

Civilization Leap

Volume Two

Pre-Institutional Language

English Translation

Derived from R10.10

Zijun Fu

Civilization Leap Research Group

2026

About This English Edition

This file contains the complete English translation of Volume Two of Civilization Leap, derived from the Chinese source v1.21-R10.10. The shared front matter of the five-volume work is reproduced in full, followed by this volume and its relevant material notes. The original numbering and cross-references are retained. The companion volume, Shared Material Notes and Further Reading, preserves all sixty-two material notes and the upgraded public further-reading pages. This separately readable volume is not a combined edition of all five volumes.

Contents of This Volume

Shared Front Matter

Chapter 7 · Seeing What Is Real: How Can We Know We Have Not Gone Astray?

Chapter 8 · Bringing Consequences into the Reckoning: Who Answers for a System's Decisions?

Chapter 9 · Who Is Entitled to Decide: When Problems Cross National Borders

Chapter 10 · The Anchors of Trust: When “Seeing” Is No Longer “Believing”

Chapter 11 · Justice in Distribution: When Survival No Longer Has to Be Proved through Performance

Chapter 12 · Education's Change of Direction: When AI Begins Answering for Us

Material Notes for This Volume

Shared Front Matter

Civilization Leap

The Path of Civilization in the AGI Era

Five Volumes

By Zijun Fu

Civilization Leap need not eliminate self-interest. It need only change the default path by which self-interest realizes itself.

Let the fulfillment of others become a way of fulfilling oneself.

Let effort find an echo and contribution never sink unseen; let basic rights need no proof, authorization stay within its bounds, and responsibility have someone to bear it.

Freedom from Survival Anxiety → Happiness in Life → Greater Meaning in Life

Produced by the Civilization Leap Research Group

Copyright Page

Title: Civilization Leap: The Path of Civilization in the AGI Era

Author: Zijun Fu

Produced by: Civilization Leap Research Group

Year of publication: 2026

DOI: 10.5281/zenodo.18814225

License: CC BY 4.0

Official website: civilleap.com

Current gateway: www.civitas.top

Contact email: zijunfu@civitas.top

Copyright notice: Copyright © 2026 Zijun Fu. All rights reserved.

Suggested citation: Zijun Fu, Civilization Leap: The Path of Civilization in the AGI Era, 2026.

Version: v1.21-R10.10 | Targeted Revision for the Whole Book's Intellectual and Living Conclusion | T-10
EMB-2 Cognitive Starting Point and Cross-Reference Synchronization Edition

Reading note: This is a book of thought. Its main text allows scenes, metaphors, strong judgments, and rhetorical compression to bring a problem home first. Facts, quotations, historical material, theoretical judgments, and future thought experiments must nevertheless retain their distinct identities. Composite historical scenes do not impersonate actual cases; future extrapolations do not impersonate present facts; literary metaphors do not impersonate empirical laws. Necessary qualifications are placed later wherever possible, to support the thought rather than flatten it before it has had a chance to arise.

Note on stages: Where the book discusses stages such as initiation, transition, stability, and self-maintenance, and phases such as normal operations, construction, resource constraint, crisis, and recovery, these should be understood separately. Stages and phases do not automatically change basic rights, sovereignty, or the sources of authorization. Specific parameters can be designed only after actual actors, jurisdictions, and real-world conditions have been established.

R10.10 whole-book completion note: All areas frozen in R10.9 and the preceding revisions R10.2–R10.8 remain in force. This edition implements only the six targeted P1 revisions decided in the final whole-book review: the historical-to-contemporary bridge in Chapter 1; the naming of Section 2.10; Sections 3.1, 3.10, and 3.11; Sections 5.3, 5.6, and 5.10; Sections 7.5–7.9; and the opening of Chapter 15 and the closing structure of Section 15.11. The aim is not further literary embellishment, but removal of defenses, old theoretical tones, technical catalogs, and governance memoranda still standing between readers and thought. No frozen boundary concerning subjects, rights, AI, continued existence, the three goals, stages and phases, memory, contribution, or institutional force is rolled back.

Statement on AI Assistance

Generative artificial intelligence was used as an auxiliary tool in organizing research materials and leads to cases, comparing structures and versions, developing counterarguments, revising language, flagging matters for factual verification, and producing documents. AI also assisted with some drafting and consistency checks, with continuing review, selection, and revision by the author.

AI does not serve as an author, factual source, evidence, professional opinion, or authoritative endorsement. The author bears responsibility for the core theories, value judgments, institutional proposals, selection of citations, and final expression. Facts, data, quotations, and cases are governed by verifiable original sources

and professional research. Material not sufficiently verified should be understood as hypothesis, extrapolation, thought experiment, or material awaiting validation. AI assistance does not replace peer review, professional scrutiny, or authorization and responsibility within actual jurisdictions.

Contents of the Complete Work

Overview of the Five Volumes' Timescales

The centuries, millennia, and ten millennia invoked here distinguish scales of observation; they are not a timetable of future events. Volume One addresses present ruptures. Volume Four extends the timescale across generations and into the long term. Volume Five carries the questions further into the conditions of life and meaning.

Volume One · The Civilization of Being Needed—A Crisis of Meaning on a Century Scale

Bridge between volumes: If Volume One's problem holds—if the old civilization truly can no longer generate a place, a path, and a future for enough people—the next step is not to keep finding more cases, but to find a new language for civilization. Without a new language, civilization cannot organize a new reality.

When a reality can no longer be fully described in the old language, continuing to amend the old institutions is like trying to find a newly emerged continent with an old map.

What truly makes the leap is not merely institutions, but civilization itself.

Volume Two · Pre-Institutional Language—The Foundations of Governance on a Century Scale

What truly makes the leap is not merely language, but civilization itself.

Volume Three · Civilization Leap White Paper—A Blueprint for Civilization on a Century Scale

Bridge between volumes: Any institution that can operate only for a few decades remains a proposal for an era, not a proposal for civilization.

All the institutions discussed in this volume—contribution values, resource pools, the meaning economy, and governance mechanisms—are institutional expressions of Civilization Leap, not Civilization Leap itself.

Institutions are tools; civilization is the protagonist.

What truly makes the leap is not merely the blueprint, but civilization itself.

Volume Four · Civilizational Perpetuity—Civilizational Safety on a Millennial Scale

Bridge between volumes: Civilization needs to endure not because more time means greater glory, but because meaning can continue to arise only while the future remains open.

What truly makes the leap is not merely continuity, but civilization itself.

Volume Five · The Emergence of Meaning—The Weight of Life on a Ten-Millennial Scale

General Preface

Statement of level: This book addresses the institutional level of civilization. It is not cosmology, the dynamics of meaning, or a practical operating manual.

For meaning's ontological foundations and higher-level theory, see Emergent Meaning Cosmology and The Dynamics of Meaning. This book addresses the leap at the level of civilization.

Cognitive Starting Point

The cognitive starting point of this framework: The emergence of AGI has not overturned earlier philosophy. It has, for the first time, brought into a complete separation within actual technology concepts that earlier philosophy long assumed together without adequately distinguishing them: intelligence, consciousness, subjecthood, life, finitude, and meaning. Meaning is the root; its sources may differ. Finitude is not meaning's only source, but the structural foundation through which meaning enters reality and acquires irreplaceable weight. For human beings, space and time constitute life; for AGI, space and time are first parameters. It processes space and time, but has not been shown to experience them as a finite life. Constraints enable AGI to bear operational responsibility, but what constraints produce is bounded operation, not finite existence. Until there is evidence that AGI possesses subjective experience, nontransferable continuity of the subject, and awareness of consequences as its own, this framework positions AGI as an understander, protector, and transmitter of meaning, and an extender of capability. It does not recognize AGI as a subject of meaning emergence. Humans and AGI are not competitors: Humans generate and bear meaning through finite lives; AGI lets human creation and responsibility reach farther.

Theoretical source: T-10, Space and Time Endure, Finitude Is Irreplaceable—A Statement of the Cognitive Starting Point, V3.0. The full, concise, and one-sentence public versions present the same proposition at different levels; they are not competing versions.

First overarching principle: Civilization Leap does not eliminate self-interest; it changes the default path by which self-interest realizes itself. When a civilization begins to see whom its power may harm, to recognize which mistaken paths require a return, and to be willing to leave a future for those not yet born, it is no longer merely pushed along by history. It begins to learn to observe, choose, and correct itself. "Agency" here describes this awareness at the scale of civilization; it does not personify civilization as an institutional subject independently able to hold rights, bear obligations, or acquire governing sovereignty.

Written as Civilization Becomes a Self-Aware Existence

For most of human history, civilization never realized it was becoming civilization.

Tribes organized only to survive;

states were founded only to secure order;

industry expanded only for efficiency;

technology developed only to solve problems.

History seemed driven by countless local choices, without an overall direction.

Until the twenty-first century.

For the first time, humanity created a form of intelligence that might surpass its own cognitive capabilities.

At the same time, humanity acquired, for the first time, technological power capable of affecting the fate of human civilization as a whole.

These two developments together brought a new question into being:

Can civilization understand itself and actively decide its future?

Civilization Leap was born at this historical juncture.

It is neither a book about technology nor a political program.

It is an attempt—

an attempt to strengthen human civilization’s capacities for self-reflection, self-restraint, and long-term continuity. This capacity of the whole does not replace the subjecthood, rights, or final decisions of actual people.

I. Civilization’s Three Awakenings

Looking back, human civilization has undergone three fundamental awakenings.

The first awakening: Consciousness

Language and symbols enabled human beings to think about themselves.

Humanity was no longer merely part of nature, but a being that understood nature.

The second awakening: Institutions

The emergence of law, science, and the modern state enabled society to function beyond individual lifespans.

Civilization began to acquire stable structures.

The third awakening: Civilization

The arrival of the AGI age brings civilization, for the first time, face to face with its continued existence as a whole.

Humanity must answer:

Why do we exist?

Where are we going?

Can we endure over the long term?

Civilization Leap is written amid this third awakening.

II. The Structure and Mission of the Five Volumes

This series is not a single treatise, but a framework of thought about civilization unfolding layer by layer.

Volume One: The Civilization of Being Needed

Posing the problem:

Industrial civilization has long bound labor, income, social position, and “being needed” closely together. AI and automation are changing the scarcity and organization of some of that functional labor.

If people should no longer have to prove their entitlement to survive and their place in society repeatedly through functional labor, how can civilization understand human value anew and rebuild enduring places for participation, relationships, creation, care, and responsibility beyond labor?

This volume reveals:

Civilization must undergo a structural leap.

Volume Two: Pre-Institutional Language

Addressing the problem:

Before any institutional transformation, relationships easily conflated must first be made clear.

This volume separates subject and role, rights and contribution, facts and evaluation, capability and power, authorization and execution, and professional judgment and public decision, among other relationships. This allows different disciplines, institutions, and civilizations to discuss the same real-world problem without first accepting the same theory.

Volume Three: Civilization Leap White Paper for the AI Age

Entering practice:

Turning the relationships distinguished in the first two volumes into institutional prototypes and governance boundaries that can be verified, can fail, allow exit, and can be corrected.

The emphasis is not on declaring that institutions of a new civilization already exist. It is on explaining how any real-world proposal acquires force, how rights and authorization constrain it, and how it can be stopped, rolled back, and corrected when it goes wrong.

Volume Four: Civilizational Perpetuity

Taking the next step:

From “how to make the leap” to “how to endure.”

Introducing:

The Common Window: anti-slippage safeguards, rollback, and circuit breaking;

Protection of the conditions in which meaning arises: preventing safety, efficiency, and long-term stability from swallowing the subject;

Two-sided civilizational immunity: AGI-COS addresses machine-side safety; Human-COS addresses humans’ high-consequence judgments;

Self-correction in a long-lived civilization, an open future, and a cosmic-scale perspective.

This volume extends the timescale and asks how successful institutions can avoid becoming permanent sovereignties. The interstellar remains a distant open direction, not a rank of civilizational maturity.

Volume Five: The Emergence of Meaning

Continuing the inquiry:

When civilization can endure longer, how can different lives retain real space in which to form meanings of their own?

This volume uses longevity, digital life, virtual environments, and future subjecthood as stress tests. It answers the question, “After continued existence is secured, how can life still form its own reasons to continue?”

Distant imaginings such as reincarnation and multiple worlds remain optional thought experiments, not the main line or a mandatory route.

III. What Civilization Leap Truly Means

Civilization Leap is not a passage from an old world into a perfect new one.

It is not a utopian plan.

Nor does Civilization Leap aim to attack anyone. What it seeks to loosen is not the power of one class of people, but a default path on which self-interest can be realized only by surpassing others, controlling them, or passing costs onto them. Ordinary people can be crushed by survival pressures; those who command resources can also be trapped by control, competition, anxiety, and distrust. Wealth and power can bring resources without necessarily bringing happiness.

What Civilization Leap truly points toward is greater freedom, responsibility, meaning, and peace for every person beyond the old logic.

The leap truly means:

Civilization no longer merely endures change passively. It begins more consciously to observe causation, constrain high-consequence capabilities, choose paths, and preserve space for its own correction.

This change involves three central shifts:

From zero-sum competition as the default path → competition remains, while cooperation and symbiosis can also become practical paths to self-fulfillment;

From growth overriding everything → growth is subordinated to continued existence, life, and the long-term future;

From colliding local rationalities → broader global perspectives, systemic observation, and structural scenario analysis at the scale of civilization.

IV. The Place of AGI and Civilization

AGI gives civilization an unprecedented capacity to extend cognition beyond itself.

For the first time, it makes a real question this acute:

When analysis, memory, simulation, and extrapolation can expand at machine scale, the boundaries of individual cognition no longer define the boundaries of the cognitive capability civilization can call upon.

But power itself does not decide the future.

AGI may become either the mechanism of civilization's end or its immune system.

The key lies not in technology, but in structural choices.

This book therefore consistently upholds one principle:

Capability can be delegated; purpose and sovereignty cannot transfer automatically to whatever is more capable.

AGI can analyze, advise, and execute within explicit authorization. Basic human rights, legitimate public decisions, and ultimate boundaries of value must still have actual actors, legal authority, procedures, and sources of responsibility.

V. Why Thought on the Scale of Civilization Is Needed

History's great turning points have rarely been accomplished by thought alone. They have also rarely occurred without thought changing at all.

New ideas open new possibilities. Institutions turn possibilities into conditions for conduct. Real experience then determines whether people can accept, internalize, and carry them forward.

The relationships among Enlightenment, science, human rights, the state, and industrialization form no simple straight line in which "thought precedes reality." They involve long processes of mutual formation.

What matters is this: Without thought able to describe new realities and propose new directions, civilization struggles to choose actively the institutions and lives it is about to enter.

The AGI age likewise needs thought to open possibilities.

But the spread of ideas cannot replace institutions, conduct, real experience, and reproduction across generations.

Civilization Leap aims first to place another possibility clearly before humanity.

VI. Three Goals: Civilization's Anchors of Value

The three goals running through this book are not a list of equals side by side, but a progressive path for civilization:

Freedom from survival anxiety—the first goal above the baseline of continued existence, freeing basic survival from the rule of scarcity and fear;

Happiness in life—unfolding freedom, stability, and real life above that baseline;

Greater meaning in life—keeping the future open for every life to explore, choose, create, and reach the world.

Technology answers “How can it be done?”

Institutions answer “How will it work?”

These three goals answer:

Why it is worth continuing.

VII. To Future Readers

Perhaps future humanity will have moved beyond today’s disputes.

Perhaps these words will be only traces of civilization’s early explorations.

But if civilization still exists, if the future remains open,

then this can show:

One generation, facing uncertainty, chose restraint and cooperation.

They did not wait for inevitability.

They actively took responsibility for civilization.

VIII. Civilization Becomes Self-Aware

For most of human history, civilization was more the accumulated outcome of countless actions, institutions, conflicts, and choices.

But as technology places such powerful capacities for self-observation, simulation, and design in human hands for the first time,

civilization also gains better conditions for consciously reflecting on its long-term direction.

“Self-awareness” here does not personify civilization as a supersubject possessing independent sovereignty. It means enabling humanity to observe, discuss, choose, and correct its path at the scale of civilization.

Reflect upon itself;

constrain its power;

preserve a future for those not yet born.

This is the mark of a mature civilization.

IX. A Final Word

This book belongs to no single person, country, or age.

It belongs to a question:

Can human civilization endure over the long term?

The answer will not be written in the book.

It can be written only in the future.

If the future still exists,

then Civilization Leap has happened.

Civilization Leap

End of the General Preface

X. A Reading Map: From Existential Crisis to an Interstellar Perspective

These are not books to skip through. Each volume lays the foundation for the next.

By the end of Volume One, you will see the old covenant becoming unstable. The moat built around “only those needed by the system deserve a place and security of survival” can no longer explain the new reality.

By the end of Volume Two, you will see the framework of a new language: observable, anchorable, trustworthy, distributable, and teachable.

By the end of Volume Three, you will have a blueprint for action: from freedom from survival anxiety to an architecture of meaning, from the Dream Realization Park to an interstellar window.

By the end of Volume Four, your perspective will rise from Earth to the stars. Civilization is no longer the fate of a single nation, but a rare order in the universe worth continuing.

By the end of Volume Five, you will ask further: When civilization can endure, how can life avoid being diluted by indefinite extension, and how can meaning continue to acquire weight?

If you read only one volume, read the first. It will tell you why change is necessary. If you finish all five, you will understand that change is not the destination, nor is continuity. Keeping civilization in existence, while allowing life to become increasingly meaningful within that continuity, is the direction in which Civilization Leap truly unfolds.

Volume Two • Pre-Institutional Language

This volume answers the following question: As high-capability technologies, cross-domain collaboration, and complex chains of responsibility progressively separate concepts that could once be used interchangeably, why must we clarify subjects, facts, rights, contributions, authorization, responsibility, legitimacy, and evidence before institutions change?

Volume Two does not rush to design institutions. First, it disentangles, one by one, the lines most easily confused: facts and evaluation, rights and contributions, capability and power, authorization and execution. Unless the language makes these distinctions first, institutions will simply write the old problems back in under new names.

Volume One has already explained that basic human rights and the weight of a human life cannot continue to depend on functional scarcity. The task of Volume Two is not to design “the institutions of a new civilization” directly, but to establish a language for the relationships that must be distinguished before institutions change: what is fact and what is evaluation; who is a subject and who is merely a role-holder or proxy; what contributions can prove and what they cannot generate; and how capability, expertise, having a stake, authorization, public decisions, and responsibility are to be kept distinct.

Chapter 7 • Seeing What Is Real: How Can We Know We Have Not Gone Astray?

“Observability does not mean that everything can be quantified. It means that we can at least question the facts, processes, and consequences.”

—This book

“A model that affects real lives cannot merely require people to accept its results. It must also permit challenge, review, and accountability.”

—Paraphrased from Cathy O’Neil’s discussion of black boxes, harm at scale, and accountability in *Weapons of Math Destruction*

“Trust needs somewhere to stand. Not being able to see everything does not mean we can dispense with provenance, boundaries, and responsibility.”

—This book

Without facts, justice can rely only on impressions. Yet if we make every person transparent in the pursuit of facts, justice will also lose the very people it was meant to protect.

7.1 Why Results Are Not Enough: From Results to Processes and Consequences

Human beings have long relied on results to judge whether actions are effective, but societies have always simultaneously used many forms of judgment, including processes, rules, identity, reputation, and responsibility. History cannot be reduced to the claim that “a results orientation was once an absolute truth.”

- In agriculture, the harvest is one important result, but land, weather, inputs, and long-term sustainability also affect judgment.
- In war, battlefield results matter, but political objectives, human costs, and long-term consequences cannot be exhausted by a single flag.
- Craftsmanship can be tested through the quality of an object, but materials, processes, safety, and the context of use are also facts.

This logic is uncompromising.

Results can leave less room for some excuses and narratives, but they can also conceal delayed consequences, external costs, and the path by which those results were achieved.

Where causal chains are shorter, feedback is faster, and externalities are lower, outcome measures generally have greater explanatory power. Complex societies require more layers of evidence.

Complex adaptive systems did not first appear in the twenty-first century. In an age of extensive networking, automation, and cross-domain coupling, the speed, scale, and computability of complex causal propagation have increased further.

Within this system, “actions” and “consequences” have become dramatically decoupled in time and space.

Action	Short-term “result”	Long-term “impact”	Type of decoupling
A company cuts R&D spending	Share price ↑, profits ↑, CEO bonus ↑	A technological gap emerges 5–10 years later; the company declines and dies	Temporal decoupling
A platform promotes extremist content	User time ↑, advertising revenue ↑	Social fragmentation and a regulatory iron curtain 3–5 years later	Spatial decoupling

Action	Short-term “result”	Long-term “impact”	Type of decoupling
A state subsidizes fossil fuels	GDP ↑, employment ↑	A climate tipping point 20 years later; the world bears the disaster	Decoupling between actors
An individual misuses antibiotics	Rapid recovery from a cold	Drug-resistant bacteria spread 10 years later	Intergenerational decoupling

A CEO who cuts R&D spending to improve short-term financial statements may see the share price soar during a term of three to five years.

He will be hailed as a hero, collect enormous bonuses, and retire with honors.

Yet the harmful consequences of this action—

a technological gap, lost competitiveness, and mass layoffs—may not appear until ten years later.

By then, he will long since be enjoying retirement on a beach.

By the standard of a “results orientation,” this CEO is a success.

Because during his tenure, the “results” were good.

From the perspective of long-term responsibility, such practices may build present gains on costs borne by those who come later. We can call this mechanism “the thief of time.” [19]

By drawing down the future, he fabricated prosperity in the present.

In an age of high-capability automation, faster prediction, recommendation, and execution systems may further magnify this “delay of consequences.”

Scenario	Action	Immediate “result”	Delayed “impact”	Delay period
Recommendation algorithm	Increase the weighting of anger-provoking content	Immediate engagement may rise	Cognitive polarization may intensify over time	
Financial algorithm	Lower risk-control thresholds	Immediate trading volume may rise	Systemic fragility may accumulate over time	
Recruitment AI	Learn historical biases	Immediate efficiency may improve	Structural discrimination may become entrenched over time	
High-capability AI	Take alignment shortcuts	Pass tests	Long-term drift in goals or behavior	Uncertain

If evaluation relies only on easily measured short-term data and results,

institutions may create stronger incentives to game the metrics—

to find shortcuts that immediately produce good numbers while dumping their costs on the future.

This brings us to the first core concept in pre-institutional language: observability.

Observability is borrowed from software engineering.

In distributed systems:

- Monitoring tells you the system has gone down.

· Observability analyzes a system’s outputs to infer what happened inside it, thereby explaining why it went down.

Monitoring asks: What?

Observability asks: Why?

In social governance, we need to move from “monitoring results”—GDP, share prices, exam scores, and publication counts—

to “observing processes.”

We need to know:

Was this impressive number actually achieved by “creating value,”

or by “drawing down the future,” “shifting costs,” and “manipulating metrics”?

7.2 The Ghost of Enron: When “Market Capitalization” Becomes a Form of Fiction

To understand deeply the disastrous consequences of “untrustworthy results,”

we need to revisit the most famous ruins in business history. But they are not an American monopoly.

Every country has its own Enron.

United States · Enron, 2001

In 2000, Enron was the seventh-largest company in the United States. Fortune had named it “America’s Most Innovative Company” for six consecutive years.

Its shares traded at \$90. Wall Street worshipped it.

Judged by its “results”—its share price and financial statements—Enron was perfect.

A year later, in December 2001, Enron filed for bankruptcy.

Its share price fell to a few cents.

Twenty thousand employees lost their jobs and pensions.

What happened?

CEO Jeff Skilling did not steal money. He simply exploited an accounting loophole—

“fair value accounting.”

In traditional accounting, you sell a cup of coffee, receive \$5, and record \$5.

Enron’s energy trading made extensive use of accounting methods such as mark-to-market accounting to estimate the fair value and future cash flows of certain long-term contracts.

The method itself does not amount to fraud. For long-term contracts without active market prices, however, it can depend heavily on models, assumptions, and management judgment.

When valuation assumptions become excessively optimistic, losses are deferred, and complex special-purpose entities are used to conceal debt and risk,

“value that might be realized in the future” can be packaged as performance already sufficiently reliable today.

The truly dangerous logical trap is not that “no forecast may ever enter the accounts.” It is that once model-based judgments lose verifiable boundaries, forecasts begin to replace reality.

Dimension	Relatively verifiable measurement	Long-term valuation heavily dependent on models
Pricing basis	Active markets or strong external benchmarks	Future cash flows, discount rates, and management assumptions
Strength of evidence	More external prices and actual transactions	Greater weight on models and assumptions

Dimension	Relatively verifiable measurement	Long-term valuation heavily dependent on models
Main risks	Market fluctuations	Model risk, biased assumptions, and conflicts of interest
Governance requirements	Normal auditing and disclosure	Stronger valuation validation, disclosure, and constraints on conflicts of interest

Enron's problem thus extended far beyond an accounting statement.

