Introduction – levels of civilisation, their historic trajectory and the future with AI visibility

The introduction has nine components

1. The Problem, the Pattern and the Aim - Defines the civilisational problem through the David example, shows that technology has forced comparable transitions before but that AI is accelerating change across more domains, and establishes **More Mores for More** as the long-term aim for shaping the architecture that follows.

2. Going upstream to the fundamentals that underpin any level of any human organisation. There are five fundamentals:

- **Our Brain.** Humans have **positive and negative predispositions**. Civilisation can draw out, manage or repurpose them.
- **Holons.** Integrated wholes within wholes that include and transcend components level-by-level as complexity increases.
- **Truth and felt truth** is indispensable and fragile.
- **Cybernetic control.** Systems require control and correction, including an effective L+1. An essential 'guard the guards' process
- **Humans are problem solving machines**. Some problems are solvable, others are paradoxes that can only be held or settled for a time.

3. Diagnostic tools - At first these fundamentals may look separate. When integrated, recursive structures begin to appear level by level. This leads to diagnostic tools.

4. Using Diagnostics. With AI we can ask a powerful question: "at what level is the architecture now operating actually?" And then: "How to change it to meet short term and long term aims?"

5. Diagnostics, visible on everyone's phone, will lead to behavioural change. Approval and disapproval are powerful animators.

6. The propositions and diagnostics lead to multiple MVP pilots that together are part of a solution-seeking process.

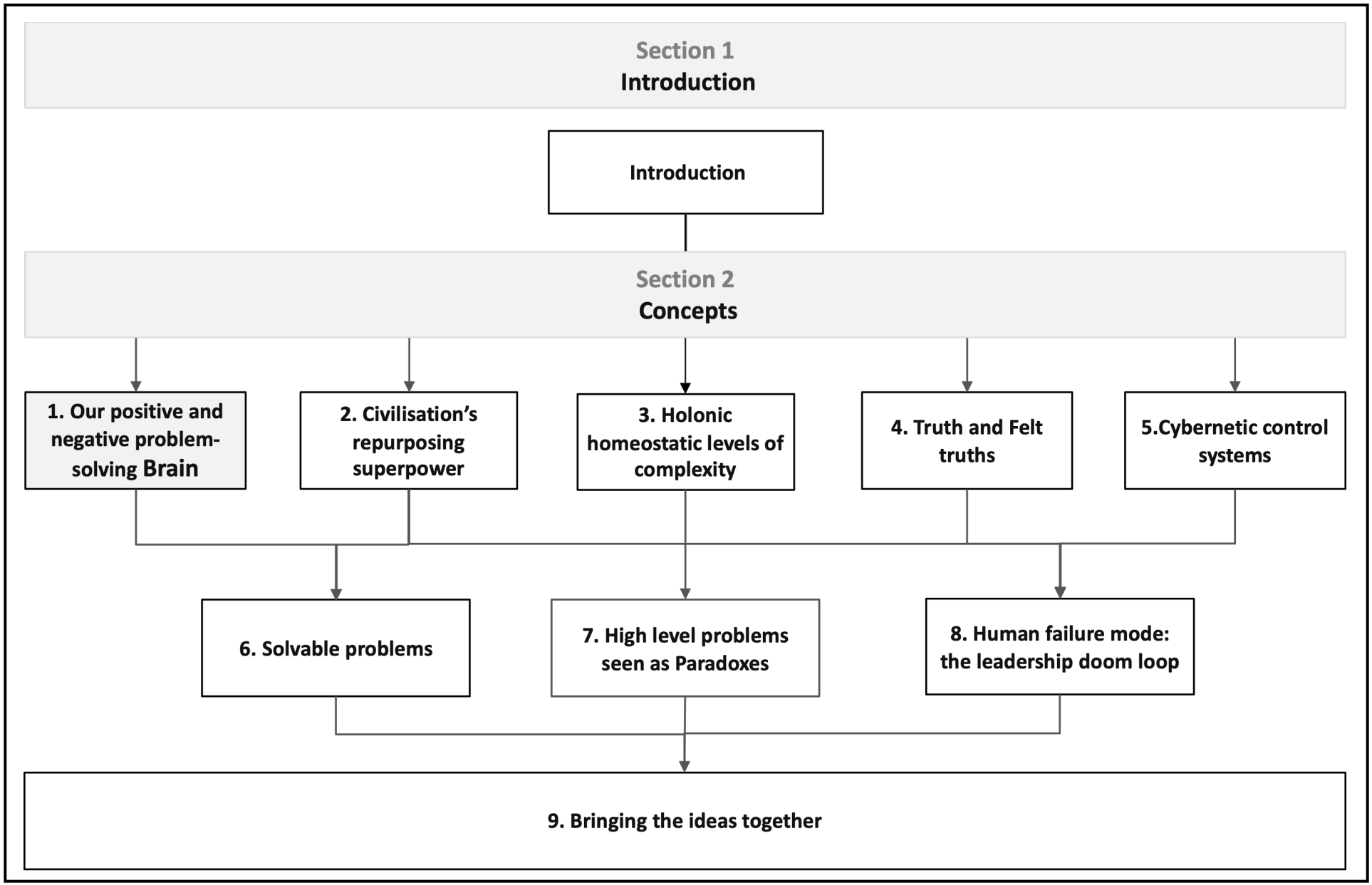
7. Multiple actors are needed to build, extend and implement pilots.

