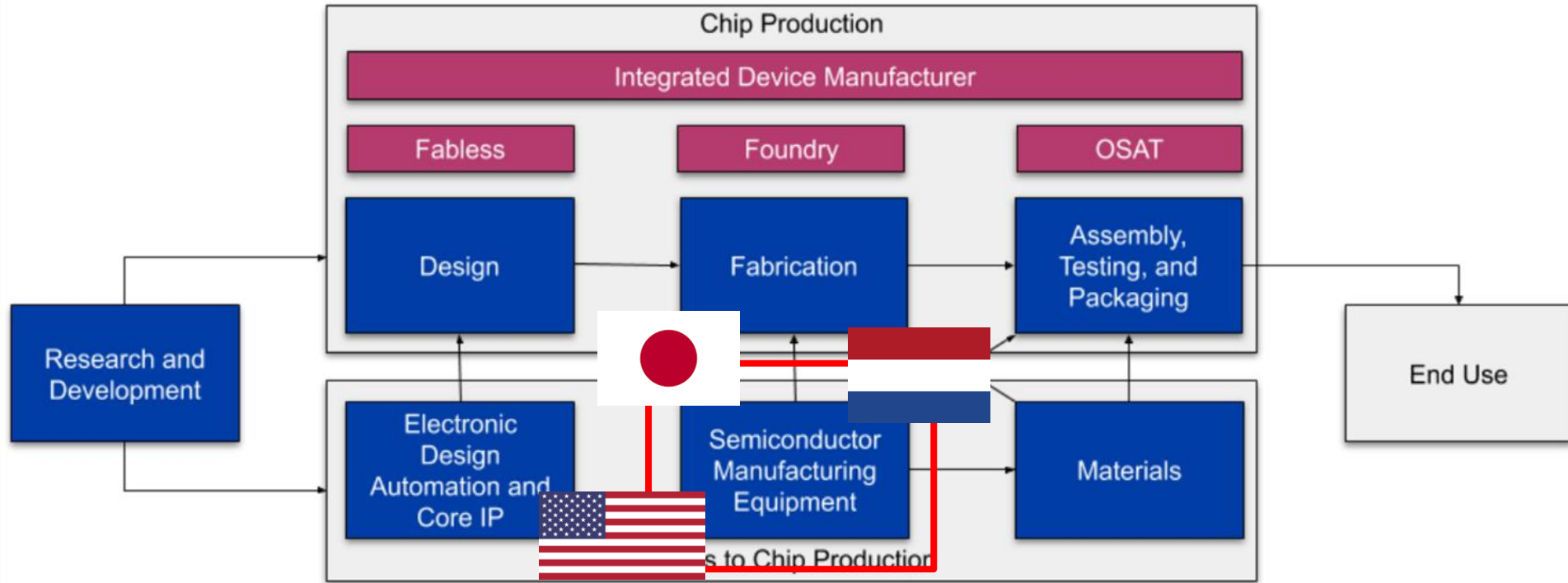


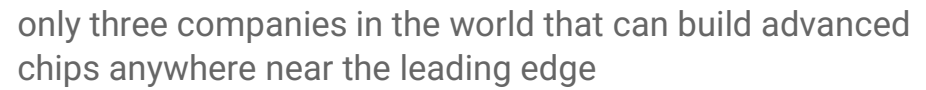
ASML

EUV lithography equipment : monopoly is due to the extreme technical and logistical complexity

Human hair width : 100 nm vs transistor width : 7nm (4nm for H100 ?)

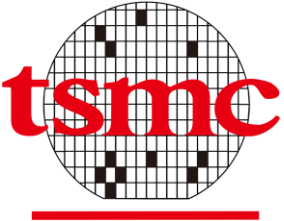






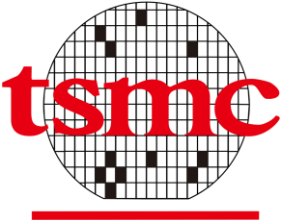
Foundries : manufacture chips designed by others





How has TSMC become such a dominant force? Why is it so difficult for any other company in the world to replicate what TSMC does?

- Manufacturing chips requires tremendous **upfront capital expenditure** :
 - In 2021, TSMC announced that it would spend \$100 billion over three years to expand its fabrication capabilities.
 - Recently spent \$20 billion to build a single fab: its legendary Fab 18, where the world's most advanced chips (including Nvidia's H100s) are built.
 - Advanced chip manufacturing requires multi-billion-dollar R&D investments year in and year out.
- **Virtuous cycle** : companies looking to get chips built choose TSMC because it offers the most advanced chipmaking capabilities -> this gives TSMC the volume required to justify and fund ongoing investments in order to maintain its lead -> and this world-leading investment budget further extends the company's advantages over rivals, making it the best choice for future customers.