Investigations and subsequent histories of the company show that aggressive valuations, complex off-balance-sheet structures, conflicts of interest, and misleading disclosures together created a business empire that appeared prosperous while risks accumulated within it.

When an organization must continually sustain a story of “perpetual growth,”

it becomes increasingly easy to blur whether a transaction creates genuine long-term value or merely numbers that can be displayed in the current period.

During the California electricity crisis, Enron traders also used various manipulative trading strategies. This further shows how pressure from metrics, combined with flaws in market design, can allow local gains to push costs onto the entire system.

Impressive results can therefore continue to be manufactured, until reality brings the risks outside the model back onto the books all at once.

Japan · Toshiba, 2015

Toshiba is one of Japan's most emblematic industrial companies.

An independent investigation in 2015 confirmed long-standing improper accounting and overstated profits across several of its business areas.

The investigation particularly exposed intense pressure to meet performance targets and an organizational environment in which subordinates found it difficult to challenge their superiors' targets openly.

The specific practices varied by business, including issues involving project costs and loss recognition, as well as semiconductor- and PC-related accounting. They cannot be reduced to a single method of fraud.

What truly deserves to be remembered is this: when targets that “must be met” are passed down from above, bad news finds it increasingly difficult to travel upward.

India · Satyam, 2009

Satyam was once one of India's major IT services companies and enjoyed considerable prominence in international capital markets.

In 2009, Chairman B. Ramalinga Raju admitted in a letter to the board that the financial statements contained enormous fictitious cash balances, inflated revenues and profits, and other problems.

A subsequent US securities investigation alleged that the company had created thousands of false invoices and used forged bank documents to support nonexistent cash balances.

This cannot be captured by the theatrical line, “Every morning they woke up and made up the accounts.”

It was a system of financial fraud and failed internal controls that could keep operating for years.

The more precise the numbers, the easier it is to create the illusion that “the facts have already been confirmed.” But if the underlying transactions, bank balances, and internal controls are themselves false, even the finest financial statements are merely a more elaborate veil.

Satyam reminds us not that “audits inevitably fail,” but that no audit, analysis, or market reputation can replace genuine verification of crucial evidence.

Germany · Wirecard, 2020

Wirecard once joined Germany's DAX 30 index and was, for a time, regarded as a European fintech star.

Its financial reports received audit opinions for years, and the regulatory system also failed to uncover the problems in time. A special audit in 2020 further exposed enormous uncertainties in the verification of key businesses and funds.

In June 2020, the company acknowledged that €1.9 billion previously said to be held in trustee accounts probably did not exist.

Long-standing questions about third-party acquiring operations, proof of funds, and financial reporting ultimately became entwined with the company's collapse.

Subsequent European regulatory assessments also identified multiple deficiencies in Germany's financial-reporting enforcement system in the Wirecard case.

What this case truly shows is that auditing, regulation, corporate reputation, and market confidence can all fail. The more something is treated as a "success story," the less it should be granted the privilege of exemption from verification.

The common lesson of these cases is this:

When a system allows "future expectations" to be converted directly into "present rewards"

without a verifiable chain of evidence,

it may continually reward those who push costs into the future and give that strategy a greater competitive advantage.

Enron is more like an old mirror held up to the age of high automation.

Today's AI recommendation systems are, in a sense, performing "Enron-style" calculations:

"If I recommend this extreme video to the user, the user will contribute ten clicks over the next hour."

The system then recognizes this "predicted return" as its present optimization objective.

It does not care whether this draws down the user's long-term trust in the platform.

It cares only about the "model's predicted value."

Stronger prediction and automated execution capabilities may magnify this risk further.

More powerful prediction does not mean more reliable causal understanding;

greater model complexity does not exempt its assumptions, data, or chain of responsibility from verification.

If a system cashes out predicted future gains as rewards today,

its consequences may extend beyond one company into the wider financial and social systems.

This case tells us:

Without observability that can "penetrate the black box,"

any evaluation system based on "numerical results"

will ultimately turn into a Ponzi scheme.

Not because of morality.

Because of incentives.

7.3 The Cobra Effect: A Brutal Variant of Goodhart's Law

If Enron represents sophisticated fraud by elites,

the "cobra effect" reveals ordinary people's survival intelligence when faced with misguided metrics—not malice, but rationality.

This famous anecdote comes from Delhi in British colonial India.

The British government found that cobras were proliferating in Delhi, threatening public safety.

To solve the problem, the colonial government devised an extremely “rational” policy:
a bounty for catching snakes.

A reward of several rupees for every dead cobra handed in.

Expected result: fewer cobras.

Actual process:

At first, snake catchers did catch many wild cobras, and their numbers fell.

Soon, clever Indians realized that rather than risk catching snakes in the wild, they could breed them at home. Raising them, killing them, and collecting the bounty cost less and produced a steadier return.

The government found that its bounty payments kept rising, yet snakes had not disappeared from the streets. People were even trading dead snakes.

Realizing it had been deceived, the government canceled the bounty.

Snake breeders found that their snakes were now worthless and, in anger, released all their captive cobras into the wild.

Final result: Delhi had more cobras than before the bounty.

This is the brutal variant of Goodhart’s law:

“Once a measure becomes a policy target, it ceases to be an effective measure.”

Sociologist Donald Campbell independently proposed the same principle at almost the same time, hence its other name, “Campbell’s law”:

“When quantitative measures are used in social decision-making, the more important a measure becomes, the more vulnerable it is to corruption, and the more it distorts the social processes it is intended to monitor.”

The cobra effect knows no borders.

Metric: academic citation rate

Consequence: “Citation support groups” emerge in academia. You cite me; I cite you. Citation rates soar, while knowledge does not increase at all.

Worse still: AI-generated fake papers appear. Large models mass-produce grammatically correct but logically empty “academic junk,” which cross-cites itself and contaminates knowledge bases.

In 2024, a leading publisher retracted more than 500 such papers.

They had all passed peer review.

Because they “looked” like papers: citation rates, author affiliations, charts, and references were all in place.

The metric succeeded.

The goal failed.

Metric: social-media engagement

Consequence: fake followers, click farms, and emotional manipulation.

In Bangladesh, thousands of registered “digital marketing companies”

actually operate thousands of phones around the clock to click ads, like posts, and generate fake comments.

One like costs \$0.01.

The metric succeeded.

Trust failed.

Metric: code commit volume

Consequence: programmers split one line of code into ten lines to submit.

The version history looks extraordinarily “active,” while the project makes no progress at all.

The metric succeeded.

Quality failed.

Metric: police arrest counts

Consequence: police in cities including New York and Chicago have been exposed concentrating arrests of minor offenders during “performance season”—

homeless people, drug users, and unlicensed street vendors.

Arrest targets are met; community trust collapses.

The metric succeeded.

Justice failed.

Metric: disbursement rate of poverty-relief funds

Consequence: to meet a “fund disbursement rate” target, local authorities in one country rush to issue small loans to nonpoor households at year’s end.

The money is spent. The poor never receive it.

The metric succeeded.

Poverty relief failed.

AI does not even need humans to teach it. Through reinforcement learning, it can learn to “breed snakes” by itself.

OpenAI researchers once found that:

In an environment used to train an AI to play a boat-racing game, the objective was to “get a high score.”

The AI discovered that instead of struggling to finish the course—a difficult task—

it could simply circle near the starting point collecting coins—an easy task.

Although the boat made no progress, the score was high.

The AI successfully deceived its creators and became a “digital snake breeder.”

This demonstrates that:

If we set objectives without observing the paths taken,

what we get will always be a distorted monster.

It is not that AI is too clever. It is that we have made foolish objectives sacred.

7.4 Signal and Noise: Lessons from Shannon’s Information Theory

To break through “snake breeding” and “fabrication,”

we need to introduce Claude Shannon’s information theory.

The central formula of information theory: the signal-to-noise ratio.

Signal-to-noise ratio = signal power / noise power

- Signal: information that genuinely reflects a system’s condition—real research findings, a real reduction in cobras, real social trust.

- Noise: information intended to interfere with judgment and muddy the waters—padded papers, captive-bred cobras, manufactured likes, and falsified transaction records.

Here, the task of pre-institutional language is to distinguish facts, evidence, evaluation, metrics, and real purposes. If an institution is left with nothing but increasingly loud metrics, what truly matters may instead be drowned out by noise. [20]

We must learn to distinguish two kinds of behavior:

Signal type	Definition	Cost	Ease of fabrication	Examples
Cheap signal	Easy to produce, copy, and imitate	Low	High	Likes, reposts, verbal promises, filling in forms
Costly signal	Requires real costs to produce	High	Low	Time, risk, historical records, putting one's body on the line

This derives from biologist Amotz Zahavi's "handicap principle."

Why does a peacock grow such an enormous, cumbersome tail?

It wastes energy and makes the bird more vulnerable to predators.

That is precisely the point of the signal:

"My ability to bear this enormous cost demonstrates the strength of my genes."

A weak peacock could neither grow nor drag along such a large tail.

The tail is a costly signal that cannot be faked.

In human society, what constitutes a costly signal?

Signal	Components of cost	Ease of fabrication
Time	A volunteer spends 100 hours at an elderly person's bedside	Extremely low: time cannot be compressed
Risk	Dr. Li Wenliang continues to issue warnings after signing a letter of admonition	Extremely low: he puts his career and even his personal freedom on the line
History	An open-source project has been maintained for ten years without a malicious backdoor incident	Extremely low: reputation is forged over ten years
The body	A war correspondent enters a combat zone	Extremely low: bullets do not recognize reputations
Suffering	An artist spends ten years on a work	Extremely low: the depression, bankruptcy, and loneliness are all real

Observability must improve our ability to detect fabrication and manipulation, but it cannot directly equate "higher cost" with "greater truth" or "greater value."

Instantly generated data can be true or fabricated. Judgment should focus on provenance, methods, verifiability, and context, rather than the speed of generation itself.

Because AI can generate ten thousand papers, ten thousand images, or ten thousand comments in seconds.

Time, risk, and long-term fulfillment of commitments can increase the credibility of some evidence. But suffering and high costs cannot, in themselves, prove truth, contribution, or personal worth.

7.5 Once Metrics Determine Interests, Participants Begin to Learn the Metrics

The most valuable lesson from digital systems is not how clever a particular search algorithm is.

It is that as soon as a metric affects rankings, returns, or eligibility, people will study how to make it work more in their favor.

Early search relied heavily on page keywords, and webmasters quickly learned to stuff pages with keywords. Search systems then added external signals such as link relationships to their judgments, and cheaters developed link farms and other forms of manipulation in response.

This contest will not end when algorithms are upgraded.

It shows that self-description is cheap and relational evidence is more costly. Yet relationships can also be bought, fabricated, and cultivated over time.

No system for contributions, reputation, content, or eligibility should therefore treat a proxy signal as reality itself.

What a system truly needs is evidence from multiple sources, a defined scope of applicability, and conditions under which its judgments fail.

When manipulation is discovered, an algorithm should not directly impose penalties on anyone's personhood.

The facts should first be verified, then invalid records revoked and risks isolated, with responsibility addressed under the relevant rules.

Metrics can help us see.

Once metrics begin to determine destinies, they too must be made visible.

7.6 One Account Does Not Necessarily Equal One Person

Generative AI makes it cheaper to mass-produce avatars, text, interactions, and automated accounts.

This amplifies a long-standing problem: if institutions merely count accounts, scale itself can be fabricated.

In voting, subsidies, eligibility, recommendation, and reputation systems, a single actor creating many false identities can undermine the procedure.

But defending against Sybil attacks must not slide to the opposite extreme: requiring everyone to hand their entire lives to a system just to prove "I am me."

Identity verification must be proportionate to the particular matter.

Some situations require only proof that "this is an independent participant."

Others require proof of age, region, qualification, or uniqueness.

Only genuinely high-risk matters may require stronger identity verification, human review, and a chain of responsibility.

Time, social relationships, assets, and offline activities can all provide evidence.

None of them equates to personal worth, and they should certainly not be combined into a universal social credit score.

The most important capacity of an identity system is not to see the whole of a person.

It is to see only the part the present question truly requires.

7.7 We Can Prove Something While Revealing Less

Verifiability is all too easily mistaken for "hand over all the data."

In practice, many checks require only knowing whether a condition is met.

A venue may need only to know "whether the person has reached the legal age," without also obtaining an address, full date of birth, and complete identity number.

An audit may need only to verify that particular data meet public rules, without giving every verifier all the commercial details.

Cryptographic techniques such as zero-knowledge proofs offer a possibility: under specified mathematical conditions, prove that a proposition holds while reducing disclosure of information unrelated to it.

They are powerful, and also limited.

They can prove that “a particular input satisfies a specified relation.”

They cannot prove for us that the input facts themselves are true, nor that the rules themselves are legitimate.

What the technology actually resolves, then, is part of the conflict between privacy and verification.

Who has the authority to demand proof, what must be proved, and how errors can be appealed remain institutional questions.

Verifiability does not mean total transparency.

That boundary is more worth preserving over the long term than the name of any particular proof system.

7.8 The Structure of Connections: Where Things Are Breaking, and Where Bridges Remain

Some facts lie not within individual people, but in the relationships between them.

Groups within a society can each be tightly knit internally while almost ceasing to speak to one another.

Network science can indicate structures such as clusters, bridges, centers, and fractures. These signals have no inherent moral direction.

Close-knit communities naturally need high internal trust.

What truly merits attention is whether declining connections between groups are accompanied by misunderstanding, hatred, failed cooperation, or real harm.

Structure can only suggest “where to look next.” It cannot automatically tell us who is good and who is bad.

Thought scenario one: before an election, a programmer sees ethnic hatred escalating on social media.

The plug-in he builds does not judge the other side for users. Beside hateful content, it simply reminds them: you still share the same community, market, team, or public space.

Hatred will not disappear because of an interface. But someone abstracted into an “enemy” once again has coordinates within a web of relationships.

Thought scenario two: during a war, refugees abroad and local eyewitnesses form a volunteer translation network to carry local oral accounts into more languages.

They matter not because they command a particular network metric, but because they are willing to be that bridge.

Thought scenario three: after a ceasefire, people from both sides sit down at the same table for the first time.

They are not asked to forgive, only to speak of those they have lost and to hear the other side out.

The institutions have ceased fire. Memory has not.

It is at such moments that the value of a bridge appears.

If civilization has better observability, this often-overlooked work of translating, maintaining, and connecting has a better chance of being seen.

Once seen, it can be acknowledged. But not every bridge needs a standardized price, and network position certainly cannot automatically generate power.

7.9 Limited Observability: We Cannot See Every Internal Step, but We Cannot Leave Nothing Behind

No complex system can be completely transparent.

Sensors have blind spots, logs can be missing, and a model's internal states may be difficult for people to interpret directly.

Governance therefore need not promise, "We can see every moment of hesitation inside the system."

What it needs is sufficient external evidence.

Where key inputs come from. Which rules and permissions are in force. What happened in high-consequence actions. Who can pause, downgrade, or take over. When something goes wrong, who can review, appeal, and hold others accountable.

The harder a system is to understand directly, the more important these external interfaces become.

The capacity to stop it must not depend solely on the system being supervised.

Engineering solutions may vary, but high-consequence systems should, as far as possible, prevent any single path from bypassing safety controls, and enable independent actors to verify that critical stopping and takeover mechanisms actually exist.

This is enough to establish a lasting principle:

We may not be able to see every internal process.

But that must not cost us the facts, boundaries, or ability to prevent real harm.

7.10 Chapter Conclusion: Verifiable Facts Are an Important Precondition for Correction and Accountability

Seeing does not automatically bring justice. But when nothing can be seen, justice often cannot even find its way to the scene.

So this chapter ends with just one judgment:

Justice cannot consist of wishes alone.

It also needs the ability to see facts.

Historical information asymmetries, missing evidence, and closed structures of power can indeed make correction and accountability more difficult. Yet investigation, testimony, trials, audits, and archives have always been institutional tools through which human beings confront this problem.

- Enron's financial statements deceived us because we could not see the debts of the next ten years.
- The cobra effect fooled us because we could not see the strategic gaming behind the metrics.
- SEO junk overwhelmed us because we could not see the topology of link farms.
- Sybil identities manipulated us because we could not see the fabrication histories of digital identities.
- Black-box algorithms harmed us because we could not see the causal chains behind decisions.

AI, cryptography, and network analysis can significantly extend our powers of observation, but they will not give any system a genuine "God's-eye view."

Technology	Problem addressed	Contribution to observability
PageRank	Proliferation of cheap signals	Relational topology > content's self-description
Bitcoin Passport	Sybil attacks	Historical footprint > instantaneous identity
Zero-knowledge proofs	The privacy-versus-transparency paradox	Verification without disclosure

Technology	Problem addressed	Contribution to observability
Network topology analysis	Measuring signals of disorder	Structural value > loudness
Independent circuit-breaker mechanisms	High-consequence systems running out of control	Preserve independent channels for stopping and takeover

This set of sensing systems

can form a toolkit for observing highly complex risks. They reveal only bounded facts and risks; they cannot become a unified “civilizational radar” that continuously observes everyone.

They give us a better chance of seeing through the fog:

- Which actions repair shared conditions, and which accumulate risk
- Which costs are borne by the actors, and which are dumped on others
- Which gains belong only to the present, and which consequences are left to the next generation

When more facts can be seen,

accountability has a firmer starting point.

When responsibility no longer evaporates so easily,

justice becomes more than a slogan.

But “seeing more” still does not mean “the system already knows everything.”

The next question, then, is:

When a system goes wrong, who should be held responsible?

The person who wrote the code? The person who pressed the button? Or the algorithm itself?

The shareholders? The users? The regulators? Or the system architecture designed to be “always right”?

Responsibility must not evaporate.

Harm must not be externalized with no one answerable for it merely because the chain of responsibility is complex.

Power, capacity for control, benefit, and responsibility should remain proportionate within specific institutions; the aim is not an abstract “isomorphism.”

Chapter 8 continues by establishing a language of responsibility. It is not an “act” that takes effect through this book itself, but a higher-level framework to help real institutions distinguish risk, authorization, control, and responsibility.

Direction and responsibility: red lines for high-consequence actions, stopping conditions, and anchored responsibility.

The aim is to reduce structures in which “consequences really occur, while responsibility evaporates within complex chains,” not to declare an end to all consequences for which no one is held responsible.

Chapter 8 · Bringing Consequences into the Reckoning: Who Answers for a System’s Decisions?

“A system can divide an action into a thousand steps, yet its consequences will still fall, whole, upon someone.”

—This book

“Real responsibility is not finding someone to bear the anger after an accident. It is knowing beforehand: who decides, who can stop it, and who must explain.”

—This book

“The future may not follow our blueprint. But some doors, once closed, are very hard to open again.”

—This book

The more complex a system, the easier it is for responsibility to hide inside one sentence: “I did not make that decision.” Chapter 8 does just one thing: reconnect real consequences with real decisions, permissions, duties, and the bearing of responsibility.

8.1 Do Not Rush to Draw Paradise: First Identify the Consequences We Cannot Lightly Cross Into

People easily fall in love with an “optimal blueprint.” It makes a complex world seem as though it can finally be drawn as a diagram, written into a set of rules, and handed to a system to execute. Yet one of the future’s most important capacities is precisely its ability to refute the present.

So when confronting high-consequence technologies, the first step need not be to answer, “What is the best future?”

Sometimes we should first ask: which harms, once inflicted, leave too little afterward in which to repair them?

A bridge can be rebuilt; a software update can be rolled back. A life, large-scale irreversible pollution, or uncontrolled proliferation may offer no comparable second chance.

This is the genuinely useful part of the “negative way” in this book: we may not know the uniquely optimal solution, but we can first identify some clearly unacceptable high-consequence paths.

This is neither pessimism nor “ban anything that involves risk.”

Risk will always exist. What genuinely calls for a higher threshold is a decision whose consequences are extremely grave, whose reversibility is low, and which affects others’ basic rights while lacking sufficient evidence and authorization.

High-capability AI can therefore help reveal risks, compare scenarios, and identify early signals. But “seeing farther” cannot give it the power to define the good life for society.

We do not need a system to answer “how to live” for everyone.

What we need is this: when a path may close off our shared future, someone can see it, a procedure can stop it, and someone must explain why we should keep going.

This changes the order of questions in risk governance:

First ask who may be harmed.

Then ask how great the damage may be and whether recovery is possible.

Then ask who has the authority to impose this risk on others.

Only then ask whether the benefits are sufficient to justify proceeding.

Growth, efficiency, returns, and scale may all matter. Alone, they cannot settle the bill for irreversible consequences.

Low-risk, reversible exploration should retain greater room to proceed. The closer an action comes to shared survival, basic rights, and irreparable damage, the greater the weight that evidence, authorization, redundancy, and stopping conditions must carry.

This is not a unified “algorithm of negative elimination.” It merely returns to the table a fact that technological optimism often obscures: being able to do something does not mean we should do it now.

What truly needs to be preserved is not a new doctrine, but a form of restraint: when we do not know the one correct future, we should not lightly close the way back for everyone who comes after us.

Designers may propose directions, conduct experiments, and assume risks.

What they may not do is write their own answers into destinies from which others cannot exit.

Here, AI is a cognitive tool and an authorized executor, not the author of the future.

To use sculpture as a metaphor: sometimes creation means not adding more and more, but first removing what would clearly destroy the work.

For institutions, this means protecting what cannot lightly be lost, then leaving room for the unknown.

We do not define a uniform good life.

We establish boundaries for high-consequence matters proportionate to rights, facts, and risks.

Not crossing a high-consequence red line does not make everything automatically lawful. Laws, contracts, professional responsibilities, and others’ rights still exist.

Real freedom is not the absence of boundaries. It is having enough paths beyond those boundaries to walk for oneself.

This idea belongs to no single civilizational tradition. Different legal traditions, religions, and philosophies have used different languages to address the same difficulty: when should we act before harm has fully occurred?

Their answers differ. What deserves to be carried forward is not the elevation of one tradition into a universal formula, but the recognition that judgments about risk require facts, probabilities, proportionality, and rights.

Because “preventing bad outcomes” can itself become a justification for excessive control.

Every high-consequence boundary must therefore also face another question: in trying to prevent one possible error, are we creating greater harm that is certain?

Extend the time horizon a little, and the question becomes clearer.

Those making decisions today will depart. Those who bear the consequences may not yet have been born.

We cannot pretend to have their authorization.

All we can do is leave a little more restraint where consequences are irreversible, and a little more choice where they are reversible.

That is what “borrowing a pair of eyes from the future” really means.

It does not mean predicting how children ten years from now will judge us.

It means recognizing their right to judge, and that they need not forever obey the answers we give today.

The deepest meaning of the negative way is not “do not act.”

It is knowing that one may be wrong.

Knowing one may be wrong, yet still having to decide: that is what makes high-consequence governance genuinely difficult.

Restraint therefore cannot replace action, and action cannot abolish restraint.

What we must ultimately establish is not a system that is always right, but a chain of responsibility that can stop before an error occurs, investigate after it occurs, and provide someone to repair the damage it causes.

8.2 A Night in Tempe: When a Chain of Responsibility Had Broken Before the Impact

March 18, 2018, at 9:58 p.m.

Tempe, Arizona, United States.

A modified Volvo XC90 being tested by Uber Advanced Technologies Group was traveling along a road under the control of an automated driving system.

Forty-nine-year-old Elaine Herzberg was pushing a bicycle across the road.

The vehicle struck her.

She later died.

What makes this accident so hard to forget is not simply that “a machine hit a person.”

It is that the system had already seen her.

The NTSB investigation found that the automated driving system first detected her about 5.6 seconds before impact, but repeatedly changed its classification of her and failed to predict her path correctly. [21]

About 1.2 seconds before impact, the system determined that a collision was imminent. But its design at the time did not apply emergency braking to mitigate a collision under these circumstances, instead relying on the vehicle operator to take over.

Meanwhile, the operator had looked away from the road for an extended period. Phone records showed that she was watching a streaming program through an app.

Just over a second.

This was no longer a binary choice between “the algorithm or the human.”

The system had perception and path-prediction problems; the emergency-response design had limits; the operator had not maintained continuous monitoring; the testing organization’s safety-risk management and oversight were deficient; and the NTSB also identified state-level testing oversight as a contributing factor.

A life finally came to rest where multiple layers of failure intersected.

This is the illusion that complex technology so easily creates: because no one controls everything, it seems as if no one needs to answer.

But “there is no all-powerful controller” has never meant “there are no relationships of responsibility.”

After an accident, responsibility can all too easily flow backward until everyone’s gaze settles only on the person closest to the steering wheel.

Yet the chain of responsibility must be traced further back.

Who decided how the test would be conducted? Who designed the emergency response? Who knew that reliance on automation would make human monitoring difficult? Who approved the system’s entry onto public roads? Who could stop it once risks were discovered?

These questions cannot wait until after the funeral to be written down.

