THE ESSAY, IN FIFTEEN SLIDES

AI Governance From First Principles

One definition—and the questions it teaches us to ask.

AIGOVSPEC

WHEN SOMEONE ASKS WHAT AI GOVERNANCE IS, WE ANSWER WITH A LIST

laws · policies · standards · audits · risk frameworks · oversight bodies

---

Those things matter. But a list of instruments does not explain the problem they are meant to solve.

It also makes the field sound distant from anyone who does not write laws, build models, or attend policy meetings.

Practice came long before agreement on a definition

2017–18	2019–21	2022–23	2024	2025–26
Naming a field	**Principles, oversight, rights**	**Inside the organisation**	**The word was everywhere**	**Dialogue and agents**
Layers of standards, ethics and law. A field of study.	OECD, IEEE, UNESCO, the African Commission.	Rules, processes and tools; ISO standards; NIST.	The EU AI Act, the Council of Europe, the UN.	Kenya, India, a UN Global Dialogue, agentic AI.

38 documents read, 2017 to 2026

13 offer a definition

19 describe an approach

6 build on the term

READING ACROSS THE RECORD

Four patterns stand out

01 Practice came first

Institutions, laws and dialogues were built while the meaning was still being worked out.

02 Definitions name the toolkit

Most pair instruments (rules, processes, practices) with a purpose: what governance is made of.

03 Adaptation is shared

Many call for agile, living governance. When to act is raised far less often.

04 Development and choice

The African Union, Latin America and the UN frame governance around inclusion and policy choice.

PART TWO · WHY A NEW DEFINITION

Everything in our definition owes something to this record. What we add is a way of holding it together.

WHAT WE TAKE

A field of inquiry · *Dafoe*

Practice someone has to run · *Mäntymäki, Kenya*

Benefits broadly shared · *UN, African Union, Montevideo*

Governance that keeps learning · *NIST, OECD, Singapore*

WHAT NO SINGLE SENTENCE HELD

1 · A name for what governance acts on

2 · Timing, written into the definition itself

OUR DEFINITION, FROM FIRST PRINCIPLES

AI governance is the field of inquiry and practice concerned with how societies should shape the incentives, institutions, infrastructures and safeguards around AI to share its benefits broadly, prevent and address harm, and maintain the capacity to change course as today's choices become tomorrow's conditions.

Breaking down the definition into smaller parts

WHO DECIDES

1 Inquiry and practice
2 Societies
3 Should shape

WHERE IT ACTS

4 Incentives
5 Institutions
6 Infrastructures
7 Safeguards
↻ *Power, across all four*

WHAT IT DELIVERS

8 Share its benefits broadly
9 Prevent and address harm

OVER TIME

10 As today's choices become tomorrow's conditions
11 Maintain the capacity to change course

One thread runs through every part: a health ministry deciding whether to buy an AI tool that helps triage patients.

PARTS 1–3 · WHO DECIDES

Governance is something we study, and something we do.

1

Inquiry and practice

learning what works, and making it count

What do we need to learn—and how will that knowledge change what someone does?

2

Societies

who gets to decide?

Who has authority, who is affected, and how can those affected influence or challenge the decision?

3

Should shape

which possibilities are we making easier?

Which possibilities are we making easier to pursue—and for whom?

Question to carry, under each part

PARTS 4–7 · WHERE GOVERNANCE ACTS

Shaping happens in particular places. Here are the four objects.

4

Incentives

what does the environment reward?

What makes that promise hold under pressure?

5

Institutions

who can act?

Who is responsible? And what makes them capable of exercising that responsibility?

6

Infrastructures

what does this depend on, and who controls it?

What are we becoming dependent on—and what alternatives will remain?

7

Safeguards

what makes protection real?

What prevents, detects, limits, or remedies failure—and what happens when it fails?

↻ POWER · WHO SETS THE TERMS?

Underneath all four lies one thing: power. Power runs through all of them.

Whoever controls incentives, institutions, infrastructure or safeguards sets terms for everyone else. And governance does not only constrain power. It creates it: every auditor, regulator and review board is a new holder of authority.

QUESTION TO CARRY

What power does this mechanism create—and who will hold it accountable?

PARTS 8–9 · WHAT GOVERNANCE SHOULD DELIVER

Our definition names both sides of the ledger: the gains, and the harm.

8

Share its benefits broadly

who gains?

Who gains, who bears the risks and costs, and what would count as a fair outcome?

9

Prevent and address harm

two verbs, two jobs

What harm could this cause, who will notice it, and what happens for the people already affected?

REAL EXAMPLE

17.7% → 46.5%

A US hospital algorithm used past spending to predict need. Correcting that bias would raise the share of Black patients flagged for extra care from 17.7% to 46.5%.

Obermeyer et al., *Science*, 2019

PARTS 10–11 · OVER TIME

Prevention and remedy both depend on when we act. And acting early means acting with incomplete knowledge.

10

As today's choices become tomorrow's conditions

why timing matters

What are we committing ourselves to, and how difficult will it become to change course?

11

Maintain the capacity to change course

governance must be able to learn

How will we discover that this arrangement is failing—and how will we change it?

THE TIMING PROBLEM

"When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time-consuming."

David Collingridge, 1980

PUT THE PARTS BACK TOGETHER

A guide to carry into the next conversation

LOOK AT	ASK
Inquiry and practice	What must we learn, and how will it change action?
Authority and participation	Who decides, who is affected, and who can challenge?
Incentives	What behavior is rewarded under pressure?
Institutions	Who can act, and do they have the capacity?
Infrastructures	What dependencies are created, and what alternatives remain?
Safeguards	What protects people, and what follows when it fails?
Power and distribution	Who gains, who bears costs, and who holds new authority?
Timing	What becomes harder to reverse if we wait?
Learning and revision	What evidence triggers review, and who can change course?

Begin with the real-world outcome a proposal is meant to shape. Then examine the arrangements behind it.

RETURN TO THE MINISTRY

What this changes in practice

BEFORE

“Is the model accurate enough?”

One question, about the tool.

AFTER

Perhaps we should begin with...

- A limited pilot, keeping existing clinical capacity
- Evidence of performance across patient groups
- Access to records, and rights to export data
- Independent scrutiny
- Conditions for expansion or suspension
- Ability to change tool or provider.

Illustrative choices, not a universal prescription.

WHEN SOMEONE ASKS WHAT AI GOVERNANCE IS, BEGIN HERE

How do societies shape the conditions around AI—and who has the power to influence those choices?

Govern while choices are still open.
And build governance that can still change its mind.

Read the essay, the history and the parts at **aigovspec.com** · A CENTPOL product · Built by Samuel Abinsinguza

AIGOVSPEC · AIGOVSPEC.COM/SLIDES

AI Governance From First Principles

Dataset 0.7.0, dated 2026-10-27 · PDF made 2026-10-27.

Research preview: independent expert review is pending.

Licensed CC BY 4.0. Share and adapt with credit to AIGovSpec, and note your changes.

Essay and slides by Samuel Abinsinguza · A CentPol product