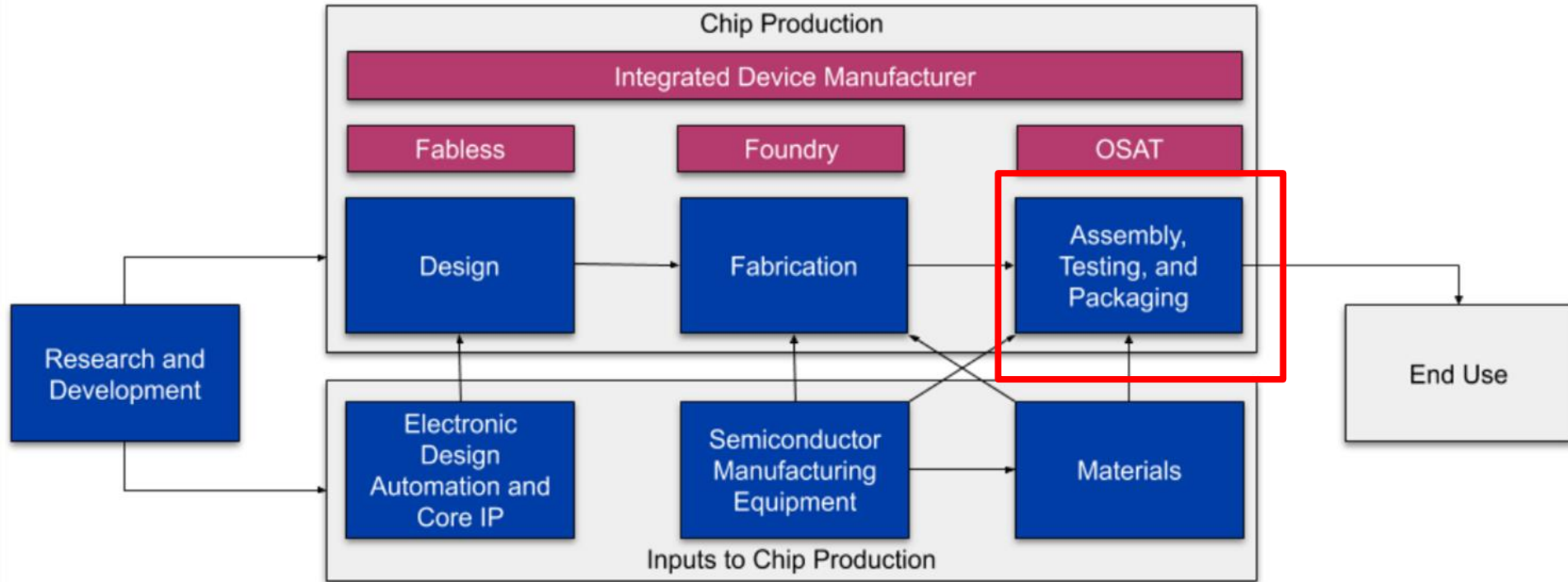


Case study : TSMC vs GlobalFoundries

- 2010s : GlobalFoundries sought to **challenge TSMC for chipmaking supremacy**
- Invested billions of dollars in order to develop leading-edge node technology and build the world's most advanced chips
- 2018 : GlobalFoundries leadership concluded that, given its scale, it would **never make financial sense** to make the multi-billion-dollar investments needed year after year to keep up with Moore's Law and **stay on the leading edge** of chip production.
- Gave up on developing leading-edge node technology, slashed its R&D costs and stopped competing with TSMC to build the most advanced chips (now focuses instead on producing lagging-edge semiconductors)