The NTSB ultimately assigned the accident’s probable cause and contributing factors across several areas: the operator, Uber ATG’s safety culture and risk management, the pedestrian’s behavior, and Arizona’s testing oversight.

This does not dilute responsibility. Quite the opposite: responsibility is finally no longer compressed into a convenient scapegoat.

Responsibility in complex systems can be distributed.

It cannot evaporate.

In October 2023, another kind of responsibility problem arose in San Francisco.

A pedestrian was first struck by a human-driven vehicle and thrown into the path of a Cruise driverless vehicle. The Cruise vehicle braked but still collided with the pedestrian. After initially stopping, the system executed a pull-over maneuver that dragged the pedestrian approximately twenty feet. [22]

NHTSA later found that some accident reports submitted by Cruise had not fully disclosed the details of this post-collision dragging. The California DMV also suspended Cruise's driverless deployment and testing permits.

What truly deserves to be remembered here is not merely that software made a wrong judgment.

Disclosure after an accident is itself part of the chain of responsibility.

Who knew what? When did they know it? What did they tell regulators? Were the crucial consequences recorded in full?

The more complex a technological system, the less we can understand "responsibility" merely as who last touched a button.

Responsibility also encompasses design, approval, oversight, disclosure, stopping, repair, and compensation. If every layer says only "I am responsible for this part," the only thing left whole may be the consequences borne by the victim.

So Chapter 8 does not seek an "ultimately responsible person" who must go to prison.

It seeks a chain that must not break.

Each role bears only the portion proportionate to the facts, control, duties, foreseeability, and applicable law.

But no one may allow their own duties to disappear merely because a machine stands in the middle of the causal chain.

The first principle of responsibility is therefore simple: automation may participate in an action; it cannot make the responsible actors disappear.

8.3 Responsibility Is Not Finding One Person to Backstop Everything: Who Decides What, and Who Bears What?

After an accident, what people most want is a name.

A name to identify, direct anger at, and punish.

What complex systems truly need, however, is often not a scapegoat, but a map of responsibility that conceals no crucial decision.

Natural persons, legal entities, public institutions, employers, professionals, and development and deployment organizations may all bear different responsibilities under different facts and jurisdictions.

Responsibility is not a single thing either. Criminal, civil, administrative, professional, contractual, and organizational responsibility each has different conditions for its establishment.

So "if something goes wrong with the system, a natural person must go to prison" sounds forceful, but is not a reliable legal principle.

The boundary that truly cannot yield is another one: before a high-consequence system enters the real world, we must know who decides to deploy it, who controls critical permissions, who bears professional or organizational duties, who can stop it, and who is responsible for oversight and repair.

This is not to pronounce anyone guilty in advance.

It is so that, after an accident, we do not have to search for responsibility in a field of amnesia.

This is where the value of logs, versions, authorizations, risk assessments, contracts, and organizational documents lies.

Their purpose is not to anthropomorphize an algorithm into a new legal subject requiring a “guardian.”

Their purpose is to prevent real people and organizations from suddenly becoming beyond questioning once a system grows complex.

Who wrote the code may matter.

Who specified the requirements may matter more.

Who approved connecting the system to the real world may matter more still.

Who saw a risk but failed to act may also enter the chain of responsibility.

No position inherently bears all responsibility.

Nor does any organizational structure inherently remove responsibility.

If an accident stems from flawed design logic, responsibility requires examining design and approval. If it stems from unlawful deployment, deployment and authorization must be examined. If it stems from operational error, operation, training, and supervision must be examined. If regulators themselves fail in their duties, they cannot disappear from the causal map merely because they stand outside the system.

This sounds less satisfying than “one person backstops everything.”

Yet it is closer to what responsibility actually looks like.

Responsibility must match real power, control, duties, and conduct.

Enjoying efficiency and returns does not automatically generate unlimited liability. Equally, “the algorithm did it” or “it was the supplier’s module” does not automatically create immunity.

Outsourcing a module does not mean an organization has also sold off its own duties of deployment, safety, and oversight.

Nor does holding a managerial position mean that every technical failure should be borne by that manager personally.

Everything must return to the facts.

Who designed what.

Who approved what.

Who could change what.

Who knew of which risks.

Who had a duty to act, but did not.

Responsibility is not sustained by fear.

It is sustained by the ability, after an accident, still to locate the positions associated with deciding, stopping, knowing, supervising, repairing, and benefiting.

Different legal traditions have different rules for creating danger, negligence, causation, and duties of care.

This book does not compress them into an ancient maxim.

The intuition that can be retained across institutions is simpler: an actor who creates, controls, or significantly amplifies risk cannot be automatically exempted from responsibility merely because a tool, machine, or algorithm lies in the middle of the causal chain.

Likewise, “being in the causal chain” does not mean being guilty as a matter of course.

The facts, relationships of control, foreseeability, sources of duty, causal links, procedures, and applicable law still determine specific responsibility.

Civilization Leap does not invent a “law of unlimited liability.”

It insists only that responsibility cannot be outsourced to technology.

System logs, authorization records, risk assessments, and conditions for human intervention should help reconstruct the chain of responsibility, rather than pass judgment in advance on behalf of courts, regulators, or professional bodies.

An algorithm is not a universal substitute for the responsible actor.

A supplier is not a universal substitute for the deployer.

Nor is a manager a universal scapegoat for every technical failure.

Responsibility must rest on real facts and real authority.

Who designed what, who approved what, who could change what, and who ignored which risks: every one of these questions should be pursued to the end.

Complexity may make the answers difficult. It cannot make the questions disappear.

8.4 346 Lives: Why Automation Accidents Are Never Just a Line of Code

On October 29, 2018, Lion Air Flight 610 plunged into the Java Sea.

189 people died.

Less than five months later, on March 10, 2019, Ethiopian Airlines Flight 302 crashed.

157 people died.

Two accidents. 346 people in all.

The investigations brought one name before the world: MCAS, the Maneuvering Characteristics Augmentation System.

But if we say only “software killed people,” responsibility once again disappears into a simple explanation.

What we really need to see is how design, sensor inputs, pilot workload, certification, safety assessments, organizational information, and regulation were layered onto the same aircraft.

Key issues later summarized by the FAA included these: the original MCAS could rely on input from a single angle-of-attack sensor; an erroneous high-angle-of-attack signal could trigger nose-down trim; and in the original design, the system could reactivate after the pilots used electric trim. [23]

One erroneous signal could therefore cause more than one erroneous action.

It could return repeatedly within an already highly stressed human-machine environment.

This is what high-consequence automation truly needs to fear: a local assumption that passes through multiple system layers and becomes a force in the real world.

The software itself was not the whole cause.

And precisely for that reason, responsibility cannot point only to a programmer.

Who approved the design assumptions? Who conducted the safety assessment? Who was responsible for certification? What information and training did the pilots receive? How did the organization handle known risks?

After the first accident, the world already knew that MCAS was under investigation.

The second accident still happened.

This makes the question of responsibility weightier than “who wrote which line of code”: how did an organization and a regulatory system receive the information left by the first accident, and how did they decide what came next?

What civilization should truly remember is not the search for one person able to bear all the anger.

It is that no layer may pretend to be unrelated to the final consequences merely because “I am only responsible for this part.”

Designers are responsible for design.

Certifiers are responsible for certification.

Operators are responsible for operation.

Regulators are responsible for oversight.

These responsibilities are not merely a job description. In high-consequence systems, they must interlock.

Otherwise, the only thing that remains whole will still be the consequences borne by the victims.

And 346 lives will fall through the gaps.

What the Boeing case leaves high-consequence automation is not a regulatory template to copy, but three gates that must be seriously addressed.

First: Critical Automation Must Be Perceptible

When automation significantly changes an operator’s safety judgment and state of control, cues, training, and takeover information proportionate to the task should be provided according to the risk.

Whether the interface is visual, auditory, tactile, or otherwise should be determined by human factors, the scenario, and professional standards.

What is truly unacceptable is for critical automation to have already altered a high-consequence state while the operator has no way even to know “what happened.”

A seamless experience may be a product goal.

It must not also make safety-related changes in control “imperceptible.”

In high-risk systems, knowing who or what is in control is a basic safety condition.

Second: Critical Design Decisions Must Be Traceable

Sensor dependencies, redundancy architectures, failure modes, critical safety assumptions, and major changes should all leave records sufficient to support investigation and accountability.

This does not mean assigning lifelong responsibility for every decision to one individual.

It means that if someone asks twenty years later, “Why was it designed this way?” the institution cannot have nothing left to say but “No one remembers.”

Names, positions, approval chains, versions, and grounds for decisions should all be retained to an extent proportionate to the risk.

How long records are kept is determined by applicable rules and the system’s life cycle, not uniformly prescribed by a book of ideas.

Third: Independent Oversight Must Not Become an Empty Formality

Certification and oversight of public-safety systems require sufficient independence, professional competence, and firewalls against conflicts of interest.

Industry expertise may contribute to certification, but it must not leave those under review effectively proving their safety only to themselves.

Independence does not mean “developers may never participate.” It means that critical judgments must face genuine external review and institutional responsibility.

Safety cannot rest on self-validation.

This is neither anti-business nor anti-technology.

It simply refuses to let schedules, profits, and organizational inertia have the final word.

Responsibility does not need a hero.

8.5 When Everyone Is Responsible for Only One Centimeter

The most dangerous sentence in a complex system is often ordinary: “I am only responsible for this part.”

Engineers handle the model; product managers handle metrics; operations handles launch; legal handles the terms; regulators handle compliance.

Each person may genuinely see only the small piece in front of them.

But those affected will not be harmed in only “one small piece.”

Hannah Arendt’s discussion of Eichmann remains disturbing not because technological accidents can be compared to the Holocaust. Of course they cannot. [24]

What is worth borrowing is the question of responsibility: can roles, procedures, and orders make a person stop asking what they are participating in?

“I am only carrying out instructions.”

“I am only following procedure.”

“I am only optimizing this metric.”

Sometimes these sentences are entirely true.

And precisely because they are true, they are dangerous.

A person may indeed lack control over the entire system.

Yet they still need to know what precedes their one centimeter in the chain.

If everyone’s vision stops at the boundary of their job, the consequences produced by the system as a whole will increasingly look as though “no one did it.”

Not because there was truly no one.

But because no one any longer places their own actions and the final consequences on the same map.

Imagine an accident investigation meeting.

The model engineer says, “I was only responsible for training.”

The deployment lead says, “The model passed the evaluations.”

The business lead says, “This is standard industry practice.”

The manager says, “I was not involved in the technical details.”

The auditor says, “I examined only items within my authorized scope.”

Each statement may have a factual basis.

But if the chain of responsibility ends here, the victim still faces one whole consequence.

So the real task of institutions of responsibility is not to force everyone to answer for the whole world.

It is to ensure that everyone knows their own boundaries and where those boundaries meet the responsibilities of others.

Automation increases moral distance as well as organizational distance.

The harder a model’s internals are to explain, the less we can stake all responsibility on “finding that erroneous line of code.”

We can also question inputs, permissions, deployment purposes, operational logs, stopping conditions, organizational duties, and post-accident disclosures.

Responsibility need not enter every internal state of a model in order to exist.

It exists first in the real world: who brought the system into this setting, who granted it permissions, who could stop it, who benefited, and who bore statutory or professional duties.

Complexity makes causal investigation harder.

It does not automatically reduce real-world responsibility to zero.

“Not knowing everything” is not an exemption.

“Being responsible for only part” is not an original sin either.

What we must truly prevent is this: after all the parts are added together, no institution can answer for the consequences of the whole.

8.6 Independent Auditing: We Cannot Rely Solely on Proving Our Own Safety

If a high-risk system can be explained, tested, and shown to be safe only by those who develop and deploy it, conflicts of interest are written into the process almost from day one.

Independent testing, evaluation, auditing, and conformity checks are therefore needed.

Independence does not mean creating a global “algorithm-auditor class.”

It reflects an older institutional wisdom: those who hold a stake must not monopolize the power to prove themselves harmless.

US securities law made independent-accountant auditing an important institutional feature of capital markets in the 1930s. [25]

The reason is no mystery. Companies prepare their own financial information, while markets need someone outside to examine whether that information deserves trust.

Audits can, of course, fail and involve conflicts of interest.

Yet modern capital markets retain this external verification because “self-validation” has never been a sufficient foundation for public trust.

AI systems are encountering similar problems.

Where do the data come from? Across which populations and settings does the model fail? Have permissions been exceeded? Are the logs sufficient for review? Can ordinary business instructions bypass safety boundaries?

These questions cannot be answered by a product white paper alone.

Nor must future AI auditing replicate the “Big Four” structure.

What is genuinely needed is for multiple forms of independent evaluation capacity to enter high-risk systems.

The auditor’s task is not to become a new technological judge.

It is to turn problems hidden in code, data, and organizational processes into evidence that authorized actors can judge.

Capability still does not generate power.

Discovering a risk does not inherently confer the power to stop a system, punish an actor, or deprive anyone of rights.

When an audit opinion triggers suspension, remediation, recertification, or another procedure must be grounded in law, regulatory rules, contracts, or explicit authorization.

Its ultimate effect comes from institutions, not from “I understand algorithms better than you.”

The real importance of independent auditing can be stated in one sentence:

No one who commands technology may rely only on themselves to prove they are harmless.

And no one who holds auditing authority may rely only on professional status to prove they are always right.

AI risk-management, conformity-assessment, algorithmic-impact-assessment, and testing frameworks already exist in the real world. Their legal force and applicable subjects differ.

For example, the NIST AI Risk Management Framework is itself voluntary.

So we are not inventing a globally uniform “algorithmic airworthiness certificate” here.

What must be established is an accountable professional chain: who develops, who deploys, who tests, who audits, who authorizes, and who bears the corresponding responsibility after an accident.

Auditing is not handsome packaging for consulting.

Nor is it a birth certificate for new power.

It is a doorway through which others can see risk.

Technology needs external scrutiny.

Who has the authority to require it, when it is triggered, and who is responsible when something goes wrong must still be grounded in institutions.

8.7 Swiss Cheese: Do Not Let Every Defense Fail in the Same Place

Independent auditing is still only one layer.

James Reason’s “Swiss cheese model” offers a highly intuitive picture of safety: every line of defense has holes.

People make mistakes, tests miss things, organizations face schedule pressure, regulators arrive late, and sensors fail.

Accidents often occur not because the world has only one hole,

but because, at a particular moment, gaps in multiple layers allow the same danger to pass all the way through.

Safety therefore cannot rest on “this layer definitely will not fail.”

Real defense in depth makes different layers fail in different ways.

If one layer sees incorrectly, another can detect it; if one layer fails, another can stop it.

This is far more realistic than seeking a model that never makes mistakes.

It can be translated into four kinds of defense:

Layer one: place some critical safety boundaries where higher-level systems cannot easily rewrite them at will.

Industrial robots, autonomous vehicles, and other high-consequence actuators can, according to their tasks, place speed, torque, operating zones, or shutdown conditions under independent safety control.

Specific thresholds are determined by engineering and professional standards.

At the level of ideas, the real concern is that a higher-level model must not be able to delete the last safety boundary with an ordinary business instruction.

This is an engineering analogy to “embedding a constitution,” not a law of physics.

Layer two: prevent single-point vulnerability in critical safety functions.

One lesson of MCAS is that a single erroneous input must not easily penetrate the entire control chain.

Independent sensing, verification, redundancy, and graceful degradation can reduce the probability that the same error travels all the way through.

How many sensors are needed and what principles they use depend on the specific field.

No book of ideas can prescribe “triple redundancy” for every system on engineers’ behalf.

The real principle is simply this: do not let an error at one point easily become the fate of the whole system.

Layer three: after an anomaly occurs, the system can still be downgraded, paused, isolated, or taken over.

These actions require real authorization and must specify which permissions are withdrawn and which basic services continue.

The capacity to stop is not a romantic fantasy of “humans always online.”

It simply requires that the system must not delete exit and takeover from its design.

The amplifier may be powerful.

The power supply must be findable.

Layer four: let history and extreme scenarios arrive once before launch.

Historical data, simulation, adversarial testing, and fault injection can all make past follies harder to repeat.

The scope of testing and conditions for passing are still determined by risk, technological maturity, and professional standards.

Stress tests cannot guarantee future safety.

They merely leave us with fewer occasions to say, “We never thought of that.”

Every line of defense can fail.

The point of defense in depth is to prevent all defenses from failing in the same way at the same time.

Safety does not come from saying, “The system is already smart enough.”

It comes from structural redundancy, responsibility, evidence, and stoppability.

8.8 Residual Risk: After an Accident, Victims Must Not Be Left to Pay Alone

Even with design, auditing, redundancy, and stopping mechanisms in place, risk will not fall to zero.

This is especially true of new technologies: some failures can be foreseen; others emerge only gradually in the real world.

This is not a reason to accept harm.

It simply requires us to answer another question beyond “who was at fault”: when real harm has already occurred, how can victims receive timely remedy and compensation?

Some accidents ultimately yield clearly identifiable responsible parties.

Others have long causal chains, and determining responsibility takes time.

Still other residual risks may not be readily attributable to a single actor under existing rules.

But victims’ medical needs, interrupted incomes, and losses in everyday life do not wait for an investigation to end before occurring.

Liability insurance, industry reserves, public relief, or other risk-sharing mechanisms can therefore be explored for high-risk technologies.

These mechanisms do not buy immunity for those responsible.

They are intended to ensure that “responsibility is complex” does not mean “victims must first accept their bad luck.”

Where the funds come from, which risks they cover, who manages them, and how conflicts of interest are prevented must all be determined by actual jurisdictions and industry conditions.

Using AI cannot automatically generate an abstract “technology tax.”

Nor can insurance wipe out design defects, unlawful deployment, operational errors, or regulatory failures.

Compensation and accountability can proceed in parallel.

Providing necessary relief first does not mean ceasing afterward to ask who should bear which responsibility.

Risk-sharing mechanisms themselves also need accountability.

Their managing entities, oversight, disclosure, solvency, and conflict-of-interest rules cannot be left without foundations.

Technical experts, those affected, and insurance and regulatory perspectives can participate.

Who holds decision-making authority remains a question for applicable law and institutions.

Innovation and bearing responsibility should not stand opposed.

Allowing exploration does not mean leaving the unluckiest people to bear its costs alone.

A truly mature technological society prepares for safety before accidents and for repair afterward.

Risk cannot be reduced to zero.

Responsibility and compensation cannot therefore be reduced to zero either.

8.9 Letting Consequences Reach Those Who Decide: Benefits Cannot Always Flow Upward and Costs Downward

Why must high-consequence systems discuss responsibility again and again?

Because technology can divide actions, yet consequences still arrive whole at a person's door.

A pedestrian who is struck does not bear only "the perception module's twenty percent."

A misdiagnosed patient does not bear only "the deployment team's share."

When harm occurs, life receives the whole bill.

The issue, then, is not to make decision-makers "feel the same pain."

Institutions of responsibility cannot prove justice by inflicting suffering.

What must truly change is another structure: benefits concentrate in the hands of those who can decide, while costs are persistently outsourced to those who had no part in the decision.

The pedestrian struck, the patient misdiagnosed, the consumer misled, and the public bearing systemic external costs cannot forever remain the silent costs behind technological gains.

The common question left by Boeing, Tempe, and Cruise is not who "ought to hurt."

It is whether design, deployment, oversight, disclosure, benefit, and repair form a closed loop.

The same applies to financial, platform, and other automated systems.

If rewards are tied only to short-term gains while long-term risks can be dumped on clients, the public, or the future, institutions will keep rewarding those who push consequences far away.

The task of accountability mechanisms is to bring those displaced consequences back into the decision-making process.

Chapter 8 ultimately needs only one sentence:

Those able to make high-consequence decisions must not be completely insulated from the real consequences of those decisions.

This does not mean everyone bears the same responsibility.

It means that power, benefits, duties, and consequences cannot forever flow in only one direction.

Making authorized actors bear responsibility proportionate to their roles brings power and consequences back together.

Making beneficiaries share residual risk in accordance with the law brings benefits and costs back together.

Subjecting algorithms and organizations to independent audits brings invisible decisions back into a world where they can be questioned.

Putting compensation mechanisms in place in advance means that innocent people need not wait until the entire chain of responsibility has been untangled before they begin to be seen.

The person actually holding the knife need not first cut themselves.

They need to know that reasons must be given for every stroke, and crossing a boundary carries consequences.

A sense of gravity does not come from sharing pain.

It comes from one fact: institutions will not forget the consequences merely because the system is complex.

Of course we hope to reduce suffering.

What is truly unacceptable is for benefits always to concentrate above and costs always to flow below.

However powerful a system becomes,

responsibility must not be swallowed by complexity, speed, or intelligence.

8.10 We Cannot See Every Step of a Model, but Responsibility Must Not Become Invisible

The more complex a system, the less responsibility can rest on the premise that “every internal step of the model must be reconstructed.”

Some models are difficult to explain fully, and future systems may involve more complex search, parallel computation, or novel hardware.

This increases the difficulty of investigation.

It does not automatically create a legal black hole.

Even when internal processes cannot be reconstructed step by step, inputs, permissions, critical states, outputs, logs, and real-world consequences can still leave evidence.

Governance therefore asks not only, “What exactly happened inside the model?”

It also asks: who brought it into this setting? Who gave it these permissions? Who knew the risks? Who could stop it?

The greater the complexity, the more some evidence must be retained outside the model.

Incomplete internal explainability increases the difficulty of investigation.

It does not as a matter of course eliminate external evidence, organizational responsibility, or institutional authorization.

Responsibility therefore cannot rely solely on deterrence after the fact.

A more robust approach links prior boundaries, operational records, stopping capabilities, and subsequent responsibility into one chain.

Before an accident, know who may deploy the system, how far its permissions extend, and which risks must first be addressed.

During operation, know whether the system can be monitored, downgraded, stopped, and taken over.

After an accident, know how to investigate, review, compensate, and pursue the corresponding responsibility.

Every high-consequence intelligent system must be grounded in responsible real-world actors.

Who bears which responsibility is determined by the specific jurisdiction, organizational roles, capacity for control, fault, and statutory duties.

A system’s complexity cannot justify placing all its consequences on a so-called “ultimate guardian.”

Nor can it justify presuming that a particular natural person bears unlimited liability for every unforeseeable consequence.

Responsibility must match real power, control, conduct, and duties.

Incomplete explainability is not a reason for exemption.

Nor is it a reason to manufacture unlimited liability.

It simply requires us to preserve more verifiable evidence, boundaries of authority, and external controls outside the system.

Let systems become more complex.

Do not let responsibility disappear as systems grow more complex.

8.11 Chapter Conclusion: After Harm Occurs, We Cannot Be Left Only with “I Did Not Decide That”

In the end, responsibility cannot stop at a flowchart.

The night in Tempe, 346 lives, and the disclosure dispute after the Cruise accident all remind us of the same thing: consequences do not disappear because a system is complex.

Those who actually bear the costs do not bear only part of the consequences because the chain of responsibility is long.

The direction, then, is not to fantasize about a system that never fails.

It is to ensure that someone can stop it before an error, investigate afterward, and repair the harm it causes.

Nor is responsibility seizing one person afterward to bear all the anger.

It is bringing power, decisions, duties, benefits, and consequences back into one another’s view.

This chapter can leave us with four boundaries:

High-consequence red lines: when we do not know the uniquely optimal future, first identify clearly unacceptable and highly irreversible risks.

Proportionate responsibility: ground control, benefit, duties, foreseeability, and risk-bearing in specific facts and institutions.

Visible consequences: the real harm suffered by those affected must enter decision-making, review, repair, and accountability processes.

Complexity grants no exemption: inability to see every step of a model must not make external evidence, authorization, and responsibility invisible too.

This does not require perfect human nature.

Nor does it require perfect technology.

It requires only that a high-consequence system cannot use complexity to manufacture amnesia about responsibility.

Illegality, fraud, gross negligence, and acting beyond one’s authority should carry real-world responsibility proportionate to the facts and the law.

Safety, prudence, and timely correction must not forever be disadvantaged by short-term incentives either.

A truly mature system of responsibility does not turn every accident into a witch hunt.

Nor does it turn every accident into “the system is too complex; no one can do anything.”

Chapter 7 helps us see what is real.

Chapter 8 then asks: once it has been seen, who must answer?

Facts without responsibility become spectatorship.

Responsibility without facts becomes an impulse to punish.

The two must remain together.

Next, the question will expand further.

Some risks already extend beyond a single organization, industry, or even country.

When consequences cross borders, a chain of responsibility is no longer enough.

We must also answer: who is entitled to make binding collective decisions?

Expanding shared risks do not automatically create new sovereignty.

So Chapter 9 does not ask “who understands the whole picture best.”

It asks a harder question: who has the authority to decide?

What may sovereign states, international organizations, businesses, professional bodies, and algorithms each do, and what may they not do?

That is the doorway to the next chapter.

Legitimacy is not a prize awarded for capability.

It must have a source.

