Compute and the Governance of AI

Lennart Heim
lennart.heim@governance.ai
Outline

1. Risks from Advanced AI Systems
2. The Promise of Compute
3. Governance Capacities Enabled by Compute
4. Examples of Compute Governance
5. Conclusion: Compute and the Governance of AI
1. Risks from Advanced AI Systems
AI capabilities are advancing rapidly

State-of-the-art AI performance on benchmarks, relative to human performance

- Handwriting recognition
- Speech recognition
- Image recognition
- Reading comprehension
- Language understanding
- Common sense completion
- Grade school math
- Code generation

Human performance = 100%

Henshall, Time, 2023
Thinking about Risks from AI

Accident Risks

Misuse Risks

Structural Risks

Zwetsloot & Dafoe, 2019
Three Regulatory Challenges Posed by Frontier AI

- Deployment Safety Problem
- Unexpected Capabilities Problem
- Proliferation Problem
The AI Governance Problem

- AI has the potential to transform the economy, science, and security at a scale.
- Alongside the benefits, there are likely serious risks.
- Transformative AI systems might be developed in our lifetime, so they warrant more attention and caution.
AI Governance Definition

“The study and shaping of local and global governance systems — including norms, policies, laws, processes, politics, and institutions — that affect the research, development, deployment, and use of existing and future AI systems in ways that positively shape societal outcomes into the future.”

Maas, 2022
2. The Promise of Compute
Compute in the AI Production Function

AI Triad

Human capital

Data

Algorithms

Compute

AI Development
Feasibility: Compute is governable

**A. Feasibility: Compute is governable**

It is possible to monitor and shape who has access to computational resources and, to some extent, how they are used.

- Rivalry and Excludability
- Features of the Compute Supply Chain
- Quantifiability
The World’s Most Complex Product: Chips
Compute Production

Inputs to Chip Production

- Electronic Design Automation and Core IP
- Semiconductor Manufacturing Equipment
- Materials

Fabrication

Assembly, Testing, and Packaging

Fabless

Foundry

OSAT

Integrated Device Manufacturer

Integrated circuits ("chips")
Compute Production

Chip/Compute Production

Integrated Device Manufacturer

Fabless

Design

Fabrication

Foundry

Assembly, Testing, and Packaging

OSAT

Integrated circuits ("chips")

Inputs to Chip Production

Electronic Design Automation and Core IP

Semiconductor Manufacturing Equipment

Materials
Compute Production

Inputs to Chip Production

- Electronic Design Automation and Core IP
- Semiconductor Manufacturing Equipment
- Materials

Chip/Compute Production

- Integrated Device Manufacturer
- Fabless
- Foundry
- OSAT

Integrated circuits ("chips")
Compute Production

Inputs to Chip Production

Chip/Compute Production

Integrated Device Manufacturer

Fabless

Design

Electronic Design Automation and Core IP

Foundry

Fabrication

Semiconductor Manufacturing Equipment

OSAT

Assembly, Testing, and Packaging

Materials

Integrated circuits ("chips")
Compute Production

Chip/Compute Production

- Integrated Device Manufacturer
- Fabless
- Foundry
- OSAT

Design → Fabrication → Assembly, Testing, and Packaging

Electronic Design Automation and Core IP
Semiconductor Manufacturing Equipment
Materials

Inputs to Chip Production

Integrated circuits ("chips")
Compute Production

Chip/Compute Production

Integrated Device Manufacturer

Fabless

Design

Fabrication

Foundry

Assembly, Testing, and Packaging

OSAT

Electronic Design Automation and Core IP

Semiconductor Manufacturing Equipment

Materials

Inputs to Chip Production

Integrated circuits ("chips")
Compute Provision and Usage

Integrated circuits ("chips")

Data Center Construction

Compute Provision / Data Center Operation

Cloud Provider

Big Tech Company

Compute Provision

Land, Cooling, Electronics, Networking

Power, Water, Connectivity

Inputs to Compute Provision

AI Training

Algorithms, Data

Inputs to AI Training

Foundation Model
Compute Provision and Usage

- **Integrated circuits ("chips")**
- **Data Center Construction**
  - Land, Cooling, Electronics, Networking
  - Inputs to Compute Provision

- **Compute Provision / Data Center Operation**
  - Power, Water, Connectivity

- **Compute Provision**
  - Big Tech Company
  - Cloud Provider

- **Compute Usage**
  - AI training
    - Algorithms, Data
    - Inputs to AI Training

- **AI Lab**
- **Foundation Model**
Compute Provision and Usage

Integrated circuits ("chips")