Not identified as a supply chain chokepoint

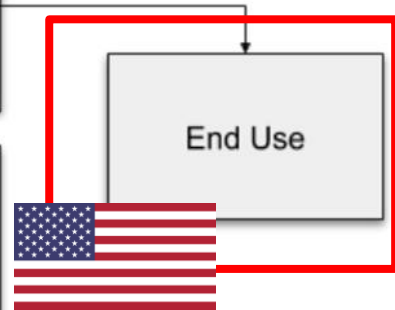
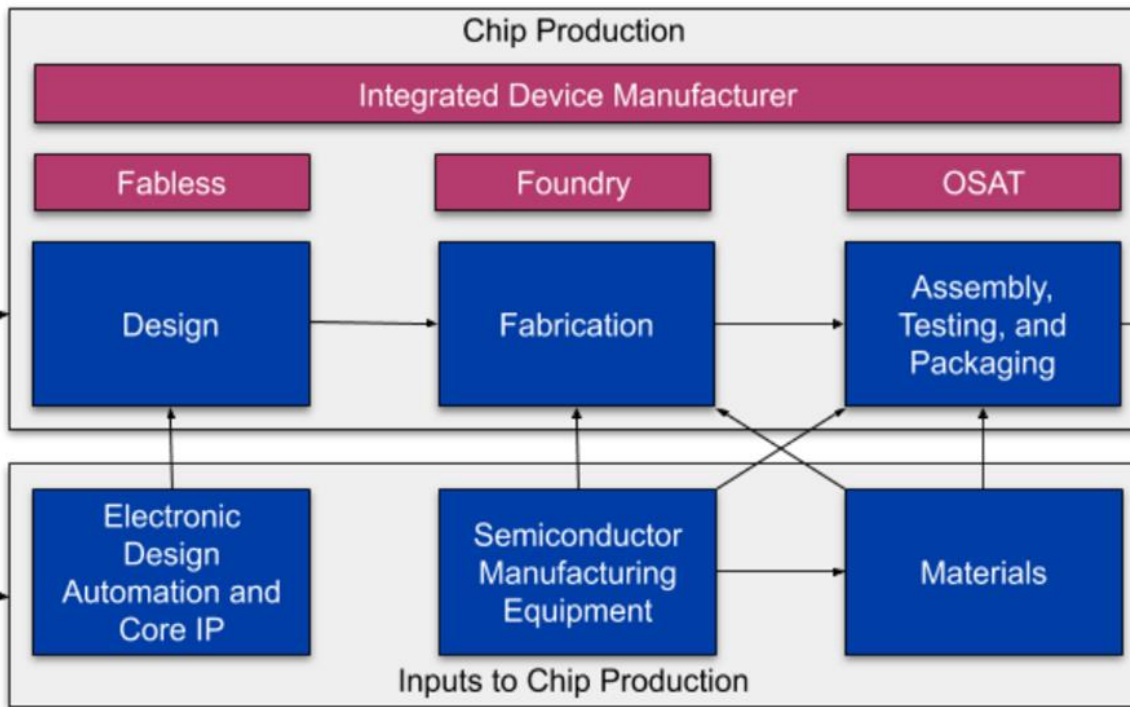
Easy to outsource





Selling access to AI chips over the Internet

3 US companies hold combined ~65% market share



Semiconductors supply chain - cloud compute providers

Supply chain : sophistication & expenses

Difficulties a country would face in manufacturing AI chips on its own:

- **Half-trillion-dollar** semiconductor supply chain is one of the world's most complex : production of a single computer chip often requires **more than 1,000 steps** passing through international borders 70 or more times before reaching an end customer
- State-of-the-art chip fabs now cost more than \$10 billion to build, making them the **most expensive factories ever built**
- China's national-level chip subsidies of \$18.8 billion by 2018 vs TSMC's \$34 billion investment on new fabs
- China total AI investment expected to reach \$38 billion by 2027 vs TSMC investing \$40 billion to build two state-of-the-art fabs in the US in 2022

Potential limitations

Small amounts of compute may pose severe security threats

- For a given AI capability, engineers tend to find ways to achieve it with less and less compute over time -> **algorithmic progress**.
- Epoch AI estimates that the **compute requirements** for AI models to achieve a given accuracy in recognizing images has **fallen by a factor of ~2.5 per year** since 2015.
- If the above trends continue (as we might expect, given ongoing innovation and incentives to reduce compute costs), then it will eventually be feasible to train AI models with dangerous capabilities, without relying on cutting-edge chips.
- As high-end hardware becomes cheaper, it tends to become more widely used, making it logistically and politically harder to regulate it.
- **It will likely become harder each year to regulate AI models of any given capability level.**

Potential limitations

Well-resourced states may carry on with advanced AI development without cutting-edge chips

- Well-resourced states **could buy large numbers of older-generation chips** and use them to train advanced AI models, despite not having access to more efficient, newer hardware.
- Strategies that rely on excluding well-resourced states from advanced AI development might be infeasible. (other compute governance strategies that do not rely on that may still work)

Potential limitations

Drastic changes in AI chip manufacturing may make export restrictions infeasible

- Some work aimed at **replacing the currently dominant approach** to building AI chips with very different approaches
 - neuromorphic computing : brain-inspired hardware
 - optical computing : hardware that uses light instead of electrons
- But : these approaches face uphill battles in competing with the immense amount of innovation that has gone into the current AI chip supply chain.
- Still : the success of an alternative hardware paradigm might mean that a small amount of hardware, or hardware that many countries can develop on their own, would suffice for dangerous AI development.

China export restrictions

From 2020 to 2023, the US and its allies have placed mounting restrictions on exporting AI chips (and the machines that make them) to China.