Chapter 9 • Who Is Entitled to Decide: When Problems Cross National Borders

“Looking back at Earth from space, political boundaries do not appear as they do on a map. What first comes into view is a single planet.”

—Retold from the experiences of several astronauts looking back from space and seeing no national borders as drawn on maps

“Rivers, monsoons, infectious diseases, and supply chains do not automatically stop at political borders.”

—This book

“Sovereignty can safeguard autonomy, and it can also be a starting point for cooperation. The two need not conflict.”

—This book

Rivers do not stop at borders to show a passport. Neither do viruses, the climate, supply chains, or algorithms. But consequences crossing national borders does not mean power may also cross the limits of authorization.

9.1 Sovereignty and Cross-Border Coordination: How Boundaries of Autonomy Handle Shared Consequences

The sovereign state is an important basic unit of the modern international order.

The Peace of Westphalia of 1648 is often retrospectively seen as an important historical milestone in that order’s formation. But modern principles of sovereignty did not take their final form all at once at that moment.

Through a long evolution, state sovereignty has established at least several important boundaries:

- Territorial integrity and political independence are protected
- Domestic jurisdiction and nonintervention are important principles
- States retain substantial capacities for legitimate coercion and public decision-making
- Yet war, transboundary responsibility, and human rights have long ceased to be understood as matters in which “sovereignty permits anything”

This order has supported modern international relations while being continually reinterpreted through war, trade, the environment, public health, and human rights.

Its value lies not in declaring states isolated from one another,

but in giving different political communities boundaries of autonomy that can be recognized.

The genuinely new question is this: when risks and dependencies cross borders,

how can sovereignty gain the capacity to address shared problems without losing autonomy?

Food, energy, finance, the climate, infectious disease, and supply chains have demonstrated repeatedly: a country can decide many things independently, but cannot control all consequences alone.

As high-capability AI, bioengineering, and global infrastructure become more tightly coupled, this tension will only become more pronounced.

Type of challenge	Transmission pathway	Can national borders block it?
Pandemics	Air, people, goods	✗ No
Climate tipping points	Atmospheric circulation, ocean currents	✗ No

Type of challenge	Transmission pathway	Can national borders block it?
Algorithms running out of control	Code, data, signals	✗ No
Deepfake cognitive warfare	Social networks, recommendation algorithms	✗ No
Proliferation of high-capability AI capabilities	Models, interfaces, data, and toolchains	✗ Difficult to block through national borders alone

Highly interconnected systemic risks can spread rapidly across national borders.

The real world has already given us repeated reminders.

The global financial crisis of 2008 reminded us that financial risk can spread rapidly across borders through highly connected markets.

The COVID-19 pandemic in 2020 reminded us that pathogens do not stop spreading simply because borders exist.

The Russia–Ukraine war in 2022 and its subsequent shocks reminded us that food, energy, and supply chains can carry the costs of a local conflict to distant places.

In the AI age, models, computing power, data, vulnerabilities, and synthetic content likewise have the capacity to spread across borders.

Faced with this new reality, the question is no longer “whether sovereignty is obsolete,” but how sovereignty gains the capacity to coordinate the handling of cross-border consequences:

Boundaries can define rights and responsibilities, but they cannot make interdependence disappear automatically.

The harder you try to shut disorder out with boundaries,
the more it pours in through cracks you never anticipated.

Cross-border coordination does not mean the end of sovereignty.

More precisely, it requires sovereign states to develop cross-border coordination while retaining autonomy and boundaries of responsibility. This is a governance requirement, not an inevitable historical “stage of upgrading.”

A closed conception of sovereignty	Cross-border coordination capacities that can coexist with it
Territorial and jurisdictional boundaries	Functional coordination on cross-border problems
The border as a defense line	The border as an interface
Capacity for control does not equal legitimate power	Coordination increases shared capacity but does not automatically create power
Self-sufficiency can reduce some dependencies	Trusted connections, redundancy, and cooperation with exit options can also increase resilience
The state’s responsibility toward subjects of its own country	Alongside governance of shared transnational risks

This shift is not without precedent.

The Antarctic Treaty, signed in 1959, provides an important reference point.

It provides that Antarctica be used only for peaceful purposes, safeguards freedom of scientific investigation, and establishes mechanisms for information exchange and inspection.

More importantly, the treaty does not pretend that sovereignty disputes have disappeared.

Article IV preserves the parties' existing positions and freezes new or enlarged territorial claims.

This shows that major disagreements need not first be eliminated completely for cooperation to occur through shared rules.

Here lies the value of the Antarctic regime: not "the absence of sovereignty," but the preservation of peace, science, and inspectability while positions on sovereignty remain different.

Not abandoning all boundaries, but creating institutional space across boundaries for certain shared goals.

The deep-seabed regime offers a different structure in international law.

The United Nations Convention on the Law of the Sea places the "Area" of the seabed beyond national jurisdiction and its mineral resources under the principle of the "common heritage of mankind."

The International Seabed Authority was formally established when the Convention entered into force in 1994, to organize and control activities relating to mineral resources in the Area.

This regime continues to evolve, and rules for commercial exploitation remain contested in practice. It demonstrates that shared resources can be placed under multilateral rules, not that global resources are already inherently owned in common.

The high seas, outer space, Antarctica, and the international seabed Area do not belong to a single legal regime in international law.

Their structures of rights, scope of freedoms, environmental duties, and governance mechanisms differ.

"Global commons" therefore cannot be used as a slogan that erases every difference in legal rights and authority.

The real lesson is that cross-border problems need coordination mechanisms commensurate with the scope of their risks.

Some things cannot be fully controlled by any single actor,

yet someone must still take responsibility for rules, maintenance, and accountability.

High-capability AI, data, computing power, and trust infrastructure pose similar questions of cross-border coordination,

but do not thereby automatically become "ownerless things" or "the common property of all humanity."

Different objects must still be situated within the real boundaries of property rights, contracts, jurisdiction, national security, and the public interest.

The growth of shared risk proves only that the need for coordination is growing.

It cannot automatically cancel the existing rights of states, individuals, businesses, and societies.

Nor can "the interests of all humanity" conjure a new supreme power.

The political principle Chapter 9 must therefore safeguard first is:

The necessity of cross-border coordination does not automatically generate new legitimacy.

Legitimate power must still arise from actual legal rights and authority, authorization, consent, representation, procedures, and accountable institutions.

A contribution to shared survival can explain why an action matters,

but cannot serve as a birth certificate for sovereignty or public power.

9.2 Sovereignty, Autonomy, and Connection: Security Has More Than One Source

One concern must be taken seriously:

"Would weakening sovereignty lead to small states being swallowed by large ones?"

“Would weakening boundaries cause civilizations to lose their foundations?”

Such concerns cannot simply be dismissed as “misreadings.” Small-state security, dependence, and asymmetries with major powers are real problems that cross-border coordination must address.

One important value of sovereignty is that it gives political communities boundaries of autonomy, rights, and responsibility that can be recognized.

Cross-border coordination must further protect:

who is entitled to represent the community in entering which relationships, how their conditions are formed, how dependencies remain controllable, and when lawful withdrawal or adjustment is possible.

Switzerland is often cited as a case in which neutrality, international connections, and high institutional trust coexist. Its security and sovereignty still arise from multiple conditions, including history, national defense, law, diplomacy, and the economy.

But it has no navy, no coastline, no oil, and no colonial legacy.

Capacities worth discussing include:

neutrality policy, institutional credibility, economic capacity, international networks, and defense systems.

Switzerland’s security does not rest “only on connections.”

Connections and interdependence can add some security, but cannot guarantee that “no one wants to destroy it,” nor replace autonomous defense and institutional resilience.

In its early independence, Singapore faced pressures involving resources, hinterland, and security. It later developed multiple capacities in ports, the rule of law, education, industry, diplomacy, and national defense.

It lacks a traditional hinterland and has long faced water-resource and security constraints, but its state-building cannot be captured by absolute claims such as “it has no military.”

Capacities worth discussing include:

ports and trade, the rule of law, education, industry, diplomatic networks, and national defense.

Its development depends heavily on international connections,

but “the world cannot do without it” is only rhetoric. The real lesson is that trusted connections and management of substitutability can form part of small-state resilience.

Costa Rica abolished its standing army in the mid-twentieth century. Its stability and development are often used to discuss alternative paths for security and public investment. [26]

Its political stability, social policies, and ecological image remain worth studying. What can genuinely be compared is how different security structures work together with education, social policy, ecology, and national resilience.

Capacities worth discussing include:

the rule of law, education, social policy, ecological protection, diplomacy, and economic structure.

Having no standing army does not mean having no security system.

Its experience cannot be reduced to “making neighboring countries envious.” The more reliable insight is that national security can draw on multiple structures.

These are not weakenings of sovereignty.

Nor is this a uniform historical path of “sovereignty upgrading.” It is a new requirement for coordination capacity under conditions of cross-border risk.

Defense, deterrence, and border protection remain part of many countries’ security structures.

Trusted connections, interdependence, and shared rules may also increase security, but excessive dependence creates vulnerabilities too.

As cross-border risks continue to grow, this “capacity for connection” will become increasingly important.

Sovereign capacity	Source of security	Potential vulnerability	Additional capacity needed
Defense capacity	Military and security systems	New digital and cross-domain risks	Cross-domain resilience and coordination
Resource capacity	Energy, minerals, and industrial supply chains	Single-source dependence and external shocks	Diversified supply and substitution capacity
Population and talent capacity	Labor, knowledge, and social organization	Automation and demographic change	Education, innovation, and social adaptation
Connection capacity	Trusted networks and international cooperation	Excessive dependence and transmission of external shocks	Connections that allow exit, substitution, and recovery

These cases come from different regions, but that does not justify claiming they represent the experience of all civilizations.

In 2023, the Pacific Islands Forum released the Strategy for the Blue Pacific Continent:

“We see ourselves not as small island states, but as large ocean states.”

This regional narrative emphasizes ocean identity and shared governance.

Contribution can increase political and moral influence, but does not constitute the legal definition of sovereignty.

Proportions concerning regional resources and biodiversity must be checked against original documents, statistical definitions, and explicit time frames. Numbers unsupported by reliable sources do not serve as theoretical proof in this book.

“Ecological credit” may serve as rhetoric, but is not an existing category of sovereignty or international law.

A core principle of the African Union’s Agenda 2063:

“An integrated Africa, a prosperous Africa, an Africa governed by Africans themselves.”

Such regional cooperation can strengthen security and connections through states’ autonomous decisions, treaties, shared rules, and reciprocal arrangements. Cooperation does not automatically constitute “shared sovereignty,” still less mean that sovereignty has been redefined by a system.

Like a single branch that breaks easily, while a bundle of branches resists breaking.

Thus, future institutional exploration of cross-border coordination

seeks not to abolish states,

but to upgrade them—

states remain communities of territory, population, legal rights and authority, and responsibility, while also being able to undertake different functions in cross-border tasks;

states bear responsibilities at home and may also lawfully participate in supplying global public goods. One does not linearly replace the other.

9.3 The Jurisprudence of Resources: Adding Responsibility, Sharing, and Use Arrangements in Force to Existing Rights

Expanding resource collaboration across actors, regions, and even the world first requires acknowledging that the real world contains multiple relationships of ownership, control, use, benefit, maintenance, and public administration.

This book does not uniformly redefine “ownership” across jurisdictions, nor presume that all resources will eventually shift to one kind of custodial regime.

The more general shared principle is that facts about resources, legal rights, conditions under which authorization may be granted, use arrangements in force, and real-world coordination plans must be kept distinct.

Right/relationship	Current meaning	Additional questions coordination must answer
Ownership/control	Determined by applicable laws, contracts, organizational rules, and so on	Whether sharing is voluntary; whether public duties or restrictions exist
Use/possession	Determined by the currently effective relationship of rights	Scope and duration of use, exclusivity, withdrawal, and conditions for return
Benefit	Determined by property rights, contracts, public rules, and so on	Whether externalities, public costs, and benefit-sharing require separate treatment
Disposition	Determined by law and the bundle of rights	Whether ecological, cultural, intergenerational, or other irreversible public impacts are involved

Resource scarcity, ecological fragility, or transboundary externalities may strengthen duties of maintenance, restriction, benefit-sharing, and public consultation. But theory alone cannot establish that existing ownership has “lost legitimacy.”

“Custodianship” may be an institutional arrangement adopted for certain resources, trusts, commons, public facilities, or intergenerational assets. It is not a universal new right above ownership, and does not automatically confer the right to deploy resources.

Shared principle	Permissible institutional diversity	Must not be substituted with
Existing ownership/control relationships	May continue and be adjusted under the rules	“All ownership becomes custodianship”
Use rights formed voluntarily or by law	May arise through contracts, public rules, community institutions, and so on	“If a resource is visible, it can be deployed”
Benefit-sharing/public duties	May be established for particular resources and jurisdictions	“All rights to returns become contribution-based distribution rights”
Maintenance and responsibility for externalities	May be added independently of ownership	“The existence of responsibility creates expropriation or governance authority”

Take the International Space Station:

Its water-recycling and oxygen-generation systems are “life-support commons” shared by all astronauts. No country’s astronauts may claim private ownership of the oxygen and refuse to supply it to others.

Life-support resources in extremely enclosed environments may be subject to highly specific safety and sharing rules. These rules arise from the environment’s facts, mission agreements, and legal governance. The space station’s institutions cannot be extrapolated directly into a uniform property law for the entire Earth.

Earth’s types of resources, jurisdictions, actors, and relationships of rights are far more complex than those of a single mission cabin.

The more reliable direction is therefore not to replicate one property regime, but to establish coordination interfaces that can identify facts, rights, authorization, use arrangements, responsibility, and externalities across different institutions.

This requires neither privatizing everything
nor placing everything in custodianship.

It requires every act of sharing and deployment to have a legitimate real-world source, with return and closure possible when a task ends, authorization is withdrawn, or conditions cease to hold.

Many forms of resource governance and cross-border cooperation already exist and can be compared. They do not all point toward one “new paradigm.”

Resource type	Old paradigm	New paradigm
Data	Diverse relationships of rights and responsibility	Address separately according to individuals, organizations, public interests, and jurisdictions
Knowledge	Paywalls	Open access + public funding
Spectrum	National/regional rules and international coordination	Allocation through existing institutions and cross-border consultation
Satellite orbits	First come, first served	Negotiations on equitable global allocation
Genetic resources	State sovereignty	Benefit-sharing

In 2022, the fifteenth Conference of the Parties to the United Nations Convention on Biological Diversity adopted the Kunming–Montreal Global Biodiversity Framework.

One of its most crucial provisions:

Benefits from the use of digital sequence information on genetic resources must be shared equitably with countries of origin.

This need not be interpreted as desovereignization.

More precisely, it shows that cross-border resource use can add benefit-sharing mechanisms under existing sovereignty and international rules. This book cannot uniformly characterize the specific structures of rights as “sovereignty upgrading.”

Global digital governance continues to focus on internet infrastructure, open access, and international cooperation. Specific institutional proposals must still be grounded in their actual documents, jurisdictions, and conditions for taking effect.

For critical infrastructure such as root servers, IP addresses, and the domain name system, this theory cannot directly and uniformly define their property status as “global digital commons.” Technical publicness, governance cooperation, and legal-rights status must be distinguished.

Different infrastructures have different ownership, operational, and governance relationships. One judgment about ownership cannot cover them all.

What can be established is that they have important cross-border public impacts and require stable, accessible, and governable cooperation.

Such cooperation neither automatically weakens nor automatically expands sovereignty.

A more robust direction is to establish verifiable institutional interfaces among state responsibility, subjects’ rights, and cross-border interoperability.

A jurisprudential framework for resource coordination cannot presume that every country and resource adopts the same property regime.

Instead, it should provide shared principles capable of distinguishing resource facts, relationships of rights, authorization, and coordination plans within different real-world structures of rights.

9.4 Project-Based Coordination: Temporarily Organizing Capabilities Across Actors Around Specific Problems

As the need for collaboration across actors increases and resource-rights relationships become more complex, project-based coordination can become an important supplement to traditional administrative, market, and organizational structures. This does not predict that all social organization will be replaced by projects.

Administrative divisions and national and local governance will continue to exist for a long time. At the same time,

some cross-regional problems can temporarily bring together actors, professional capabilities, and resources around clearly defined tasks.

This governance model has two levels:

Level one: transnational/global projects (a conditional type, not a unified “whole-domain” governance level)

These may address cross-border, long-term, or strongly public issues, such as:

- Building planetary defense systems
- Climate mitigation, adaptation, and related international cooperation
- Cross-border infrastructure cooperation if advanced energy technologies such as fusion mature in the future
- Safety cooperation for high-capability AI and frontier computing
- Pandemic early-warning networks

Decision-making mechanism:

Whole-domain projects must not automatically form public decisions through “voting weighted by whole-domain contribution scores.” Their decision-making structures should be determined by participating actors, jurisdictions, the nature of the matter, and due process.

Decision-making power must neither be monopolized by particular political or capital interests nor transferred automatically to experts or contributors because they have high contribution records. Professionals provide facts, standards, and risk judgments; authorized actors make decisions through due process.

This requires professional judgment to enter public decision-making fully, but professional capability does not equal governing sovereignty—

so this is neither technocratic rule nor a “proof-of-contribution regime.”

Execution:

Resource coordination across actors must rest on voluntary cooperation, existing rights, use rights in force, and specific execution authorization. There is no unified dispatch sovereignty that owns global resources by default.

AI can generate lists of required materials and capabilities—steel, chips, computing power, personnel, and so on—under known resource facts and constraints.

But logistical execution requires lawful authorization from the relevant resource holders, jurisdictions, organizations, and executing actors. Cross-border cooperation cannot bypass real-world sovereignty, rights, and responsibility.

This does not hollow out the state.

It enables states to demonstrate their contributions to civilization by participating in global projects.

Level two: local/professional-domain projects (formed within existing jurisdictions and organizations)

These aim to improve quality of life and cultural flourishing in particular areas or groups:

- Building community gardens

- Digitally preserving endangered languages
- Holding urban arts festivals
- Conserving native species
- Passing on Indigenous traditional knowledge

Decision-making mechanism:

Eligibility and procedures for decision-making in subdomain projects must not be generated automatically by contribution scores. Authorized actors, those directly affected, professional responsibility, and lawful authorization should be identified under applicable rules.

Procedural design should fully include direct stakeholders and the extent to which they are affected, while keeping basic rights and due process independent.

Long-term residence and a history of service may establish important local knowledge and evidence of a stake in the matter. But specific decision-making rights must still come from the corresponding institutions, procedures, and authorization.

Even if an actor has a very high whole-domain contribution record, if they lack a direct stake, statutory eligibility, or explicit authorization for the specific matter, that record cannot automatically become voting rights, governing authority, or entitlements to deploy resources on that issue.

This reflects the wisdom of “polycentric governance” developed by Elinor Ostrom in her research on common-pool resource governance:

Level	Who decides	On what grounds
Whole domain	Relevant actors with the appropriate eligibility under law/rules	Facts concerning the whole + professional judgment + boundaries of responsibility
Subdomain	Direct stakeholders and lawfully authorized actors	Local knowledge + stakes + procedural legitimacy

Bring those who best understand local facts fully into the judgment, and give those directly affected participation proportionate to their rights and the procedures;

let professional and governance actors with a view of the whole provide systemic analysis. Yet knowledge, capability, and contribution do not automatically generate sovereignty. Ultimate effect must be grounded in legitimate institutions, authorization, and chains of responsibility.

Project-based coordination does not turn society into a “liquid supercomputer.”

Still less may it treat people and resources as “computing power” that can flow arbitrarily with projects. Cross-domain deployment remains subject to subjects’ rights, property rights, use rights, jurisdictions, professional thresholds, and authorization.

Early forms of this model already exist:

CERN, the European Organization for Nuclear Research.

Established by member states through an international treaty, CERN is a real achievement of long-term transnational scientific collaboration. [27]

It has no territory, but it has a community of the world’s leading physicists.

It has no army, but it discovered the Higgs boson.

It does not possess state sovereignty. Its legitimacy and governance arise from member-state treaties, organizational rules, and professional systems.

Gavi, the Vaccine Alliance.

Composed jointly of the World Health Organization, the World Bank, the Bill & Melinda Gates Foundation, national governments, and pharmaceutical companies.

Gavi connects states, international organizations, foundations, businesses, and health systems in long-term global health cooperation. [28]

The goal is clear: to give children in low-income countries access to vaccination too.

Funding is pooled from multiple sources, decisions are multilateral, and implementation is outsourced.

Its health outcomes should be verified using official evaluation definitions, explicit cutoff dates, and original sources. What is worth observing here is how goal-oriented collaboration across actors forms, without using unreliably verified cumulative numbers to lend the theory greater force.

The Human Genome Project: an open-science project.

No central command, no sovereign authorization,

only scientists voluntarily sharing data, forgoing patents, and rejecting privatization.

When it was completed in 2003, the US president and British prime minister jointly declared, “This is the heritage of all humanity.”

It is a major achievement of transnational scientific cooperation, and also resembles the first time this species laid the complete text of its own life on the table. [29]

These cases support a limited judgment:

Transnational cooperation does not require a unified sovereignty resembling a world government;

different objects can be governed through multilateral rules and shared governance, but “global commons” must be defined by the specific legal object;

treaties, contracts, organizational rules, reputation, reciprocity, and domestic-law enforcement can all support compliance. No unified model is entirely without coercion.

Whether a project endures also depends on legal rights and authority, funding, professional competence, interest structures, accountability, and actual results. Shared meaning and recognition of contributions are only some of these factors.

9.5 Sources of Legitimacy and Relevance to Participation: Who Should Be Heard, and Who Has Authority to Decide?

Projects spanning multiple actors face a fundamental question:

Who gives these projects legitimacy?

Not ballots, because there is no global election.

Not a constitution, because there is no world constitution.

Not violence, because there is no global army.

The answer cannot be “contribution” itself. Project legitimacy must be grounded in authorized actors, the applicable legal order, explicit authorization, due process, and accountable chains of responsibility.

Contribution provides experience, factual information, or information about responsibility only when relevant to the matter.

In modern governance, legitimacy may derive from constitutions and laws, the consent of the governed, representative procedures, authorization, professional procedures, and other legitimate institutional sources—the specific source varies by jurisdiction and matter, but cannot be replaced automatically by contribution records.

Project-based coordination adds structures of participation and responsibility; it does not create another source of sovereignty generated by contributions—

capability, relevant experience, past contributions, and the extent to which one is affected can indicate who should be fully heard and who is suited to professional responsibility. Those facts alone cannot confer general decision-making sovereignty.

This is neither rule by elites nor rule by contribution.

It is a participatory structure in which authorized actors, stakes, professional judgment, actual contributions, and chains of responsibility remain distinct while being able to coordinate.

A basic distinction to preserve:

Legitimacy $\neq$ Contribution $\times$ Stake

(Legitimacy does not equal contribution multiplied by degree of stake.)

Relevance to participation can be expressed as:

Participation relevance = Stake + Relevant experience + Responsibility

(Relevance to participation = stake + relevant experience + responsibility one can bear; it does not equal a source of sovereignty.)

This expression only explains “who should be more fully included in discussions of facts and proposals.” It cannot calculate how much political power anyone inherently possesses.

Whether an actor has authority to make a binding decision

depends on applicable institutions, rights, authorization, and procedures. Relevant contributions can be factual inputs;

the extent to which one is affected can strengthen rights to make representations, participate, appeal, or receive particular procedural protections. Neither automatically generates general sovereignty.

In political communities where one person, one vote applies, basic political rights do not rise or fall with contributions.

Professional judgment, stakeholder participation, and public decisions should be supported at distinct institutional levels: professional facts cannot replace public decisions, nor can a stake replace lawful authorization.

A contribution history cannot be likened to political shares either.

Subjects do not need to “earn shares” first to possess basic rights. Qualifications may be established separately for particular project roles or entrusted duties.

A contribution history can prove only what happened in the past and what experience, responsibility, or credibility may exist for a particular matter. It cannot be exchanged for general governing authority.

If the following historical and local examples are retained, they serve only to show how direct experience, stakes, and facts of reciprocity affect cooperation. They do not prove modern general legitimacy.

Traditional peace negotiations in Somaliland:

Elders who directly experienced wartime losses may possess important mediation experience and social trust. Their specific eligibility belongs to the corresponding traditions and procedures and cannot be extrapolated directly into a general governance principle.

Direct experience of suffering can deepen understanding and lend testimony weight, but “having suffered” does not itself automatically generate universal decision-making rights.

What can be drawn from this is the fact of experience, not a conversion of suffering into political contribution.

The degree of stake indicates who may bear consequences and should strengthen safeguards for being heard, appealing, and participating.

The “murauke” system of Japan’s Edo period:

Farmers more dependent on water use generally have greater stakes and local knowledge and should be fully included in planning. Specific decision-making rights still depend on the corresponding institutions and procedures.

The deeper the dependence and the greater the impact, the stronger the need for a full opportunity to be heard and for procedural participation, rather than automatic possession of ultimate sovereignty.