Data Center Construction

Compute Provision

Land, Cooling, Electronics, Networking

Power, Water, Connectivity

Compute Provision / Data Center Operation

Cloud Provider

Big Tech Company

Compute Usage

AI lab

AI Training

Algorithms, Data

Inputs to Compute Provision

Inputs to AI Training

Cloud Provider

Foundation Model
Compute Provision and Usage

Integrated circuits ("chips")

Data Center Construction

Compute Provision

Cloud Provider

Big Tech Company

Compute Provision / Data Center Operation

Land, Cooling, Electronics, Networking

Power, Water, Connectivity

Inputs to Compute Provision

AI Training

Algorithms, Data

Inputs to AI Training

Foundation Model
Feasibility: Compute is governable

A. Feasibility: *Compute is governable*

It is possible to monitor and shape who has access to computational resources and, to some extent, how they are used.

- Rivalry and Excludability
- Features of the Compute Supply Chain
- Quantifiability
Training compute doubles every ≈6 months.
Efficacy: Compute is indicative of AI capabilities

A. Feasibility: *Compute is governable*

- Rivalry and Excludability
- Features of the Compute Supply Chain
- Quantifiability

B. Efficacy: *Compute is indicative of AI capabilities*

By observing, regulating, or influencing an entity’s access to compute, one can predict and modulate actors’ access to AI capabilities.
Why Governing Compute is Promising for Governing AI

A. Feasibility: *Compute is governable*

- Rivalry and Excludability
- Features of the Compute Supply Chain
- Quantifiability

By governing compute, you can govern AI capabilities.

B. Efficacy: *Compute is indicative of AI capabilities*
3. Governance Capacities Enabled by Compute
1. Knowledge
2. Shaping
3. Enforcement
1. Knowledge
2. Shaping
3. Enforcement

How actors use, develop, and deploy AI—and which actors are relevant.
1. Knowledge
2. Shaping
3. Enforcement

Direct and influence the trajectory of AI development and the distribution of AI capabilities among different actors.
1. Knowledge
2. Shaping
3. Enforcement

Respond to potential violations, such as an actor training an excessively risky AI system.
4. Examples of Compute Governance
US Semiconductor Export Restrictions

1. Block access to high-end AI chips
2. Block designing AI chips domestically
3. Block from manufacturing advanced chips
4. Block from domestically producing semiconductor manufacturing equipment
5. Block “US persons” from supporting chip development
Leverage Compute for Verification Mechanisms

- Assurances & verifiable commitments (across nations and actors)
- Transparency, e.g., transparent use of compute
- Shared control, e.g., on a joint AI project
- Sanctions and restricted access
Examples of Verification Mechanisms

- Proof-of-learning / training
- Proof-of-inference / deployment
- Proof-of-data
- (Verification of) properties of training runs
- Or proof-of-non-learning?
5. Compute and the Governance of AI
AI Governance Definition

“The study and shaping of local and global governance systems — including norms, policies, laws, processes, politics, and institutions — that affect the research, development, deployment, and use of existing and future AI systems in ways that positively shape societal outcomes into the future.”

Maas, 2022
Governance throughout the AI Lifecycle

Development

- Design
- Model Architecture
- Training
- Trained Model
- Deployment
- Fine-tuning, distillation, ...
- Inference
- Deployed Model

Deployment
Governance throughout the AI Lifecycle

**Development**

- Pre-emptive / pre-training authorization

**Design**

- Model Architecture

**Training**

- Trained Model

**Training Compute Verification**

- Proof that model only used X FLOP

**AI Chip Export Restrictions**

- Stop/hinder actors from training such systems

**Training Compute Threshold**

- Which models are of concern?

**Actor’s Compute Capacity**

- Who’s able to train this model?
Governance throughout the AI Lifecycle

**Licensing of AI Systems**
Cloud providers verify the (electronic) license of an AI system before deploying.

**Pre-Deployment Notification**
Governments and other entities receive a notification.

**Verify Model Authenticity during Evaluations** (which is required for deployment)
Governance throughout the AI Lifecycle

**Compute Providers Shut Down AI System**
to prevent continuous harm

**Compute Providers Identify Developer**
(enable post-incident attribution)

**Deployment**

**Compute Providers Identify Deployer**
(enable post-incident attribution)

**Trained Model**

**Deployed Model**

**Retrospective Deployment Correction**
Conclusions

- Governing compute is **feasible, effective and valuable** but **alone not sufficient**
- Enabling AI governance capacities that would otherwise be difficult to achieve: **knowledge, shaping, enforcement**
- Mechanisms for verifiable claims that can enable more trust across actors
- Compute is already being used as a governance node — we should improve our understanding and build **more nuanced instruments**

Lennart Heim
{lennart.heim@governance.ai}
@ohlennart
https://heim.xyz