- 2020 : Dutch government, reportedly under US pressure, declined to renew ASML's license to export its most advanced chip-making machines to China.
- 2022 : US government unilaterally established sweeping restrictions on exporting AI chip technology to China -> asserted authority over some non-US-based companies, on the grounds that they were using US-made software and equipment or US employees.
- While US export controls on AI chips were initially unilateral, some actors with key roles in the AI chip supply chain (Netherlands and Japan, +Germany in discussion) have since imposed related export controls.

China export restrictions

From 2020 to 2023, the US and its allies have placed mounting restrictions on exporting AI chips (and the machines that make them) to China.

- 2022 : Biden administration banned the export of all high-end AI chips to any entity operating in China. (95% of all AI chips used in China today are Nvidia GPUs)
- Identified a number of other strategic chokepoints without which AI chip production cannot be sustained—and cut off China's access to these as well.
 - Electronic design automation
 - Semiconductor manufacturing equipment

China export restrictions

More broadly, compute governance can slow the proliferation of the ability to develop frontier AI or deploy it at mass scale, ideally limiting the access of malicious or reckless actors.

Compute gov for international cooperation

- AI chip policy so far has largely occurred in the adversarial context of export controls, that is not the only possibility for governing AI chips.
- AI chips might also be used to **verify compliance with international agreements** on AI, or to otherwise **advance cooperation**.
- For example, **mechanisms built onto AI hardware** may be able to verify and enforce compliance with AI regulations or agreements, in privacy-preserving ways.
- A third potential application of AI governance—in addition to restricting access and facilitating cooperation—**compute subsidies can promote beneficial AI projects**, such as safety research.

Hardware enabled mechanisms

[Home](#) / [News](#) / [New IAEA Uranium Enrichment Monitor to Verify Iran's Commitments under JCPOA](#)

New IAEA Uranium Enrichment Monitor to Verify Iran's Commitments under JCPOA

Modern example : monitor installed by IAEA checks that Iran's uranium isn't being enriched above a certain threshold (3.67%) -> limits nuclear risk while "allowing Iran to use nuclear technology for peaceful purposes"

-> provide the world with assurance that Iran fulfils its nuclear-related commitments

Hardware enabled mechanisms : assurances

- **Monitoring the supply of chips:** Verifying that actors have access to a certain number of chips by running proof-of-work challenges or requesting unique IDs of chips.
- **Monitoring workloads:** Ensuring that AI systems are only trained up to a certain size or within specific parameters.
- **Verifying compliance:** Ensuring that monitoring has not been tampered with by conducting on-site inspections or using technical tools to verify the location of chips (could be done by on-site inspections)
- **Enforcing compliance:** Remotely turning off chips or denying unsigned workloads, which could help prevent the misuse of AI systems.

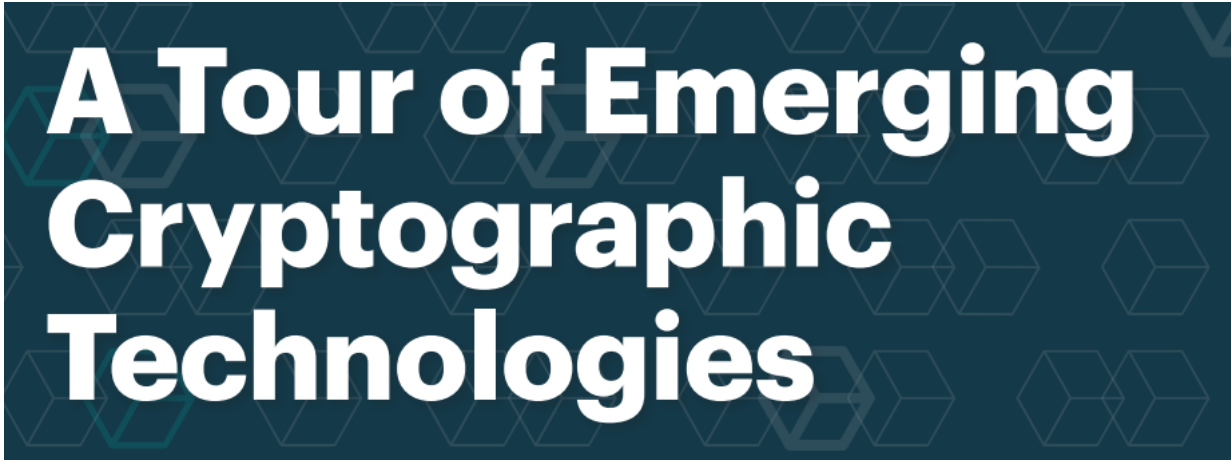
Compute monitoring

Reading :

WHAT DOES IT TAKE TO CATCH A CHINCHILLA?
VERIFYING RULES ON LARGE-SCALE NEURAL NETWORK
TRAINING VIA COMPUTE MONITORING

Privacy-preserving compliance verification

Bonus reading :



A Tour of Emerging Cryptographic Technologies

Chapter 2