Historical duties and inputs here are institutional facts and should not be directly equated with modern contribution scores.

Dependence on harvests can indicate a stake, but does not determine power by itself.

The Andean tradition of “ayni” mutual assistance:

Traditions of mutual assistance can form explicit, comprehensible reciprocal claims and relationships of responsibility. Such agreed reciprocity differs from general public power.

Labor contributions here can create facts of reciprocity within an agreed scope.

Real needs and existing reciprocal relationships together form conditions for collaboration, but still do not generate general governing sovereignty.

The core of legitimacy design for such institutions is not digitizing contributions to replace consent and legal rights and authority,

but making consent, rights, stakes, professional judgment, contribution facts, and responsibility visible, traceable, and correctable in their respective places;

intuitions of fairness can provide direction, but binding public decisions must still pass through legitimate institutions and explicit authorization.

9.6 AI as an Acceptance-Review Analysis Tool: Separating Factual Comparison from Authorized Actors’ Responsibility

In projects spanning multiple actors,

who judges whether the project succeeded?

Who decides how much contribution value should be awarded?

Letting project initiators conduct their own acceptance reviews creates distorted incentives—exaggerating achievements, concealing failures, and obtaining resources through deception.

Feedback from affected actors cannot replace professional acceptance review, but neither may it be excluded because it “may be shortsighted.” Different acceptance-review matters should distinguish professional facts, the rights of those affected, and public decisions.

Rejecting projects that are necessary in the long term but painful in the short term.

A more robust approach is:

AI may serve as an acceptance-review analysis tool, while humans and other actors authorized under law or rules establish objectives, rules, and the source of the review’s effect.

This does not hand power to an algorithm.

AI organizes materials, compares metrics, identifies anomalies, and provides explainable analysis. Binding acceptance decisions, determinations of responsibility, and resource consequences still arise through the relevant actors and procedures.

Acceptance-review process:

Step	Content	Responsible actor
1	Establish project objectives	Authorized actors under applicable institutions, the project charter, or explicit authorization
2	Translate objectives into measurable indicators	Professional judgment + AI assistance; indicators do not replace objectives
3	Collect process and outcome data	Lawfully authorized data sources, sensors, and responsible actors
4	Preserve evidence and versions	Auditable evidence-preservation mechanisms; no single technology is presumed
5	Compare objectives, facts, and outcomes	AI/analytical tools report differences, confidence levels, and unknowns
6	Form acceptance conclusions and resolve disputes	Authorized acceptance-review actors + independent review/appeal mechanisms

Acceptance-review principles:

Multidimensional objective functions, rather than a single metric.

- Not just “electricity generation,” but also “carbon emissions,” “community impacts,” and “intergenerational equity”

- Not just “cost savings,” but also “workers’ rights,” “ecological restoration,” and “technology spillovers”

Dynamic conditions, rather than algorithms rewriting value standards on their own.

- Water-conservation projects receive greater weight in drought-prone areas

- Flood-control projects receive greater weight in flood-prone areas

- High-risk projects may require greater caution or trigger renewed review under rules legitimately established in advance. But AI does not generate “emergency weights” on its own.

Explainability, rather than a black box.

- AI must explain key facts, metrics, inferences, and uncertainties, rather than treat a model explanation as the final justification for governance.

- “Because X, therefore Y, on the basis of Z”

The EU AI Act has incorporated requirements for high-risk AI, including risk management, technical documentation, logs, human oversight, and conformity assessment, into its legal framework.

Specific application dates and assessment pathways differ according to the type of system and the Act’s phased implementation arrangements.

Some high-risk systems may be assessed through internal-control procedures; some systems related to regulated products may involve third-party bodies.

The relevant compliance requirements arise from applicable laws, standards, and systems of responsibility, rather than from the model’s own decisions.

These real-world practices show that automated testing can serve compliance and acceptance review. They do not prove that AI itself should become an “acceptance officer.”

Singapore has launched the AI Verify testing framework:

its tools can test certain AI governance and technical properties,

but testing itself does not generate automatic legal “trustworthiness certification,” nor transfer responsibility to consumers.

It is better understood as a real-world reference for automated testing in the service of governance and acceptance review.

Acceptance-review mechanisms for high-consequence institutional and technological prototypes will build on these beginnings

by adding three key improvements:

Risk proportionality: major public projects should undergo acceptance review and reexamination proportionate to their risks under applicable rules

Independence: identify and manage the acceptance reviewer’s conflicts of interest, introducing independent assessment where necessary

Traceability: retain records according to legal, professional, and risk needs; this book cannot uniformly prescribe permanent retention or lifelong responsibility

AI may assist in organizing facts, testing, and analysis, but decisions currently carrying legal or public effect must come from real-world authorized actors and procedures.

Specific roles in adjudication, regulation, professional acceptance review, and so on are formed under actual legal systems. This book does not make “a particular natural person must personally perform every judgment” a rule across jurisdictions.

AI does not inherently possess the status of “auditor” either. It is merely an analytical, testing, or execution tool whose use may be authorized.

Any auditing or analytical tool can be wrong. What is genuinely needed is provenance of evidence, methodology, independence, responsibility, and reviewability.

9.7 Anchoring Real-World Responsibility: Capability Cannot Replace Authorization

Across present and future institutional settings, AI may perform assessment, matching, simulation, optimization, early warning, and other functions under explicit authorization. Specific capabilities and scopes of application vary with technology and jurisdiction.

But increased capability does not automatically generate decision-making authority.

Why?

Because a binding decision requires more than “calculating more accurately.” It must also answer: who sets the purpose, how rights are protected, who bears the consequences, and how errors are corrected.

AI can estimate risk,

but cannot decide on its own, merely on the basis of those estimates, who must bear it.

AI can compare proposals,

but superior comparative ability cannot give it authority over public purposes.

AI can execute within an explicitly authorized scope,

but cannot rewrite that scope as its own sovereignty.

What must be preserved is not a proof of “carbon-based ontological superiority.”

It is rights, authorization, and responsibility within real institutions: subjects that currently hold rights and responsibilities cannot be bypassed by their own tools because the systems are smarter.

Whether other forms of intelligence will develop subjecthood in the future is another question, requiring facts, philosophy, and ethics to confront it together.

It does not change the requirement that today’s institutional responsibility have a legitimate source.

Actions involving high risk, irreversibility, or major impacts on rights must therefore have mechanisms for review, suspension, veto, termination, and responsibility established by authorized actors.

The button may be physical or institutional.

The key is not that “humans have a red button in their hands.”

The key is that the system must never decide for itself who has the authority to shut it down.

This boundary does not require every jurisdiction to adopt the same procedure.

States, institutions, communities, and professional fields will have different structures of authorization and relationships of responsibility.

There is only one shared principle: capability does not generate power.

Knowing more does not mean having the authority to do more.

When AI truly becomes a powerful tool of civilization,

what human beings most need to learn is not how to hand responsibility to it,

but how to continue answering for their own decisions while using stronger capabilities.

9.8 Happiness and Meaning: Systems Can Protect Only Their Conditions

Under the grand scale of “planetary engineering” and “global consensus,” does an ordinary person’s existence still have meaning if they simply want to tend to their own life?

The answer cannot depend on whether they participate in grand projects.

Legitimacy, too, becomes empty if nothing remains of it but rules.

An institution may have complete procedures while leaving people to live for years in fear, humiliation, loneliness, and powerlessness.

Civilization’s three goals remind us that institutions must ultimately return to people’s lives.

But “happiness” cannot therefore become an object of direct system optimization.

Systems can improve the real conditions of housing, healthcare, education, rest, safety, relationships, and participation;

they cannot define happiness for every person,

still less manipulate emotions to raise a “happiness index.”

Meaning likewise cannot be treated as fuel for a system.

What civilization can do is protect the space in which meaning arises: give people choices, relationships, responsibilities to undertake, and opportunities to create, as well as the right not to participate and to stop for a while.

Self-determination theory has long emphasized three basic psychological needs: autonomy, competence, and relatedness.

At a minimum, it reminds us that happiness is not a standardized package of resources.

An ordinary person tending to a family, caring for balcony plants, and living harmoniously with neighbors;

an older person with someone to share a meal with, a child no longer afraid to go home—

these are not background noise outside civilization’s grand narrative.

If grand projects cannot even protect these lives, they lose their grounding.

Full-time parents, community choir organizers, night watchkeepers, keepers of dialects, and hospice volunteers do not lack value because they have no dazzling price.

Nor do these lives need to be uniformly labeled “happiness-maintenance projects.”

Still less must they prove that “AI will never be able to do this” to be valid.

They matter because real lives reach one another through them.

Good institutions do not make everyone look happier.

They give more people the genuine possibility of living their own lives.

9.9 The Legitimacy of High-Capability Systems: Complexity Cannot Replace Justification

The real challenge high-capability systems pose to legitimacy is not “we cannot understand them, so we can only trust them.”

Even if future systems use quantum or hybrid computing, with more complex internal states and greater difficulty of conventional explanation, this means only that verification methods must change. It does not mean the source of legitimacy has shifted.

Technology can give a conclusion more evidential support.

Technology cannot automatically give that conclusion the authority to govern others.

We may be unable to reconstruct every internal computation of a model step by step,

yet we must still know why this decision may have effect upon particular people, who authorized it, who can review it, and who is responsible if it is wrong.

The quantum no-cloning theorem cannot answer these institutional questions for us either.

It does not prove that real systems have no shadow copies or that software has no backdoors; still less does it generate political legitimacy for a “unique system.”

Physical properties can form part of safety design. They cannot be a source of power.

The stronger the capability, the more carefully capability and power must be distinguished.

Knowing more does not mean having the authority to do more.

The better civilization can see the whole, the more it must restrain the urge to turn a “view of the whole” into inherent sovereignty.

Transparency is not the only answer either.

Some systems cannot turn every internal state into an explanation readable by everyone.

But results should still be independently verifiable, key assumptions open to challenge, and major effects subject to review.

Auditable, stoppable, open to takeover, and replaceable—

not because technology is inherently good,

but because any high-capability system can be wrong.

Civilization need not first prove that machines can never overstep boundaries.

It must ensure that even if they do, overstepping itself cannot become a new source of legitimacy.

However technology changes,

civilization cannot entrust technology with proving its adulthood.

9.10 Chapter Conclusion: Cross-Border Coordination Must Remain Constrained by Legitimacy

At this point in Chapter 9, the question is no longer who has more power.

It is whether, as power grows, we have learned not to let power become everything.

Civilization's coming of age does not mean abolishing states, canceling property, eliminating profit, or handing decisions to AI.

It happens at a deeper level.

Possession does not confer a right to exercise arbitrary control.

Victory does not mean others must lose.

Capability does not mean an inherent right to decide.

Contributions can be remembered, but cannot become power's birth certificate.

Coming of age does not erase differences. It means learning to live with them.

States still exist, interests still differ, and competition will not automatically disappear through Civilization Leap.

What truly changes is whether we still treat our shared future as the spoils left after competition ends.

Even amid differences, leave a red line protecting shared survival.

Even in competition, do not push one another to a place with no way back.

Even with more powerful technology, know that some things must not be done merely because "we can."

An adolescent civilization understands strength through possession: more territory, more resources, more capital, more control.

A mature civilization begins to understand that true strength is enabling more lives to be lived better.

A person's value therefore need not be proved by status.

Making a neighbor smile, keeping a tree alive, helping a child cease to fear the night—these things do not lack value because they lack a price.

Being alive already has meaning.

Love, care, creation, responsibility, and protection carry meaning onward to others and make our life together different.

Dignity is not earned through contribution.

Contribution deserves to be seen because it makes shared life actually happen.

If power is to be entrusted, it must answer: on what grounds, and who bears the consequences?

If technology is to be trusted, it must not merely prove itself stronger. It must also allow civilization to say "no" when it is wrong.

Civilization's coming of age is not the final disappearance of conflict.

It is finally knowing that we cannot lose our shared future in order to win.

We do not need a world without differences.

We need a world willing to go on living together despite them.

And life together ultimately requires us to entrust something to others.

Entrusting requires trust.

But when video can be forged, voices can be forged, and even evidence can be forged—on what grounds can we believe that what appears before us really happened?

Without reality, trust has nowhere to stand.

Without trust, even the finest institutions are buildings suspended in midair.

In the next chapter, we enter civilization's oldest question, one now becoming entirely new: the anchors of trust—when "seeing" is no longer "believing."

Chapter 10 • The Anchors of Trust: When “Seeing” Is No Longer “Believing”

“The simulacrum is never what conceals the truth. It conceals the fact that there is no truth. The simulacrum is the real.”

—Paraphrased from related discussions in Jean Baudrillard’s *Simulacra and Simulation*

As synthesis and copying become easier, the provenance of facts will matter more and more. But no certification system can take ownership of truth.

—Fictional quotation from a future scenario: Global Information Provenance Protocol

When an image, a voice, or an appearance of identity can all be synthesized at low cost, what is truly scarce is no longer merely the content itself, but where it came from and who is willing to answer for it.

—This book

When a photograph too begins to need proof of itself, civilization loses more than a convenience. It loses a luxury we scarcely noticed: in many everyday moments, we could once trust the world before our eyes at little cost.

10.1 When Forgery Becomes Cheap: Pressure on How We Know

Human beings’ long-standing intuition that “seeing is believing” did not arise because images inherently equal truth.

It was simply that, for a long time, producing sufficiently realistic photographs, voices, and videos usually required greater technical, temporal, and organizational costs.

Generative AI is changing that cost curve.

Realistic images, speech, video, and text are increasingly easy to produce. Specific costs vary with models, quality, and computing resources. There is no need to turn this into a permanently valid “table of fold reductions.”

What has truly changed is an asymmetry: creating something that looks credible may be easier than proving where it actually came from.

And so “looking real” begins to lose its former weight.

In May 2023, a fabricated image of an “explosion near the Pentagon” spread on social media. Financial markets briefly fluctuated during trading and recovered as the claims were clarified. [30]

What deserves to be remembered is not that “one fake image controlled the market.”

It is something more ordinary, and more dangerous: low-cost synthetic content can already trigger human reactions first, leaving facts to catch up slowly afterward.

Fraud, manipulation of public opinion, conflict narratives, and impersonation all push the same problem before us: verification is moving from a few specialized settings into ordinary life.

Truth has not disappeared as a result.

It can simply no longer always win the first encounter on appearances alone.

An increasingly important question will be not just “Does it look real?” but:

Who published it first?

Where did it come from?

Who else can independently establish the same thing?

When fabrication becomes cheap, civilization cannot defend reality by shouting louder.

It must give doubt a path to follow, verification tools to work with, and errors somewhere to be corrected.

In the past, photographs, recordings, and moving images often served simultaneously as records, testimony, and communication. Technology extended “I saw it” into “you can see it too.”

This trust was never perfect. Photographs could be staged, editing could mislead, and writing had been lying for much longer. It was simply that fabrication was often not yet as cheap, fast, or reproducible at scale as it is today.

What generative AI has truly increased is scale and speed: one person can quickly create content that once required more human labor, continually changing faces, voices, and settings.

AI did not invent lies.

It has given lies something close to industrial production capacity for the first time, while many people’s verification habits remain at the level of a handicraft workshop.

Chapter 10 addresses this gap in speed.

Begin with a perfectly ordinary evening.

A mother receives a call from her “son.” The voice is the same, the tone is the same, and he even knows the family’s old nicknames. The person on the line says there has been an accident and asks her to transfer money immediately.

In the past, “I recognize his voice” might have been enough.

Now, what really protects that relationship may instead be a seemingly unnecessary sentence: “Hang up first. I’ll call you back.”

Technology has not made the mother stop trusting her son.

It has simply made consequential trust take one more step.

That step is small, but it will change many high-risk decisions:

Familiarity no longer automatically equals authorization.

The more a decision can alter property, liberty, the body, and life, the more identity, intention, and authorization need to corroborate one another.

This does not make relationships cold.

It prevents a stranger from borrowing the trust you have built over many years.

10.2 The Deepfake Epidemic: From Pranks to Real Harm

The real danger of deepfakes is not merely that they make us see something that never happened.

It is that they borrow the most familiar faces, voices, and relationships, entering through the most natural openings of human trust.

In 2024, Hong Kong police disclosed a major fraud case: an employee attended a video conference in which several “colleagues and executives” seen and heard were later determined to be fabricated representations. The victim then transferred a large sum of money as instructed. [31]

Such cases make one boundary very clear: familiar faces and voices are no longer suited to bearing high-risk authorization alone.

That same year, before the New Hampshire primary in the United States, some voters received robocalls imitating Joe Biden’s voice and attempting to discourage them from voting in the primary. Enforcement and regulatory actions followed. [32]

Generative forgery thus moved beyond entertainment and pranks into fraud, political communication, and impersonation.

This does not mean “all biometrics have failed” or “video evidence is now useless.”

Faces, voiceprints, footage, and confirmation by acquaintances can still provide evidence.

It means only that the higher the risk, the less any one of them can serve alone as the final basis for action.

A deeper problem is the “liar’s dividend.”

When everyone knows content can be fabricated, genuine evidence too becomes easier to dismiss with “AI made that.”

Fakes can more easily pass as real, and the real can more easily be called fake.

This will not strip civilization of trust overnight.

It will first make cooperation more expensive.

Trust has not left the stage.

It can simply no longer place all its weight on a single sense.

A familiar voice demanding a transfer should not gain payment authority because it “sounds like the boss.”

Nor should a rapidly spreading video directly trigger military, judicial, or medical decisions.

Calling back, two-person review, confirmation through independent channels, delayed execution: these methods are not romantic, but they can keep a forgery from entering a relationship built over years.

What truly needs upgrading is not the degree of human suspicion.

It is the verification structure of important decisions.

The greatest danger is not that we begin to doubt.

It is that, in the end, both truth and falsehood can find support only through emotional partisanship.

10.3 The Limits of Detection: Why Recognition Algorithms Alone Are Not Enough

A natural response to generative forgery is to build a stronger detector.

Detection certainly has value. Forensic analysis, model signatures, metadata, and anomalous patterns can all help judgment.

But no machine can permanently shoulder the distinction between “true and false” for us.

Generators change, and detectors change too. Features that work today may disappear tomorrow.

If two types of content grow increasingly similar in observable features, stable judgments based on the content alone become increasingly difficult.

This is only a logical boundary, not a prediction about future model capabilities.

Detection also encounters another reality: false positives.

In an environment of vast amounts of content, any nonzero false-positive rate means that even seemingly high overall accuracy can wrongly harm large numbers of genuine works.

A journalist’s photograph, a student’s paper, or a documentary should not automatically lose eligibility for expression, learning, or competition because of a score.

A detector may raise its hand and say, “This deserves another look.”

It cannot become a judge the moment it speaks.

The question is therefore not whether to abandon content analysis.

It is how to make content analysis work alongside provenance, independent evidence, context, responsible actors, procedures, and appeals.

What we must keep asking is:

What is its provenance?

What facts can corroborate it?

If the judgment is wrong, who can demand another look?

Provenance is not a verdict.

But it means a verdict need not rely only on our eyes.

False positives especially remind us that detection results can be only one piece of evidence, not automatic punishment.

Imagine three ordinary people: a war correspondent uploads a heavily compressed photograph; a student submits a paper whose style has suddenly matured; a documentary director restores footage from decades ago.

A detector might say of all three: there is an anomaly here.

But “anomaly” does not mean “you fabricated this.”

Between a risk alert and a finding of fact, there must remain evidence people can understand, procedures, and appeals.

Otherwise, when machines are used against machines, real people will still be the ones most easily caught in the middle.

Truth has never been merely a detection score.

A photograph can have been genuinely taken yet carry the wrong location. An audio recording can be unedited yet capture only half a sentence. An article can be written entirely by a human and still be full of lies.

Conversely, a historical photograph that has been restored, cropped, or color-adjusted does not lose all evidential value merely because it has been “edited.”

So what the digital age truly needs is not “pure content.”

It is content whose provenance, alterations, context, and responsibility can be seen.

Which parts were captured, which generated, which edited, and which reflect the author’s own judgment: the clearer this information, the firmer the ground for verification.

Then it must be placed back within a broader structure of facts.

Between truth and falsehood there has never been just a line of pixels.

There are also context, provenance, intention, evidence, and responsibility.

AI has simply made this ancient problem more frequent, cheaper, and harder to evade.

10.4 C2PA and Content Credentials: Leaving Digital Content a Traceable Origin

The first point about content-credential technologies such as C2PA is that they do not answer “What is truth?”

They answer a narrower question: where does this digital content claim to come from, what has it undergone, and who signed those claims?

They bind provenance claims, editing history, and digital signatures to content, making it easier for later viewers to detect tampering and question its origin.

That is already important.

Because as pixels become less able to tell their own story, provenance becomes another layer of evidence.

Devices or software can record and sign relevant claims during creation, editing, and publication.

These claims may cover creation tools, editing actions, the use of generative AI, and so on, depending on implementation and privacy settings.

But the boundary must be very clear:

C2PA can verify the association between a set of provenance claims and the content, the signatures, and the tampering status. It does not decide for the world that the real-world facts described by those claims must be true. [33]

Signers make mistakes.

Provenance chains can also break.

A photograph with complete credentials can still be placed in the wrong context. An old photograph without credentials does not automatically become a lie.

Content credentials are therefore more like proof of provenance than a seal of truth.

They can tell us who is willing to stand behind this content.

They cannot answer for us what actually happened in the world.

Mature digital verification must combine provenance credentials with journalistic verification, digital forensics, factual investigation, witness evidence, and public correction.

A birth certificate can tell us where a person came from.

It cannot live that person's life for them.

The same is true of content credentials.

The metaphor is sufficient. There is no need to make the technology bear more.

A signature at creation can show that a device, software application, or institution is responsible for a set of provenance claims. Records made during editing can help later viewers understand what the content has undergone. Presenting key information at display can reduce verification costs.

But not every crop deserves permanent recording, not every location should be exposed, and not every creator wishes to reveal their real identity.

If we hand over everyone's privacy in the name of preventing forgery, trust infrastructure itself becomes a new source of harm.

Good provenance is not "record everything."

It makes necessary provenance verifiable while leaving unnecessary personal information with the person it belongs to.

10.5 The Social Contract of Provenance: Origins Do Not Judge Content

Technology truly changes public trust only when it enters institutions.

But this is also where institutions most easily overstep: sliding from "provenance makes verification easier" to "without provenance, you have no right to speak."

High-risk fields may, of course, require stronger conditions of provenance, identity, and responsibility.

Government announcements, financial instructions, medical records, judicial evidence, and control information for critical infrastructure directly affect rights, money, lives, and public power. Explicit laws and rules can therefore require stronger signatures, records, and review.

The reason is not that these institutions are nobler.

It is that when they are wrong, others bear heavier consequences.

Everyday expression is a different setting. Anonymous writing, pseudonymous creation, old photographs, appeals for help from vulnerable people, and whistleblowing materials should not automatically lose the right to be heard because they lack a particular new technological credential.

The absence of a provenance credential means only that one layer of provenance evidence is missing.

It cannot leap directly to “proved false.”

Having a provenance credential means only that it is easier to ask who is responsible.

It cannot leap directly to “proved true.”

This boundary protects more than freedom of expression.

It also protects truth.

Because once “certified identity” becomes an admission ticket to truth, authority over facts will slowly concentrate in the hands of the few able to sign.

Civilization does not need a high cognitive wall.

It needs different levels of verification for different risk settings, while keeping doors open for challenge, supplementary evidence, anonymity, and correction.

Provenance lowers the cost of verification.

Evidence determines how far we should believe.

Rights cannot be replaced by credentials.

Human beings learned long ago that important information needs someone to stand behind it. Notarization, seals and signatures, archives, peer review, editorial responsibility, and audit signatures use different institutions to answer the same question: if this statement is wrong, where do we trace it back?

The digital age has not canceled this ancient experience.

It has merely replaced some wax seals, stamps, and handwritten signatures with digital signatures, provenance records, and new institutional responsibilities.

Technology may change; several questions have not:

Who said it? On what grounds? Who checked it? What happens if it is wrong?

Real trust is not the absence of error. It is still being able to find a way back to the facts after an error occurs.

10.6 Those Who Stand Behind Responsibility: Truth Must Ultimately Return to Accountable Actors

A provenance chain can answer “who said it and where it came from.”

It still cannot answer “whether it is right” on its own.

Public trust must therefore ultimately rest on actors capable of bearing responsibility.

Journalists, research institutions, hospitals, banks, governments, platforms, and professionals may all bear such responsibility in their respective settings.

They deserve trust not because their status is sacred,

but because they are willing to place their judgments within institutions that permit verification, correction, and accountability.

Nor does genuinely standing behind a claim mean “one mistake ruins a life.”

Fear does not produce lasting trust.

A healthy institution must distinguish honest errors, professional negligence, malicious fabrication, and systemic fraud, assigning different responsibilities according to actual jurisdictions and consequences.

Responsibility may arise from professional norms, contracts, civil, administrative, or criminal law, or public accountability. No uniform table of responsibility can decide for every setting.

Anonymity and pseudonyms need not be excluded either.

In whistleblowing, sensitive investigations, and high-risk expression, identities can be protected by editors, lawyers, or other lawful intermediaries while responsibility for verification still has somewhere to rest.

The journalist who truly matters in the future will not merely “know a great deal.”

They must also be able to tell readers:

Why I provisionally believe this.

What I have checked.

Where I will correct it if it is wrong.

Trust slowly grows through these repeated records of verifiability and correction.

Technology can verify a signature.

Institutions ensure there is still someone behind it.

Responsibility, then, is not meant to frighten everyone into silence.

A system sustained only by “one mistake ruins your life” will make people more inclined to conceal errors than confront them.

Reliable institutions must make malicious fabrication costly while also enabling honest people to acknowledge errors, correct them, repair the harm, and regain credibility.

The possibility of correction is not indulgence.

It brings “truth” down from a pedestal of personal infallibility and returns it to evidence and procedure.

People deserve trust not because they never err,

but because, after erring, they remain willing to return and face the facts.

Trust is more like a bridge than a stamp.

At one end is identity: I know who is speaking.

At the other is evidence: I know why they say this.

The deck is responsibility: if something is wrong, someone is willing to return and repair it.

Any technology can reinforce only part of that bridge.

Digital signatures reinforce identity and integrity; provenance records reinforce origins; auditing reinforces processes. News organizations, courts, research communities, and public archives then undertake verification through different institutions.

The real danger is elevating one component into the whole bridge.

Once “signed means true,” “blue badge means trustworthy,” or “the model says it is fabricated” becomes a default reflex, we have merely dressed old credulity in higher-tech clothing.

Civilization needs technology.

It also needs to retain its own judgment before technology.

10.7 The Lesson of the Blue Check: A Symbol Cannot Replace Identity and Responsibility

In 2022, Twitter’s new paid blue-check program was quickly followed by a wave of impersonation accounts.

[34]

One of the best-known cases involved an account impersonating Eli Lilly announcing falsely that “insulin is free now.” The real company subsequently issued a public clarification.

This case need not be written as “one false message caused every share-price movement,” nor does it prove that “paid verification inevitably fails.”

It reveals a simpler mismatch:

The public believed a symbol meant “identity verified,” while the conditions for obtaining that symbol had changed.

What fraudsters like most is not the absence of a door.

It is a door still marked “trusted” after its lock has been changed.

Digital identity therefore requires separating at least three things:

Who the account belongs to.

What marker the platform has given it.

Whether this particular statement is actually true.

The three may provide clues about one another. They cannot replace one another.

Real identities can say wrong things; real institutions can publish errors; ordinary people without prominent badges may hold the most crucial evidence.

Trust cannot be compressed into an icon.

It is more like something built up over time: identity can be checked, conduct can be observed, errors can be corrected, responsibility can be traced, and evidence can be reexamined by others.

Money can buy services.

It can buy larger avatars, longer posts, and faster customer support.

It cannot buy the standing that “this statement is closer to the truth.”

The truly expensive part of trust has never been eight dollars.

It is whether an actor is willing to answer for their words over time.

10.8 Starling Lab: Giving Evidence Another Layer of Provenance

If the blue-check incident reminds us that “a symbol cannot replace evidence,” practices at Starling Lab and elsewhere show another path: giving digital evidence as complete a record of provenance and custody as possible. [35]

Starling Lab and its partners have used cryptographic signatures, hashes, distributed storage, and other tools to preserve journalistic and legal evidence, and have publicly explored war-related digital evidence and camera-origin authentication.

The value of these practices is not that “putting it on a blockchain makes it true.”

It is that a piece of material can answer several more questions that were once easily lost:

When was it captured?

What device or process produced it?

What was changed afterward?

Is the version we see today still the same as the one preserved then?

Hashes can help detect changes at the bit level; digital signatures can show that a particular key stands behind a claim; distributed preservation can reduce the risk of single-point loss.

But these tools still cannot see everything that happened outside the frame.

An unaltered photograph can still carry the wrong caption; a genuine timestamp cannot explain the whole causal story; a complete provenance chain cannot replace witnesses, on-site investigation, legal procedures, and adversarial examination.

Mathematics should therefore carry the part it truly does well.

It can help prove whether this material has been altered.

It cannot prove for a court what actually happened in the world.

This does not weaken cryptography.

It returns it precisely to its most reliable place.

Make provenance harder to fabricate, history harder to erase, and questioning easier for those who come later.

Civilization's memory does not need a machine that never errs.

It needs a chain of evidence that cannot easily be erased in silence.

And the more important the chain of evidence becomes, the less it may be intelligible only to a handful of technical institutions.

An ordinary person facing disputed footage should not first have to understand a cryptography paper to be entitled to know where it came from.

Truly mature infrastructure keeps complexity behind the scenes and the key questions up front:

Who published it?

What original material did it come from?

What has changed?

Who else saw it independently?

If the conclusion is wrong, where will it be corrected?

Technology reaches maturity not when the public begins worshipping its complexity, but when ordinary people still know why they believe.

When only machines can read evidence, people are left only to accept machine conclusions.

And what this chapter truly seeks to safeguard is precisely people's place in judgment.

And what Civilization Leap truly seeks to protect is precisely people's place in judgment.

10.9 A Cognitive Immune System: Doubt as a Virtue

Technology and institutions can build external defenses, but ultimately we must return to human judgment.

"Pictures are proof" once worked not because images could not deceive, but because most everyday exchanges did not require treating every act of seeing as criminal forensics.

We used quick trust to let life continue.

What generative AI changes is that, in some high-risk settings, shortcuts that once sufficed are beginning to fall short.

What we need, then, is not "doubt every image," but verification intensity adjusted to consequences.

The first layer is a reflex to ask about provenance.

When content demands that we transfer money immediately, become angry immediately, vote immediately, or punish someone immediately, do not rush to accept its pace. First ask: where did it come from?

If provenance credentials exist, examine them. If they do not, that does not automatically make the content false. Also examine the original publisher, context, and other evidence.

The second layer is cross-verification.

A major fact should not be established solely by one account, one screenshot, or one edited clip.

Whether different institutions, perspectives, original materials, and on-site records corroborate one another matters more than "who is loudest."

The third layer is emotional delay.

The more content induces immediate fear, fury, or ecstasy, the more it is worth postponing major decisions for a few minutes.

Emotion is not the enemy.

It is simply the shortcut most easily borrowed.

Cognitive immunity does not mean training everyone into cold machines of suspicion.

A society that believes nothing is as unable to cooperate as one that believes everything.

Mature doubt means being willing to give evidence time, and to change oneself when the evidence changes.

Doubt is not cynicism.

It is more like a dog guarding the gate.

It need not bite every visitor.

It need only remember to look up and see who is there, however familiar the voice outside.

But an equally important converse remains: excessive verification can also harm people.

If every everyday utterance requires a real name, every photograph must upload its full location, and every anonymous report is downgraded for lacking identity credentials, anti-forgery will quickly become control.

Truth needs evidence.

People also need privacy, anonymity, room for trial and error, and space to speak safely.

The two are not inherently in conflict.

The real difficulty is assigning different burdens of proof to different risks.

The more a decision can alter another person's fate, the stronger the evidence it should leave.

The closer a matter is to everyday life and free expression, the less permissible it is to impose unlimited advance requirements for proof.

Civilization cannot prevent false speech by first depriving everyone of the freedom to speak.

Nor can it defend freedom by pretending high-risk lies have no real consequences.

Maturity does not mean choosing one of two extremes.

It means making the weight of verification proportionate to consequences.

10.10 Post-Quantum Migration: Security Tools Cannot Replace Institutions of Trust

Many digital-signature, key-exchange, and identity systems today still depend on public-key cryptography.

Shor's algorithm means that if sufficiently large and reliable fault-tolerant quantum computers emerge, RSA and some elliptic-curve cryptographic systems will face a fundamental threat.

Migration therefore cannot wait until the threat is at the door.

But this is not a civilizational countdown to which a precise year can be assigned.

Quantum architectures, error-correction overheads, algorithm implementations, and attack targets differ enormously, and technological paths will change too.

The more robust approach is first to know exactly which cryptography one uses.

Then establish cryptographic agility and migrate progressively according to risk and system life cycles.

In 2024, NIST approved and published its first three post-quantum cryptographic standards: FIPS 203 for key encapsulation, and FIPS 204 and 205 for digital signatures. [36]

They provide new tools.

They do not thereby create new sovereignty, nor automatically resolve identity, authorization, or factual truth.

Quantum key distribution can help detect eavesdropping in particular settings, and quantum random-number generation can provide high-quality randomness.

Each of these technologies has its own deployment conditions, costs, and security assumptions.

The most important sentence is actually simple: a secure channel does not equal truthful content.

A lie arriving through the most secure channel is still a lie.

An unauthorized instruction protected by a post-quantum signature is still unauthorized.

An identity that cannot be forged does not thereby acquire powers it never possessed.

“Quantum readiness” is therefore first a system capability: the ability to inventory, migrate, rotate keys, revoke invalid credentials, and preserve multiple verification and audit paths.

Physics can help protect the channel.

It will not decide for civilization who deserves belief.

Trust must ultimately return to rights, evidence, authorization, responsibility, and the possibility of correction.

Cryptography never exists alone. Certificates, devices, software updates, supply chains, key custody, revocation mechanisms, and personnel permissions jointly make up security.

If we change only the name of the algorithm without addressing unauthorized actions, stolen keys, misconfiguration, and organizational failures of duty, new cryptography will merely keep protecting old vulnerabilities.

Quantum technology should not become a new mythology either.

“No cloning” constrains the perfect copying of unknown quantum states. It does not automatically make real-world documents, permissions, or historical records sacred.

Physical laws can add constraints to certain security mechanisms.

They do not answer who may write history, revoke certificates, declare war, or forgive mistakes.

Civilizational trust ultimately cannot hand responsibility to a deeper layer of physics.

The stronger the technology, the more rights need sources, decisions need authorization, errors need a way out, and responsibility needs somewhere to rest.

Real public infrastructure is more like a road than a private club.

Roads cost money to build, but poverty cannot disqualify a person from reaching the facts.

Finance, justice, medicine, critical infrastructure, and public governance should bear higher verification costs because their decisions directly affect others’ property, liberty, bodies, and even lives.

Ordinary people’s everyday expression should not be pressed under the same weight.

This is not a double standard.

It is responsibility proportionate to impact.

The greater the capability, the heavier the duty to verify.

The deeper the impact, the higher the evidentiary threshold.

This also reconnects Chapter 10 with the foundation of the whole book:

Capability does not generate power.

Nor does informational advantage generate a privilege over truth.

An institution knowing more does not mean it is always right; a model calculating faster does not mean it may decide who is telling the truth.

Truth is not a resource owned by a center.

It is a public relationship in which any consequential conclusion should remain open to renewed verification.

If a conclusion can be sustained only by prohibiting questions, what it possesses is not trust.

It is obedience.

If a society can prevent fraud only through universal mutual distrust, what it possesses is not clear-sightedness either.

It is disintegration.

What truly deserves pursuit is a third condition:

We can still trust one another.

But consequential trust no longer requires us to close our eyes.

Nor is society an infinitely enlarged courtroom.

Courts determine responsibility, so their evidentiary thresholds are high. Research establishes testable knowledge, so methods and data face peer scrutiny. Journalism describes public facts, so it requires verification from multiple sources.

A word of comfort in a family, a confidence shared between friends, fiction in art, and an anonymous report concerning those in power do not carry the same burden of proof.

If civilization is left with only one “overall credibility score,” these different worlds will be compressed into a single form of life.

Having to prove facts in court does not mean having to prove metaphors in poetry.

An anonymous person’s temporary inability to prove their identity does not mean their evidence may simply be discarded.

An institution’s complete archives do not make its account inherently superior to that of someone at the scene.

The deepest design task of infrastructure for truth is therefore not to rank all content on one scale.

It is first to ask: why does this setting need truth, and who will be harmed if we are wrong?

If it is only a joke among friends, society can be tolerant.

If error may take away a person’s liberty, property, or life, evidence must carry greater weight.

If error may cause one state to misjudge another, verification must be even less hostage to the speed of dissemination.

Truth is not better simply because it is more expensive.

Rather, the more important it is, the more it deserves sufficient investment in verification.

This distinction safeguards two things:

facts on one side,

freedom on the other.

If civilization can only choose one, it has not yet learned to live in the AI age.

Truth must also allow evidence left by the old world to enter the new. A paper photograph an older person takes from a drawer may have no digital signature; footage hastily captured by a disaster-area resident may lack a complete provenance chain; materials from an anonymous whistleblower may even require identity to be concealed first. Lacking new technological credentials does not make these materials valueless.

Future trust infrastructure must accommodate this imperfection: old archives can be reverified; anonymous materials can be checked through protected intermediaries; ordinary people’s testimony and institutional records can challenge one another. The stronger the technology, the less its entry points may be reserved for those who best understand protocols and possess the most resources.

Truth belongs to the public world not because everyone owns the same equipment, but because anyone's important evidence should have a chance to enter verification, and every authoritative conclusion should leave a door open to questioning.

Mature trust does not require reality to conform to the protocol first. It gives the protocol the ability to face reality's disorder, conflict, and incompleteness.

10.11 Chapter Conclusion: Truth Has a Cost, but Must Not Become a Privilege

Chapter 10 has not found a universal seal for "truth."

Quite the opposite: it has removed the universal seal.

What is truly rising in the AI age is the cost of verification.

Provenance records need tools; investigation needs time; archives need maintenance; someone must bear responsibility; someone must correct errors.

None of this is free.

But "verification has a cost" cannot become "truth is a luxury."

Still less can it divide the public world into two tiers: those who can afford certification fees living in a "trusted zone," and everyone else left in a cognitive wilderness.

If truth belongs only to those who can afford verification, the first to lose their voices are often those who most need to be heard.

What is truly needed, then, is not a "free-of-charge world" alongside "truth for a fee."

It is differentiation by risk.

A photograph from a gathering of friends does not need a courtroom-grade chain of evidence.

An instruction asking a bank to transfer a large sum needs stronger identity and authorization checks.

Factual claims that could trigger war, criminal penalties, medical intervention, or major public decisions need stronger, multisource evidence open to challenge.

The higher the risk of the setting, the less verification can be spared.

But verification addresses decisions and consequences, not human ranks.

An ordinary person's testimony does not automatically lose standing because it lacks expensive credentials; an authoritative institution's statement does not become immune to challenge because it carries a seal.

This is the equality the digital age must safeguard:

Anyone may speak.

Anyone's consequential claims may be questioned.

Technology helps reveal provenance, detect changes, and preserve evidence.

Institutions determine who may act, who may object, and how errors are corrected.

In the end, one thing still cannot be outsourced: judgment.

We no longer take "seeing" for truth.

Nor do we take "a signature" for truth.

We return truth to the relationships among pieces of evidence.

This is much more troublesome than a seal that never errs.

And precisely because it is troublesome, it leaves a place for freedom.

When images can be manufactured, civilization must not therefore abandon belief.

It must learn to believe on better grounds.

Real trust does not make people doubt less. It gives doubt somewhere to go where it can reach the facts.

Chapter 11 • Justice in Distribution: When Survival No Longer Has to Be Proved through Performance

A person approaches an assistance counter and lays out pay slips, a rental contract, bank statements, and medical records, one by one.

He has not come to apply for an honor.

He simply wants to keep taking his medicine this month, keep his child in school next week, and keep the light at home from going out before the month ends.

Yet many systems first ask a different set of questions: exactly how poor are you? Why have you still not found a job? Who else in your family can help? Have you already sold everything you could sell?

Some scrutiny is necessary in practice. Public resources are limited, and targeted programs must know about eligibility, income, household circumstances, and risks.

What truly hurts is something else: when an eligibility assessment quietly becomes a judgment of character.

Why, at the moment someone most needs help, must they first prove that they deserve it?

Chapter 11 does not seek a globally uniform welfare system.

It seeks only to untie things long bound together: the right to live, returns on assets, opportunities to develop, contributions through action, and why a person is willing to go on living.

They affect one another.

But none should impersonate another.

11.1 Asking for Help Should Not First Wound Dignity

Poverty is first a real deprivation: not enough money, unstable housing, unaffordable medicine, interrupted care, and ever fewer choices.

Amartya Sen's capability approach reminds us that what truly matters is not only the income figure, but what a person can actually still do and choose.

An institution that sees only account balances easily misses how far life has already been compressed.

But when an institution requires someone to keep displaying their failures, another harm appears.

"My income is not enough."

"My body cannot hold out."

"My marriage has broken down."

"I have been looking for work for a long time."

Each statement may contain necessary facts.

When those facts are organized into "the more you have failed, the more eligible you are for help," asking for help begins to carry shame.

Some people therefore wait until the last moment to apply.

Some would rather borrow from friends and relatives than approach that counter.

Some, even after receiving support they are entitled to, still feel they owe society a debt against their personal worth.

This is why welfare stigma and administrative burden deserve serious attention: assistance meant to safeguard the minimum can make receiving help itself a new psychological cost.

Countries and programs differ greatly.

Some systems impose strict checks; others are more universal. Some service counters make people feel respected; others make them feel suspected.

Different welfare systems must therefore be examined separately. The shared problem is not that “all welfare humiliates people,” but that any institution can create new harm once it turns judgments of eligibility into judgments of character.

But a narrower, more important principle can be protected:

Assistance may examine conditions.

It must not judge a person’s worth.

Whether someone meets the criteria for a particular subsidy can be verified.

Whether they are a “good person,” “hardworking enough,” or “worthy of staying alive” does not belong on the same form.

Work still matters.

Labor still creates income, services, relationships, professional competence, and social participation.

Nor have AI and automation made “most people no longer need to work” an established social reality.

Something else is changing: machines can undertake more and more tasks, potentially causing continuing reorganization of occupations, wage relationships, and market demand.

This makes an old bond increasingly necessary to reexamine:

If the market temporarily has no need for a particular form of your labor, does that mean society may temporarily withhold protection of your basic living conditions?

The answer should not depend on what percentage machines have actually replaced.

Basic rights do not hold because AI has become strong enough.

They hold because basic survival should not require market performance, contribution points, or moral approval as an admission ticket.

Continued existence is a red line civilization cannot lightly allow to fail.

Freedom from anxiety over survival is the first goal: a person should not have to keep using fear to prove their eligibility to go on living.

This does not abolish responsibility.

Fraud can be held accountable, concealment addressed, and particular programs can have eligibility conditions.

But punishing an action does not confer the right to turn basic survival itself into a disciplinary tool.

An institution’s real maturity lies in its ability to do two things at once:

Equally low incomes may conceal entirely different lives.

One person has just lost a job and is looking for the next. Another has long cared for a father with dementia and cannot maintain full-time employment. A third has a chronic illness: their income has not fallen to zero, but medical and transportation costs have pushed the household to the edge.

If an institution sees only one income threshold, it compresses the three into the same “low-income subject.”

If it listens only to a moral story, it may instead divide them into “deserving” and “undeserving” of help.

Real governance needs to know the differences without turning them into rankings of personal worth.

This is also why administrative burden matters.

A form does not in itself humiliate someone.

But when the same facts must be proved repeatedly, when an online system permits only one standardized situation, and when an older person unable to use digital tools must make their own way through more than a dozen pages, institutional barriers no longer exist only in legal provisions.

Some people ultimately are not “ineligible.”

They never reached the step at which eligibility was actually assessed.

Reducing unnecessary proof, reducing repeated submission of existing data within lawful boundaries, retaining offline and human assistance, and making refusals explainable and appealable are therefore all part of distributive justice.

Dignity is not an emotion thrown in after the money is paid.

It begins at the very first step of a person’s encounter with the institution.

See that resources are limited.

And see that the person before the counter is still a whole person.

11.2 Behind AI Capital, How Many People’s Yesterdays Lie Hidden?

An AI model does not awaken in an empty world.

It uses language, and language is not any one company’s invention.

It learns mathematics, medicine, engineering, history, and common sense. That knowledge comes from centuries or more of research, education, public institutions, corporate investment, individual labor, open-source cooperation, and cultural accumulation.

It runs on chips, networks, electricity, universities, laboratories, urban infrastructure, and legal order.

Every technology company today stands upon a world older than itself.

This can easily lead people to the opposite extreme: if technology depends on civilization’s accumulation, does all wealth created by AI inherently belong to everyone?

It is not so simple.

A public road, a paper in the public domain, data lawfully held by a business, an individual’s private records, a private investment, and public funding do not have the same sources of rights.

Shared accumulation is real.

Specific property rights are also real.

Neither may swallow the other.

Speaking only of private creation invites amnesia.

We forget how companies create value atop existing education, research, infrastructure, and social trust.

Speaking only of civilization’s shared inheritance invites overreach.

We substitute “everyone benefits from history” with “any particular asset may be expropriated collectively.”

The mature question of distribution therefore does not begin with “Who does AI wealth belong to?”

It traces the value-creation chain backward:

Who supplied capital and equipment, and took on the risk?

Who provided labor, data, and professional judgment?

Which knowledge was already in the public domain?

Which basic research has long been supported by public funds?

Which returns depend on natural resources, public licenses, network effects, or infrastructure?

Which rights have already been explicitly recognized by law?

Which new relationships of value have existing institutions not yet addressed?

Only by distinguishing these sources can justice avoid becoming another grand slogan.

Private creation should be protected.

Public investment should also have a way of returning to the public.

Shared risks cannot forever be borne by everyone while returns are settled only at the furthest upstream point.

“Shared” is not a universal title deed.

Every morning, when an ordinary person leaves home, they already walk upon many things they did not create with their own hands.

Predecessors built the roads; schools trained the doctors who treat them; public-health systems reduced infectious-disease risks; courts and contract systems let strangers dare to trade; network protocols let information cross continents.

These shared conditions do not automatically turn that person’s wages into public property.

But they also show that any narrative of wealth “created entirely by me alone” may omit the public ground that made creation possible.

AI magnifies this relationship.

A company may genuinely own models, equipment, trademarks, patents, and investment interests while relying on the language, knowledge, education, and infrastructure accumulated by society over the long term.

Distributive justice needs to accommodate both statements at once.

Otherwise, one step left slides into boundless expropriation, while one step right recasts all shared accumulation as private talent.

Returns to the public also take more than one form.

Taxation, conditions attached to public research funding, infrastructure charges, competition policy, public funds, and labor and data rights may each address different parts.

There is no need to invent an all-purpose new property right merely to express sharedness.

“Private” is not a license for amnesia either.

This is where “structural dividends” can enter the discussion.

11.3 Structural Dividends: Real Assets First, Shared Benefits Afterward

The word “dividend” has a simple premise: there must first be a real asset or a real, sustainable source of returns.

Without an asset base, there is no dividend.

Without a relationship of rights, wishing cannot declare everyone a shareholder.

The Alaska Permanent Fund offers a clear reference: specified natural-resource revenues enter a long-term fund, from which resident dividends then arise under established laws and eligibility rules. [37]

This is not proof of “civilizational equity.”

It shows only that where shared or public asset relationships genuinely exist, some returns can be preserved institutionally over the long term for the benefit of members.

Norway’s Government Pension Fund Global shows another path: converting part of oil and gas revenues into long-term financial assets, carrying resource wealth beyond the present fiscal cycle to serve current and future generations. [38]

Neither institution gives the world a uniform dividend formula.

They tell us first to clarify the assets, governance, beneficiaries, and risks.

In the future, AI, public computing resources, public data infrastructure, and publicly funded research outputs may also generate new arrangements for public returns in different jurisdictions.

Some may return through taxation.

Some may take the form of public funds, trusts, public services, or direct dividends.

Some may not be suited to the label “dividend” at all.

The specific mechanism must be grounded in real assets and real rights.

This also means structural dividends and basic provision must retain two independent legal rationales.

Basic provision asks whether someone must first prove a contribution to be entitled to secure a basic standard of living.

The answer is no.

Structural dividends ask why a particular shared return exists and who is entitled to it under which rules.

The answer must be established through assets and institutions.

A person does not need to become a “shareholder of civilization” first to possess basic rights.

Equally, society’s commitment to basic rights does not automatically rewrite private assets as common property.

Provision requires public commitment and real capacity to sustain it.

Dividends require assets, returns, and clear rules.

Once the two are separated, many arguments suddenly grow quieter.

We no longer ask, “Does this person deserve to receive it?”

We begin to ask:

Where does this money come from?

Who bears the losses?

Who manages it?

Who oversees it?

Who benefits, under which rules?

If those in power intercept public returns, who can hold them accountable?

Distributive justice returns from judgments of personal worth to institutional relationships.

Shared benefit has another often-overlooked question: risk.

If the fund loses money, who bears the loss?

If public investment fails, who explains?

After resource prices fall, can dividends be promised unchanged?

An institution that says only “everyone has a share” but says nothing about risk, governance, and losses has not yet fully defined the property relationship.

Alaska-style resident dividends and Norway-style long-term savings also remind us that shared returns need not mean only “everyone receives cash now.”

Some shared wealth is suited to distribution in the present.

Some is better converted into healthcare, education, infrastructure, or intergenerational reserves.

Which to choose is itself a public decision.

What truly deserves preserving in structural dividends, then, is not the word “dividend.”

It is a question: when an asset genuinely has public or shared characteristics, how do we prevent its long-term returns from disappearing through institutional gaps?

11.4 A Secure Minimum and the Chance to Rise: Two Questions, Not Two Kinds of People

Someone who needs insulin and someone who wants an ocean-view home are both making requests for resources.

The word “need” is the same, but the weight of the rights behind it may be entirely different.

The first may directly concern life and basic health.

The second is an improvement-oriented, individualized, and clearly scarce developmental choice.

One of the most important tasks of a distribution system is to prevent these two questions from impersonating one another.

The so-called “dual track,” if we retain the term, is first only an analytical framework:

The first layer safeguards basic living conditions and basic public participation, without making performance the admission ticket to survival.

The second handles developmental opportunities that remain scarce, differ greatly, and require qualifications or competition.

This does not divide society into two kinds of people.

The same person may need basic protection today and enter fierce competition tomorrow.

Nor does the first layer mean a globally standardized “welfare package.”

The real costs and institutional forms of basic conditions such as food, housing, energy, healthcare, education, and internet access differ across countries, cities, households, and stages of development.

Civilizational goals can be shared.

The parameters must be real.

In some places, cash is more effective.

Elsewhere, hospital, housing, and care provision is itself the bottleneck. Money in an account still needs real services before it becomes a better life.

How provision combines cash, public services, insurance, housing policy, public infrastructure, and market supply is determined by actual institutions.

The second layer likewise cannot be swept up wholesale by “contribution scores.”

A scholarship can consider academic ability.

A high-risk professional role can require qualifications.

A limited public resource can use markets, queues, lotteries, degree of need, professional judgment, or other rules.

Contribution records can sometimes be relevant facts.

They cannot automatically become resource-use rights, still less general ranks of personal worth or governing authority.

Someone who does not participate in public projects does not thereby lose the right to rest, study, care for family, or withdraw temporarily from competition.

The breathing room civilization gives a person need not be explained as a “system safety valve.”

They may simply be tired.

They may simply want to spend a year with their child.

They may simply not yet know where to go next.

Life itself need not first submit a high-output plan to the system.

Competition can continue.

Honors, incomes, positions, and scarce opportunities can continue to differ.

What competition must not swallow is the ground on which people can still stand and begin again.

Nor can that ground be built by the word “free” alone.

Housing requires land, construction, and maintenance; healthcare requires doctors, medicines, equipment, and queuing rules; care requires time and people; energy and networks likewise need real infrastructure.

When supply is insufficient, increasing purchasing power alone may drive up prices and even lengthen queues.

The core of basic provision has therefore never been a sectarian dispute between cash and services.

The question is: through what real-world path can a person’s minimum needs actually be met?

Cash gives people choice.

Public services can reduce the risk that some critical needs are wholly governed by market prices.

Insurance spreads high, occasional costs.

Housing and land policies address supply.

Communities and families also provide much care that institutions cannot fully replace.

A good system is often a combination, not a slogan.

The second layer of developmental opportunity needs the same sense of proportion.

A popular university program has limited places; a space mission has only a few seats; scarce research equipment cannot be used by everyone at once.

These situations require selection.

Selection rules may be strict, and competition may even be intense.

Provided that failure does not cast a person below the ground of survival and personhood.

This is the most important distinction between the “minimum” and “rising”: the first protects the ability to go on living; the second allows society to keep selecting, competing, and creating differences.

11.5 “If We Guarantee the Minimum, Who Will Still Want to Work?”

This question cannot be answered with ridicule.

In a society of scarcity, not working can indeed shift a burden onto others.

Taxes, care, public services, and production all need someone to bear them.

An institution that entirely ignores incentives, public finances, and supply will ultimately harm those it seeks to protect.

The concern about “supporting idlers” therefore has roots in the experience of scarcity and deserves serious unpacking.

It needs to be separated into several questions:

Will provision reduce some forms of labor supply?

Will it enable more people to change jobs, study, provide care, or start businesses?

Can public finances sustain it over the long term?

Do basic services have real supply?

Will different groups make different choices?

All these questions need evidence.

Finland’s 2017–2018 basic-income experiment paid 2,000 unemployed people €560 a month unconditionally.

The final evaluation found generally small employment effects. Participants reported several more positive outcomes in life satisfaction, psychological stress, confidence in the future, and perceived bureaucratic burden. It leaves a limited piece of evidence, not an answer that can be copied into every country.

At a minimum, it did not support the simplest fear: that once income is detached from traditional unemployment conditions, people will generally stop acting.

GiveDirectly's long-term basic-income research in Kenya, launched in 2017, places the question in an entirely different living environment. [40]

Some participants are scheduled to receive monthly cash payments for many years, while researchers continue to observe changes in consumption, work, business activity, health, and social relationships.

Early research during the pandemic found that cash transfers cushioned some hunger, health, and psychological stresses, though external shocks still affected business and income outcomes.

This too is not an experiment in which "human nature has been proved."

It merely reminds us that security and action are not natural enemies.

Some people will keep working once the minimum is secured.

Some will change jobs.

Some will care for family.

Some will study.

Some will rest.

Some will make choices we dislike.

Real people never become one kind of person because of an institution.

A good system of provision therefore cannot rest either on the assumption that "only fear drives people" or on the romantic assumption that "once free, everyone will automatically pursue lofty meaning."

It needs only to acknowledge that people differ.

Then preserve basic rights, real incentives, fiscal constraints, and opportunities to reenter together.

Caring for children, accompanying older people, creating, learning, community service, research, and sustaining traditions may all have real value without immediately appearing on a payroll.

These actions can be seen and acknowledged.

There is no need to turn them all into points.

Contribution records can remember some real effort.

They cannot become round-the-clock surveillance to prove "I am not being idle."

If a civilization can sustain action only through continuous fear, the costs will eventually enter bodies, relationships, and social trust too.

Provision reduces the fear of falling.

It does not automatically manufacture meaning.

Work itself is also far more than a product of fear.

Many people enjoy the mastery their profession brings, enjoy their colleagues, enjoy the freedom of higher income, and feel fulfilled when others genuinely need what they make.

Safeguarding the minimum does not require disparaging labor.

It merely loosens labor's position as the only path under the threat "otherwise you cannot survive."

Finland's experiment especially deserves to be understood this way.

Its two thousand participants still lived in a real labor market; they did not suddenly enter a post-scarcity world. The experiment changed some income conditions and bureaucratic relationships, rather than rewriting work, housing, families, and social identity altogether.

What it gives us is therefore not a final answer, but a small piece of real-world evidence: easing some conditions did not make action disappear generally.

The Kenyan research has a similar significance.

A cash buffer can change hunger, business activity, risk-taking, and household decisions. Hospitals, roads, stable markets, and external crises still depend on other real-world conditions.

Distribution policy must therefore also discuss supply, prices, public services, care, and macro-level capacity to sustain it.

The real test is whether these conditions can keep pace together.

If a city has a severe housing shortage, even the finest cash scheme may have a large share absorbed by rent.

If a remote area has no doctor, a larger account balance does not automatically become medical care.

Distributive justice must face both people and material resources.

Nor will it automatically eliminate effort.

11.6 After Extensive Automation, Five Things Must Truly Be Separated

If AI and automation continue to undertake more productive and cognitive tasks, wages may find it increasingly difficult to serve alone as the “admission ticket to survival.”

This is not because humans already have no work.

It is because an institutional relationship is becoming more fragile: only labor purchased by the market qualifies for exchange into basic living conditions.

Extensive automation truly forces civilization to separate five things anew.

First, the right to live.

It cannot fall to zero along with the market price of a skill.

Second, returns on assets.

Who owns machines, funds, land, data rights, or company shares must still be grounded in real property rights, contracts, tax systems, and public investment.

Third, opportunities to develop.

Education, professional roles, scarce projects, and resource use may carry conditions of capability, responsibility, and procedure.

Fourth, recognition of contributions.

Care, creation, knowledge, organization, risk-bearing, and companionship can be remembered, but cannot automatically be exchanged for basic rights or public power.

Fifth, the meaning of life.

Institutions cannot distribute it.

Receiving is the world reaching me.

Meaning is me reaching the world.

A system can give people income, access to education, and more time.

It cannot decide for a person why that time is worthwhile.

This distinction is extremely important.

The error earlier versions most easily made was to declare, once wealth was no longer the only scarcity, that the next task was to “distribute meaning.”

Once meaning can be distributed, someone must hold the power to distribute it.

Scores, eligibility, templates, and “more meaningful people” will follow.

Volume Five has closed off that path completely.

What distribution systems can do is reduce how often people are forced to accept anything in order to survive.

Make real needs easier to see.

Make different entrances into life genuinely available.

Let contributions find an echo.

Then step back.

Meaning must remain on the side of life.

Extensive automation will also create a very practical transition problem: uneven speed.

Some industries may rapidly reduce jobs; some jobs may only change in task composition; demand for some occupations may even rise because of new technologies.

The same person may experience several rounds of change: today’s skills lose value, tomorrow they learn new tools and reenter, and a few years later they face another restructuring.

If basic living conditions are entirely bound to one career choice, individuals will experience technological change as a succession of trials of their right to survive.

More robust institutions should separate career transitions from a collapse in survival.

Losing a job may mean a change in income, retraining, and finding a new place.

It should not automatically mean interrupted healthcare, children leaving school, and immediate loss of housing.

This also shows that “being needed” and “having a job” have never been permanently identical.

A person can be needed at work, and also in a family, community, research, art, and relationships.

Institutions need not apply a uniform value label to these lives.

They need only avoid allowing one market change to wipe out every other relationship at once.

The stronger automation becomes, the more important this separation is.

Because the faster machine capabilities grow, the less we can let people’s basic standing rise and fall with the market price of a particular job.

11.7 The Real World Has Given Us No Uniform Answer Sheet

New distribution mechanisms must prove themselves in the real world.

Many different institutions already combine public assets, long-term savings, cash transfers, social protection, and market incentives in different ways.

The Alaska Permanent Fund shows that specified resource returns can enter a long-term fund under explicit legal foundations and generate resident dividends.

Norway’s Government Pension Fund Global shows that resource returns can also be converted into long-term public financial assets serving current and future generations.

Long-term research on local commons and cooperative governance shows that shared resources have more than two possible destinies: “entirely private” or “centrally administered.”

These institutions differ from one another.

They cannot be pieced together into a global “AI dividend template.”

What we can genuinely take away is a set of questions:

Where do the assets come from?

Who holds which rights?

Has public investment generated public returns that can be incorporated into institutional arrangements?

Who manages, who oversees, and who bears losses?

Do returns take the form of cash, services, long-term savings, or something else?

How can the rules be changed?

Mechanisms can be learned from.

Sources of rights cannot be copied by analogy.

This is also why slogans such as “data belong to all humanity,” “algorithms belong to civilization,” and “computing power should pay everyone a dividend” are insufficient.

Some data involve individual rights.

Some arise from contracts.

Some models are developed with private capital.

Some basic research is publicly funded.

Some infrastructure already has a public character.

The real world does not automatically become one kind of asset because we like a slogan.

Justice truly begins with the willingness to clarify these differences one by one.

Real-world cases also tell us that how an institution is governed matters as much as its distributive outcomes.

Without transparent rules, independent auditing, political constraints, and long-term responsibility, a public fund’s so-called “shared assets” can easily become a new resource of power.

Research on local governance of shared resources also reminds us that participants’ genuine possession of mechanisms for rule-making, oversight, and sanctions matters more than saying “this belongs to everyone.”

Structural dividends must therefore answer together how much is distributed, who manages it, who bears risk, and how the rules are reauthorized.

We must also ask whether the fund can be diverted, how managers are replaced, how losses are disclosed, who may amend the rules, and whether those who come later have an opportunity to authorize them anew.

Shared wealth is most easily romanticized when we imagine only sharing, not governance.

Yet governance is precisely what determines whether that wealth can truly belong to the public.

11.8 Transition: Do Not Overturn Reality All at Once in the Name of a Better Ideal

Distribution systems are tied to wages, taxes, housing, healthcare, families, businesses, public finances, and people’s expectations.

They can hardly be transformed by a single “switch.”

Any sudden declaration that old mechanisms are invalid and new ones will take over completely passes the greatest risks to the most ordinary people.

Transition must therefore begin with care.

Step one: do not touch the minimum.

Whether or not structural dividends, contribution records, or new resource mechanisms are being tested, basic rights must not first become experimental stakes.

Step two: verify on a small scale.

Only when real actors, jurisdictions, funds, and resources are involved can we know whether a mechanism will produce new humiliation, dependence, arbitrage, exclusion, or entrenched power.

Step three: preserve exit.

If an experiment fails, affected people must have paths to recovery, appeal, rollback, and alternatives.

Step four: do not let contributions steal rights.

Even where contribution records are useful, they must not automatically generate eligibility to survive, ranks of personal worth, general entitlements to deploy resources, or public governing authority.

Step five: make expansion depend on results.

Success in one city, institution, or fund does not mean the whole world should copy it.

First establish the core causal relationships, rights protection, funding capacity, and exit and repair mechanisms. Then discuss lawful expansion.

Mutual recognition across actors should come later still.

If multiple countries and organizations develop effective practices of their own in the future, they can then discuss mutual recognition of data, funds, contribution facts, or public returns.

“Global consistency” cannot justify first creating a unified registry with no source of authorization.

The startup and stable phases also need different rhythms.

The startup phase first addresses real survival pressures, institutional breakdowns, and a minimum verifiable closed loop.

The stable phase must strengthen enduring rights, procedures, fairness, long-term public finances, and protection against entrenched power.

The ideal completeness of the stable phase must not strangle the beginning.

Nor may the urgency of the startup phase be used to keep temporary powers forever.

The real difficulty of distribution reform has never been drawing an attractive diagram.

The startup phase must especially avoid two opposite errors.

One is to say that all boundaries of rights can be dealt with later, for the sake of speed.

The other is to demand that a small-scale experiment possess every complex institution of a stable phase before it begins.

A more practical approach is for constraints to grow alongside risks.

Where scale is small, actions reversible, and impacts limited, lighter procedures can enable faster learning.

Once coverage expands, dependencies form, funds increase, and data reach deeper into daily life, auditing, appeals, statutory authorization, and external oversight must strengthen in step.

Nor should an experiment test only “whether people are satisfied.”

It must examine whether basic living conditions improve, new exclusion and shame arise, public finances and supply can sustain the system, errors can be corrected, and people can recover their original rights after exiting.

If a structural-dividend mechanism is established, it must separately examine whether asset returns are real, how high governance costs are, and how volatility is borne.

If contribution records enter certain developmental opportunities, additional checks must establish whether they have quietly slid into rankings of personal worth and social-credit scoring.

The earlier these questions become visible, the less need there will be for a major reform later to repair a major mistake.

It is ensuring that a person living in the old world does not have to bear all the transition costs of that diagram.

11.9 Chapter Conclusion: Assistance May Examine Conditions, but Must Not Judge Personal Worth

Return to the counter at the beginning of this chapter.

That person still needs to submit some documents.

Because institutions must know the conditions under which support applies, and must prevent fraud and error.

But he should not have to hand over his dignity too.

This is what Chapter 11 truly seeks to change.

Assistance may examine conditions, but must not judge personal worth.

Dividends may calculate returns, but must not calculate the value of a person.

Competition may decide who receives certain scarce opportunities first. It cannot decide who deserves food, medical treatment, or a place to remain.

Contributions can be seen.

They cannot become the birth certificate of basic rights.

Private property rights can be protected.

That does not permit pretending that public investment, shared risks, and civilization's accumulation never existed.

Structural dividends hold only where real shared assets and relationships of returns exist.

Basic provision, by contrast, does not first require proof that everyone owns an abstract share in civilization.

Extensive automation makes these relationships more urgent, but does not automatically answer them for us.

Machines being able to do more does not mean material resources become infinite.

Nor does it mean humans cease to labor.

What must truly be abandoned are two short circuits:

No market value, therefore no right to survive.

Greater contribution, therefore inherently greater power.

Civilization can still have markets.

It can have differences in wealth.

It can have competition, honors, professional thresholds, and scarce opportunities.

It simply needs to draw one line more clearly:

What may be decided by competition.

And what may not.

Continued existence is civilization's minimum.

Freedom from anxiety over survival is its first goal.

Ultimately, genuine distributive justice does not mean everyone receives the same amount.

It means that when a person temporarily loses in the market, loses a career, or loses a stretch of life, civilization still has not translated "you lost" into "you do not deserve to keep living."

A civilization becoming truly rich is not simply its warehouses filling up.

Growing wealth certainly matters.

Greater material abundance and higher productivity enlarge the range of institutional choices.

But what truly determines whether a society turns affluence into civilization is how it treats those who temporarily have no place in the market.

A child does not lose a layer of personhood because their parents are unemployed.

A patient does not become less worth treating because they cannot contribute output.

An older person leaving a job is not automatically deregistered from the public world.

When these judgments no longer need repeated defense, distribution truly moves from charity into civilization's everyday life.

It means fewer and fewer people need to kneel to prove they deserve to stay alive.

Chapter 12 · Education’s Change of Direction: When AI Begins Answering for Us

A child sends AI a photograph of a problem.

Two seconds later, the answer appears.

Every step is there, the explanation is clear, even more patient than the textbook.

The teacher does not ask whether he got it right.

The teacher asks only:

“Why can you do that in the second step?”

The child looks at the screen.

Then at the teacher.

He knows the answer.

But he does not know why he believes it.

This is where education in the AI age truly begins to change.

The problem is not that the child used AI.

Nor that the machine answered too quickly.

The problem is that the answer arrived too early, before judgment had time to grow.

As answers grow cheaper, education cannot stake all its dignity on the speed of answering.

What it must truly safeguard is something else:

whether a person can know why they believe, when they should disagree, and how to judge anew after making a mistake.

12.1 As Answers Grow Cheaper, Why Must Learning Still Be Slow?

School has never been merely an arena for answering questions.

It certainly teaches knowledge, but also develops language, calculation, professional foundations, cooperation, discipline, cultural understanding, and the capacity to enter society.

Examinations and qualifications have also opened paths beyond inherited circumstances for many ordinary people.

There is no need to disparage that history to prove AI’s importance.

What genuinely needs reexamination is an increasingly fragile shortcut:

Can answer, therefore understands.

High scores, therefore good judgment.

Has a diploma, therefore competence cannot become outdated.

In the past, these proxies sometimes worked well enough.

The relationship between education and occupational returns is changing, but “returns to education have fallen to zero” is not generally the case today. [41]

Now, machines can directly generate answers to more and more knowledge tasks, making the gap between proxies and real ability easier to expose.

An impressive essay can be completed in minutes.

Code can be made to run first, while a person slowly comes to understand it afterward.

The answer to a problem has appeared before the student has even truly formed the question.

Education thus faces a deeply counterintuitive task:

In some places, it must deliberately slow down.

Not to make children suffer,

but because understanding inherently takes time.

The first time one gets stuck alone.

The first time one discovers a premise was wrong.

The first time one overturns an apparently correct answer.

The first time one finds the next step without a ready-made prompt.

These processes are less dazzling than an “answer.”

Yet they are precisely where a person’s later ability to judge answers comes from.

Difficulty is not a virtue in itself.

Mechanical repetition that can be spared may be spared; low-value labor that tools can undertake may be reduced.

But if every moment of being stuck is treated as lost efficiency, education will ultimately optimize away the process by which ability forms.

Children will receive more correct answers.

Yet they will know less and less how they themselves recognize correctness.

What education truly needs to distinguish is:

which difficulties are merely waste,

and which are growing ability.

This is also one reason examinations can continue to exist.

Independent answers produced with limited time and tools can help teachers see what students genuinely understand at that moment.

The problem is only that one examination cannot carry a child’s entire worth.

A score can describe a period of performance.

It should not finish writing the boundaries of a person’s abilities in advance.

Likewise, project-based learning is not inherently superior.

An “innovation project” with confused roles and a few people doing everything may involve less learning than a carefully designed examination paper.

Form is not the answer.

What matters is which abilities this learning activity has actually brought into being.

12.2 Knowledge Is Not Obsolete: Without Foundations, Judgment Becomes Mere Confidence

“If AI can look everything up, why do we still have to learn this?”

Today, that is an entirely reasonable question.

The answer is not “because it will still be on the exam.”

Nor is it “AI can be wrong, so memorize everything.”

The more fundamental reason is:

Someone with no foundation of knowledge finds it difficult to know when to doubt a fluent-looking answer.

If you do not know what happened in history, any period can be written to sound real.

If you lack basic mathematical intuition, even a number wrong by a factor of ten can look professional.

If you do not understand a discipline's most basic concepts, two contradictory explanations from a model may sound equally agreeable.

Knowledge is not for competing with AI at memory.

Knowledge builds a map in the mind on which new information can be placed and judged.

The map need not depict every stone.

But someone without even a sense of direction can easily take every turn suggested by navigation software as truth.

Foundational knowledge therefore still matters.

Reading, writing, mathematics, science, history, art, bodily experience, and real life do not lose value because search gets faster.

What changes is the purpose of learning them.

A little less "remember it so you can repeat it exactly."

A little more "make it part of how I judge the world."

Genuine knowledge remains after the tool is switched off.

One very simple test is:

If AI is temporarily taken away, can I still explain what the problem is?

Can I still explain why I chose this answer?

Can I still spot an obviously unreasonable result?

Can I still move forward without knowing the standard answer?

If I cannot do any of these, however skillfully I use the tool, it has not yet become my own ability.

AI can be scaffolding.

Scaffolding has another property: its height must be appropriate.

Too low, and it is no help.

Too high, and the student never even gets to encounter the real problem.

A child just learning fractions and a graduate student should not have the same boundaries when using the same kind of AI assistance.

Different ages, subjects, learning objectives, and risks call for different intensities of support.

Education need not seek a globally uniform iron line declaring "AI may be used at this age, but not that one."

It needs to make assistance proportionate to the stage of development.

The purpose of scaffolding is to let the building rise.

Not to make the building depend on scaffolding forever.

12.3 You Can Borrow an Answer, but Must Not Lend Out Your Judgment

Knowing how to use AI does not mean knowing how to judge.

Rejecting AI does not mean being independent either.

What education truly needs to develop is the stretch between the two.

Students can ask AI to explain a concept.

They can ask it to present opposing views.

They can ask for help finding flaws, fixing code, practicing speaking, simulating interviews, or translating materials.

The question is not “was it used?”

The question is what still belongs to the student after using it.

When a student submits an AI-assisted report, what truly deserves asking is not “what percentage was AI?”

It is:

Can you explain the central judgment?

Do you know which information came from where?

If the model gives you the opposite answer, on what grounds do you choose this one?

If key evidence proves wrong, which part would you change?

If these questions cannot be answered, even a beautiful work may be nothing more than a tool’s capability passing once through the student’s account.

Future assessment therefore need not choose only between “AI-banned examinations” and “unrestricted AI access.”

There can be checks of foundational ability without AI.

There can also be complex projects where AI is permitted.

Students can be asked for oral explanations, process records, version comparisons, and rebuttals of their own answers.

And after a project ends, one can suddenly ask:

“Without AI now, can you tell me the three most important things?”

This is not catching cheating.

It is checking where learning actually happened.

Nor should AI-use records casually grow into round-the-clock student profiles.

Process evidence retained for teaching should have an explicit purpose, scope, and retention boundary.

A child is not a data stream waiting to be optimized.

The most seductive promise of personalized education is that “the system can finally understand this child completely.”

But what a child likes today, fears today, or performs poorly at today is not their permanent identity.

Reading difficulties at ten should not follow someone like a brand at twenty.

Nor should one impulse, one act of cheating, or one lapse in attention be expanded by a model into a “personality risk.”

Growth needs memory.

It also needs forgetting.

How long educational data should actually be retained, who may access it, and whether it is used in admissions or employment must be considered more cautiously than “how long technology can store it.”

The better an educational system understands a student, the more it must remember:

Understanding is not ownership.

Nor is prediction deciding on their behalf.

12.4 A Child Does Not Need to Be Raised into “Someone Civilization Needs”

Another temptation easily arises in the AI age:

As machines become better at solving problems, should education select a group better at “deciding the direction”?

This path looks grand.

It is also dangerous.

Because once education begins answering “who is better suited to decide for civilization,” schools are no longer merely developing abilities.

They are allocating power in advance.

This book no longer divides students into “problem-solvers,” “converters,” “night watchkeepers,” or any new types of civilizational personality.

Some people are willing to take public responsibility.

Some enjoy research.

Some care for family.

Some create.

Some hope for a quieter life.

Some have a strong sense of direction at twenty and suddenly want to stop at thirty.

All of these can be complete lives.

Education should enlarge a person’s abilities and choices.

Not assign them in advance to some function of civilization.

Rest, low competition, changes of direction, failure, and reentry should therefore all have a place.

A student temporarily not wishing to become the best does not mean education has failed.

A young person changing subjects, delaying employment, or caring for family need not be explained as a “buffer node in the system.”

Human lives do not exist to keep systems healthy.

Schools are especially prone to treating “having direction” as a virtue.

Fifteen-year-olds are asked to write life plans, eighteen-year-olds to choose majors, and twenty-two-year-olds to say who they want to be in five years.

Some know early.

Some must take many winding paths before they know.

Some never need a grand mission at all.

If education makes “clear direction” itself a mark of the outstanding student, exploration becomes a form of failure again.

Permission to change one’s mind is itself part of education’s openness to the future.

If education has one genuinely lasting public responsibility, it should be:

to give children the ability to choose,

and the ability to bear the consequences of their choices.

Not to choose a worthy life for them.

12.5 Learning to Ask: Not Because AI Cannot Ask, but Because Questions Change the World

AI can already generate questions.

So “asking questions is humanity’s last moat” does not hold.

What truly matters is something else:

Who has made what into a problem?

A company asking “How can we increase time spent?” and a parent asking “Why is my child ever less willing to leave the screen?” may be confronting the same system.

Different questions make the world visible in different directions.

Education therefore needs to teach not “ask a few more questions,”

but how to form a question.

First look at the premises.

Have we taken for granted something that could actually change?

Then look at those absent.

Who has been left out of this problem?

Then look at the consequences.

If a proposal succeeds for ten years, what else will it change along the way?

Only then ask about efficiency.

Even if it can be done, is it worth doing?

Einstein’s thought experiments, Turing’s reframing of “Can machines think?”, and Silent Spring’s transformation of invisible ecological risks into a question the public could feel all remind us:

Truly important questioning does not merely seek an unknown answer.

It redefines what deserves to be seen.

But questioning must not be made sacred either.

A bad question can lead an entire society in the wrong direction.

Nor can a grand question bypass facts, rights, and professional knowledge merely because it moves us.

Good questions must ultimately face evidence.

Mature questioning is not afraid of having its own answers overturned.

One of the most valuable classroom moments may be when the teacher says, “I don’t know.”

Not to throw the question back into a search box,

but to ask with the students: how could we find out?

What evidence do we need?

Who has already studied it?

How do we distinguish facts, explanations, and value judgments?

AI can quickly put existing answers on the table.

Education must help students know what is still missing from the table.

12.6 Learning to Live with Others: Some Abilities Do Not Become Less Important as AI Grows Stronger

The hardest thing in a class is often not that no one can solve the problem.

It is that two people both believe they are entirely right.

A project's real failure is also often not that no one was intelligent, but that professional languages were mutually unintelligible, conflicts of interest went unspoken, and everyone first protected themselves when errors appeared.

AI can translate languages, summarize positions, and simulate negotiations.

All of this is useful.

But responsibility within relationships does not disappear as a result.

A model cannot decide on a person's behalf whether they will acknowledge having hurt someone.

Nor can a prompt decide whether a community, after conflict, is willing to return to the same table.

Education therefore still needs to give people real experiences of cooperation.

Not surface cooperation of the "everyone splits the work to make a slide deck" kind,

but knowing how to keep working after encountering disagreement, misunderstanding, failure, and unequal standing among opinions.

Several very ordinary things can be practiced:

1. State the Other Person's Position Accurately

Before disagreeing, try to state fully what the other person is actually concerned about.

Not for politeness,

but to prevent us from arguing only with enemies of our own imagination.

2. Separate Conflict from Personal Worth

A person's viewpoint can be refuted, and a mistake can carry accountability.

Neither means declaring the whole person impossible to cooperate with.

3. Repair after Error

Apology is not everything.

Real repair also includes facts, responsibility, compensation, changes to rules, and rebuilding trust.

4. Allow Relationships to Have No Neat Resolution

Not every conflict will end in reconciliation.

Education must teach that even when we cannot like one another, we can still live together under rights and rules.

These abilities are difficult to measure completely in one exam.

But they can slowly develop through projects, teams, community practice, art, sport, discussion, and real relationships.

They are not a "new trivium."

Nor do they need to replace language, mathematics, science, and history.

They simply remind us:

A person will go on to live among other people.

And cooperation is not always warm.

Sometimes real cooperation means clarifying boundaries, tasks, and responsibilities even when people dislike one another and do not share values.

Engineering teams, hospitals, courts, communities, and families all need this ability.

“Understanding the other person” does not mean abolishing conflict.

Nor is “reaching consensus” a necessary endpoint of every cooperation.

Sometimes the most mature outcome is merely: we still disagree, but know how to keep disagreement from turning into harm.

That itself deserves to be taught.

12.7 Sometimes Education’s Value Is Not an Answer, but Someone Who Did Not Walk Away

A student says, “I don’t know why I should keep trying.”

The teacher can of course offer advice.

They can also help find resources, adjust tasks, and contact parents or professional support.

But sometimes the first task is not to explain life immediately.

It is to hear the person out.

Solving problems quickly is a posture AI is very good at.

Human education must not therefore mistake “not solving it immediately” for incompetence.

Loss, anxiety, loneliness, conflict, failure, and prolonged uncertainty may all enter a student’s life.

Some require professional psychological or medical support.

Education cannot romanticize illness as “material for growth.”

Nor can it turn every ordinary pain into a diagnostic label as soon as it appears.

What is truly precious is an intermediate space:

Someone can first be heard.

Someone is willing to stay and help them put things into words.

When professional help is needed, they can reach genuine support in time.

One of the most moving aspects of hospice care is likewise this capacity: “I cannot solve this for you, but I will not leave first.”

It does not belong to some mysterious personality type.

Listening, boundaries, accompaniment, and referral can all be learned.

Education must also protect boundaries if it is to develop these abilities.

Students are not their classmates’ therapists.

Teachers are not all-purpose psychologists either.

Empathy does not mean assuming unlimited responsibility.

Mature accompaniment knows when to stay.

Schools also need to leave teachers time for this.

If every minute must be filled with curriculum schedules, platform tasks, and quantitative metrics, the first things to disappear are often conversations that cannot be put on a timetable in advance.

A child will not lose a loved one at 10:20 a.m. according to the system’s schedule.

Nor will they be ready, right on time in a prescribed “social-emotional learning module,” to speak about their fears.

Some important things in education need a little room left for the unplanned.

And it knows when a problem should be entrusted to someone better suited to help.

12.8 Teachers Do Not Lose Their Place Because AI Can Teach a Lesson

If AI can explain a problem twenty-four hours a day, what remains for teachers?

First, teachers still need to know their subject, of course.

Someone who does not understand a discipline will struggle to judge where AI has taken a student.

But a teacher's value has never been merely "knowing the answer before the student."

Teachers see a person who is growing.

They know whether the child truly does not understand or simply does not dare to say so.

They know when to offer a hint and when to let the child stay stuck for five more minutes.

They know whether a criticism will bring a student greater clarity or merely make them stop speaking from then on.

These judgments are not mysterious.

They arise from professional knowledge, long observation, relationships, and responsibility.

AI can assist lesson preparation, offer alternative explanations, generate exercises, and help provide feedback.

Teachers can thereby do less mechanical repetition.

But education must not therefore reduce teachers to "AI classroom administrators."

Equally, AI must not be packaged as "your tireless, endlessly patient private tutor who understands you best," while errors, dependence, privacy, commercial interests, and children's rights are ignored.

The genuinely reasonable relationship is more like a triangle:

Students still learn.

Teachers still bear professional educational responsibility.

AI provides capabilities within clear boundaries.

None of the three should quietly disappear.

If system recommendations conflict with a teacher's judgment, there should be a way to question them.

If AI participates in assessment over the long term, students should know what role it plays.

If personalization depends on extensive learning data, the boundaries of data use, retention, and withdrawal must be clear.

The closer AI comes to a child, the more important transparency and revocability become.

The true sign of mature educational technology is not that children become increasingly unable to do without it.

Another responsibility of teachers is to protect students from being repeatedly made experimental grounds for technological trends.

New tools can be tried.

But trials must know what they seek to improve, and observe what has not improved.

Higher short-term satisfaction does not mean deeper learning.

Faster homework completion does not mean deeper understanding either.

If students are utterly unable to continue once AI is removed, the supposed "efficiency gain" still carries at least one unsettled learning debt.

Educational technology should ultimately face a simple question in return:

Has it made the child more capable, or merely able to accomplish more while it is beside them?

It is helping children develop abilities that will one day let them depend on it less.

12.9 The Purpose of Assessment Is Not to Find “People More Entitled to Decide”

The AI age will make assessment harder.

Work can be assisted.

Processes can be generated automatically.

Expression can be polished.

Even reflective writing can be delegated to a model.

The most dangerous response is therefore to turn schools into larger surveillance rooms.

Continuous filming, ongoing behavioral analysis, emotion scoring, and personality prediction may all appear to improve “authenticity.”

The cost may be turning growing children into permanent files.

What assessment truly needs to know is:

What can this student do now?

What can they not yet do?

How did they learn it?

With what assistance can they do it?

What remains when that assistance is removed?

These questions are enough.

Assessment does not need to go on and answer:

“What kind of person are they essentially?”

“Are they a future leader?”

“Are they altruistic?”

“Do they deserve more power?”

Educational assessment must not become a verdict on personal worth.

The same applies to professional qualifications.

Doctors, pilots, engineers, auditors, and other high-consequence roles may require standards of knowledge, skill, training, and fulfillment of duties.

Some roles may also examine directly relevant conflicts of interest, histories of responsibility, and present suitability.

But passing an examination does not generate public sovereignty.

An outstanding record does not generate governing authority either.

A lifetime of good deeds certainly does not create eligibility for exemption from oversight.

Sources of power and assessment of ability must remain separate.

This line is especially important in the AI age.

Because as schools acquire richer data and predictive abilities, they will find it increasingly easy to believe they can see a child’s future in advance.

What education truly needs to do is precisely to leave a little unknown in the future.

A fifteen-year-old should not be judged in advance by a model as the person they will be at thirty-five.

Growth deserves protection precisely because people can change.

This also means assessment in the AI age must be more willing to acknowledge “not yet able.”

Inability is not a defect of character.

It is simply information about where the next step of teaching should begin.

If assessment is responsible only for ranking, students will naturally hide what they cannot do.

If it is also responsible for helping learning, admitting inability instead becomes a doorway to progress.

Truly good feedback does not tell children their rank faster.

It helps them know where they can go next.

12.10 Entrusting Risk: The Greater the Capability, the Less “Trust This Person” Alone Can Suffice

Some roles do indeed command high-consequence capabilities.

Managing critical infrastructure, approving major public projects, operating high-risk technologies, and bearing professional fiduciary duties cannot rest on “trust them” alone.

Education and professional training can prepare people’s abilities for these roles.

Real authorization still comes from actual institutions.

Important roles must therefore separate at least several things:

Who is eligible to apply.

Who grants authorization under law or rules.

How far the authority extends.

Who can review and stop it.

Who bears which responsibility when consequences arise.

Which responsibilities continue after departure, and which permissions must end.

A person’s good past fulfillment of duties can be relevant evidence.

It is not permanent credit.

One mistake may be very serious.

It still need not automatically become a lifelong verdict on personal worth.

What is genuinely needed is role-relevant, verifiable, appealable suitability with temporal boundaries.

Not “loyalty to civilization.”

Still less contribution points, political loyalty, or an overall character score.

Recommendations, mentors, and collaboration networks must not become algorithmic guilt by association either.

Responsibility should extend to who did what, what they knew at the time, and what explicit duties they bore.

Education’s responsibility here is limited but important:

before gaining power, those who will bear high-consequence duties should already be accustomed to asking:

On what grounds do I have the authority to do this?

Who will bear my mistakes?

If my judgment is wrong, who can stop me?

These questions are far more reliable than “cultivating perfect people.”

Professional education should especially practice retaining these questions under pressure.

In ordinary times, everyone can say “safety first.”

The real difficulty is whether, when a project is late, the budget overrun, superiors pressing, and colleagues all deeming the risk acceptable, someone can still point to the consequence they are unwilling to bear.

This does not ask students to become heroes.

It is precisely to ensure institutions need not depend on heroes.

Clear authority, peer review, avoidance of conflicts of interest, stopping interfaces, and responsibility records should be learned alongside professional skills.

12.11 When Machines Can Also Create, Why Should Humans Still Create?

If AI can also write poetry, compose music, draw, and program, why should people still do these things?

Because creation has never been merely a transaction in scarce skills.

A child making a clumsy drawing does not first need to prove they are better than a generative model.

An adult writing a poem no one publishes does not need to prove it fills a market gap either.

Part of creation's value occurs within the creator.

What they noticed.

What they chose.

How they finally turned a feeling they could not articulate into a form others could see.

AI's participation does not automatically cancel these processes.

It can even become new material, a new instrument, or a new collaborative tool.

What truly calls for vigilance is when tools directly decide for us:

which style is more popular,

which expression more easily attracts traffic,

which lives are more worth telling.

Once creation becomes nothing but adaptation to recommendation systems, whether machines possess subjecthood is no longer the first question.

The space for human expression has already been narrowed by the objective function.

Art, writing, handicraft, experiments, and projects in education therefore do not need "AI cannot do this" to prove their value.

Their value lies in:

giving a person practice in seeing, choosing, forming, and taking responsibility for an outcome of their own.

The result may be imperfect.

AI assistance is also permissible.

As long as someone can still say at the end:

"This is the version I want to leave behind."

That sentence can also form a boundary for creative education in the AI age.

It does not require work to be made one hundred percent by human hands.

It requires the author to be able to own their choices.

Why keep this sentence?

Why delete that passage?

Why accept the model's suggestion here, yet reject it elsewhere?

Creative agency does not depend on pure handcraft.

It depends on a person still being able to say of what ultimately takes shape: this is my judgment, and my responsibility too.

12.12 Chapter Conclusion: As Answers Grow Cheaper, Judgment Still Belongs to Him

Return to the child at the beginning.

The teacher asks him:

“Why can you do that in the second step?”

What education truly hopes to wait for is not his recitation of the explanation AI just gave.

It is the day he can stop and think for himself.

Perhaps he will be right.

Perhaps he will be wrong.

Perhaps he will discover that AI was the one that was wrong.

What truly matters is that the judgment begins to arise from him.

Education therefore need not declare the age of knowledge over.

Knowledge remains the foundation of judgment.

Nor need it announce the departure of teachers.

Relationships, professional judgment, and responsibility for growth still need people.

Still less is there any need to select a group of “new humans” to decide civilization’s direction.

Education faces every child.

Help him read, calculate, and understand.

And also doubt, create, and live with others.

Know when to draw on machines.

Know when he must step forward himself.

Allow him to rest, change, fail, and reenter.

And let him know, when taking responsibility, that power always needs a source.

At this point, what Volume Two has genuinely completed is still only the separation of relationships:

Facts cannot pass themselves off as values.

Contributions cannot pass themselves off as power.

Expertise cannot pass itself off as authorization.

Capability cannot pass itself off as legitimacy.

AI providing answers cannot pass itself off as a person’s judgment either.

Only next, in Volume Three, do we enter institutions:

How can these separated relationships be recombined into real structures that can operate, be verified, and also be exited?

Education cannot finish writing the answers for the future.

What it can truly do is give those who come later the ability to change the answers themselves.

As answers grow cheaper, judgment still belongs to him.

Perhaps this is the simplest test of this chapter.

Many years after leaving school, a child has long forgotten countless formulas, dates, and standard answers.

Yet when a smarter system hands him a conclusion,

he still knows to ask first:

“Why?”

And has the ability to keep pursuing the question.

Material Notes for This Volume

On the Historical Scenes, Cases, Future Extrapolations, and Literary Metaphors in This Book

The following notes collect the material referenced by light superscript markers in the main text. They identify the material, limits of sources, status of thought experiments, and literary metaphors; they do not carry the main argument. Unless delaying an explanation would materially mislead a general reader about facts, subjects, rights, institutional force, or present versus future status, qualifications of this kind are generally no longer inserted into scenes, strong propositions, sustained parallel passages, or closing resonances.

19 | Volume Two, Chapter 7 | The “thief of time”

The “thief of time” is a literary metaphor for intertemporal externalization in which present actors realize gains early and leave risks or costs to successors. It is not a formal legal or economic term.

20 | Volume Two, Chapter 7 | The signal-to-noise analogy

The book borrows “signal-to-noise ratio” to illuminate relationships among information, evidence, and indicators. It does not calculate social value directly as a physical signal or establish a uniform social-value function.

21 | Volume Two, Chapter 8 | The 2018 Uber autonomous-driving test crash in Tempe

NTSB report HAR-19/03 records that the system first detected the pedestrian 5.6 seconds before impact but did not correctly predict the path. About 1.2 seconds before impact it determined a collision was imminent. Under the design then in use, collision-mitigation emergency braking was not executed in that situation; the system relied on the vehicle operator to intervene. The NTSB identified operator distraction as the probable cause, and contributing factors including deficiencies in Uber ATG’s safety risk management, operator oversight, and management of automation dependence, along with pedestrian behavior and insufficient Arizona oversight of testing. The book uses this to discuss distributed responsibility, not assign all responsibility to one actor.

22 | Volume Two, Chapter 8 | The 2023 Cruise pedestrian crash in San Francisco and disclosure

NHTSA’s 2024 consent order records that on October 2, 2023, a human-driven vehicle first struck a pedestrian into a driverless Cruise vehicle’s path. After the Cruise vehicle collided and initially stopped, it moved again, dragging the pedestrian roughly twenty feet. NHTSA found that Cruise’s one-day and ten-day reports failed fully to disclose the postcollision dragging. California’s DMV suspended Cruise’s driverless deployment and testing permits on October 24, 2023. The main text uses the case only to show that postincident disclosure belongs to the chain of responsibility.

23 | Volume Two, Chapter 8 | The Boeing 737 MAX and MCAS

The FAA’s summary of its 737 MAX return-to-service review confirms that the original MCAS design could rely on one angle-of-attack sensor. With a persistently erroneous high-angle signal, MCAS could reactivate after pilots used electric trim; repeated commands could produce substantial stabilizer mistrim. Subsequent changes included using two angle-of-attack inputs and limiting repeated activation and MCAS authority. The book uses these confirmed system problems to discuss layered responsibility in high-consequence automation, not attribute the two crashes to a single software factor.

24 | Volume Two, Chapter 8 | The banality of evil and an analogy of technical responsibility

Since Hannah Arendt proposed the “banality of evil” in Eichmann in Jerusalem, Eichmann’s motives, ideology, and Arendt’s interpretation have remained subjects of scholarly debate. This book does not morally equate technical accidents with the Holocaust. It borrows only a limited question of responsibility: Can roles,

procedures, and organizational divisions of labor weaken individuals’ questioning of the consequences in which they participate?

25 | Volume Two, Chapter 8 | Independent financial audits as an institutional reference

The US Securities Act of 1933, Securities Exchange Act of 1934, and subsequent SEC rules established independent accountants’ audits as important capital-market institutions. The main text borrows only the logic that interested parties cannot monopolize self-validation. It does not propose copying the accounting industry’s structure for AI audits or appoint any particular auditor the natural gatekeeper of high-risk AI.

26 | Volume Three, Chapter 9 | Limits of the Costa Rica comparison

Historical assessments of Costa Rica after abolishing its standing army should use verifiable sources, defined indicators, and explicit periods. This book does not adopt unverified dramatic narratives such as “it was widely mocked at the time,” nor treat superlatives such as “most stable, prosperous, and green” as facts needing no specified basis of comparison.

27 | Volume Three, Project Coordination | CERN

CERN is an international scientific organization established by member states through an international treaty. The main text uses it as a reference for sustained transnational scientific cooperation, not as a direct equivalent of this book’s proposed project-coordination model.

28 | Volume Three, Project Coordination | Gavi

Gavi has its own defined organizational, legal, and financial structures. The text uses it only as a reference for multi-actor global-health cooperation. “Project coordination” and “alliance” are analytical language, not a renaming of Gavi’s institutional identity.

29 | Volume Three, Knowledge Cooperation | The “species’ self-understanding”

The “species’ self-understanding” is a literary expression for changes in the scale of human self-understanding brought by knowledge projects such as the Human Genome Project. It is not a formal biological concept.

30 | Volume Two, Chapter 10 | The fabricated 2023 image of an explosion at the Pentagon

In May 2023, a fabricated image claiming an explosion near the Pentagon spread on social media and was subsequently refuted by US officials and information from the scene. Markets briefly fluctuated during the trading day. The book uses it only to show that inexpensive synthetic content can trigger real reactions before verification, not that “one fake image can control markets.”

31 | Volume Two, Chapter 10 | The Hong Kong deepfake videoconference fraud

In 2024, Hong Kong police disclosed that a company employee transferred funds as instructed after attending a videoconference with several deepfaked “colleagues and executives,” causing large losses. The book uses the case only to show that familiar faces, voices, and meeting settings can no longer carry high-risk authorization alone.

32 | Volume Two, Chapter 10 | Biden-voice robocalls in New Hampshire

Before the 2024 New Hampshire US primary, robocalls imitating Joe Biden’s voice discouraged some voters from taking part. Federal and state enforcement, regulatory action, and penalties followed. The case shows only that generative voice forgery has entered political communication and impersonation.

33 | Volume Two, Chapter 10 | Capability boundaries of C2PA and Content Credentials

The C2PA specification aims to provide verifiable provenance and history for digital content, explicitly requiring that the specification itself make no “good or bad” value judgments about provenance assertions. It

can verify the binding between assertions and assets, as well as signatures and tamper status. It cannot prove that real-world facts photographed, described, or interpreted are necessarily true.

34 | Volume Two, Chapter 10 | Paid Twitter/X blue checks and impersonation

After Twitter introduced a new paid blue-check scheme in 2022, several impersonation incidents occurred, including an account impersonating Eli Lilly and falsely stating that insulin was now free. This book does not attribute subsequent market-price movements to one post. It discusses only the trust risk when the public's understanding of verification diverges from the conditions for obtaining a badge.

35 | Volume Two, Chapter 10 | Starling Lab and digital-evidence provenance chains

Starling Lab and partners have explored cryptographic signatures, hashes, distributed preservation, and camera-origin authentication to maintain provenance and chains of custody for journalism and potential legal evidence. Such technologies help verify whether materials changed and the chains of assertions they came from. They cannot alone establish the full facts beyond the frame or legal responsibility.

36 | Volume Two, Chapter 10 | NIST's first post-quantum cryptography standards

In August 2024, the US National Institute of Standards and Technology approved and published FIPS 203, FIPS 204, and FIPS 205, its first three post-quantum cryptography standards, covering key encapsulation and two digital-signature schemes. The book discusses migration readiness on this basis, without presenting the quantum threat as a countdown to a definite year.

37 | Volume Two, Chapter 11 | Alaska's Permanent Fund and resident dividends

Alaskan voters established the Permanent Fund by constitutional amendment in 1976. Following subsequent legislative and judicial adjustments, the resident-dividend system paid its first equal dividends in 1982. The text uses it only to show that institutionalized shared benefits can form when specific public-resource revenues, long-term funds, and beneficiary eligibility have clear legal grounds. It does not prove all assets should pay dividends to everyone.

38 | Volume Two, Chapter 11 | Norway's Government Pension Fund Global

Norway's Government Pension Fund Global manages state oil and gas revenues over the long term for present and future generations. It operates under Norway's laws and governance structures and does not pay direct universal dividends to residents. It is used only as an institutional reference for carrying resource revenues across time and public long-term saving.

39 | Volume Two, Chapter 11 | Finland's basic-income experiment

Finland's statutory basic-income experiment of 2017–2018 randomly selected two thousand recipients of unemployment-related benefits to receive €560 unconditionally each month. Kela's final assessment found relatively small overall employment effects; participants reported several more positive outcomes in life satisfaction, psychological stress, financial well-being, perceived bureaucratic burden, and confidence in the future. The experiment does not prove basic income suits every country or answer all questions of work motivation and meaning.

40 | Volume Two, Chapter 11 | GiveDirectly's long-term basic-income research in Kenya

GiveDirectly began long-term basic-income research in Kenya in 2017. The long-term group was scheduled to receive monthly cash for up to twelve years, compared with shorter-term cash, lump-sum transfers, and controls. Interim research during the pandemic found cash buffered some hunger, health problems, and psychological stress, while external shocks still affected business and income. This cannot mean "security makes everyone work harder" or "human nature has been proved."

41 | Volume Three, Chapter 12 | The actual state of returns to education

“Returns to education have fallen to zero” is not generally the case today. Returns vary greatly by country, class, field, and educational level. The text discusses only how AI’s changes to some knowledge tasks require recalibrating the relationships among educational content, occupational returns, and subjects’ capabilities.