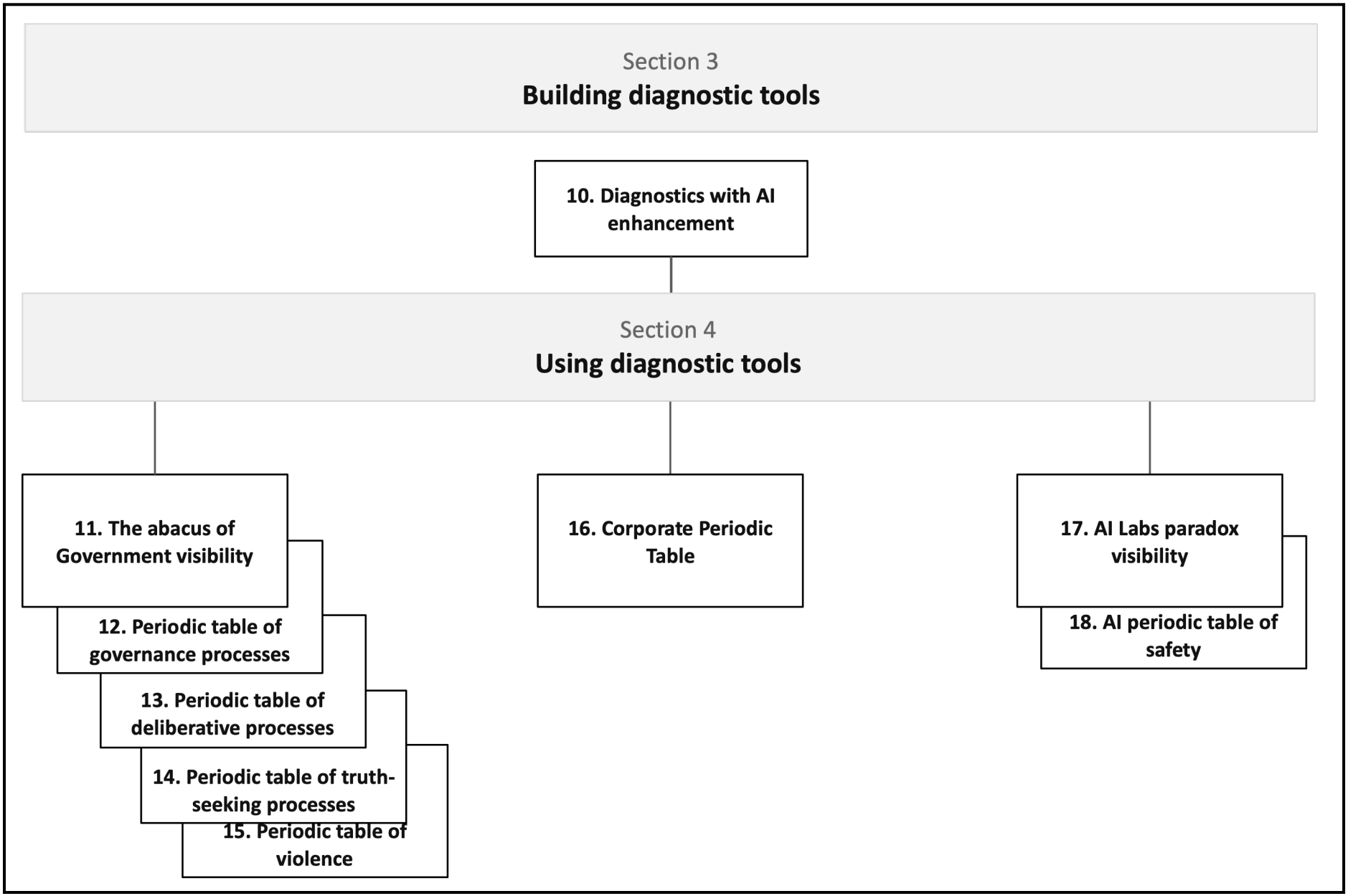
8. AI implications: How can AI help Humans?

9. AI implication: How can AI help control AI?

Schematics of the workbook Exhibit 0.1

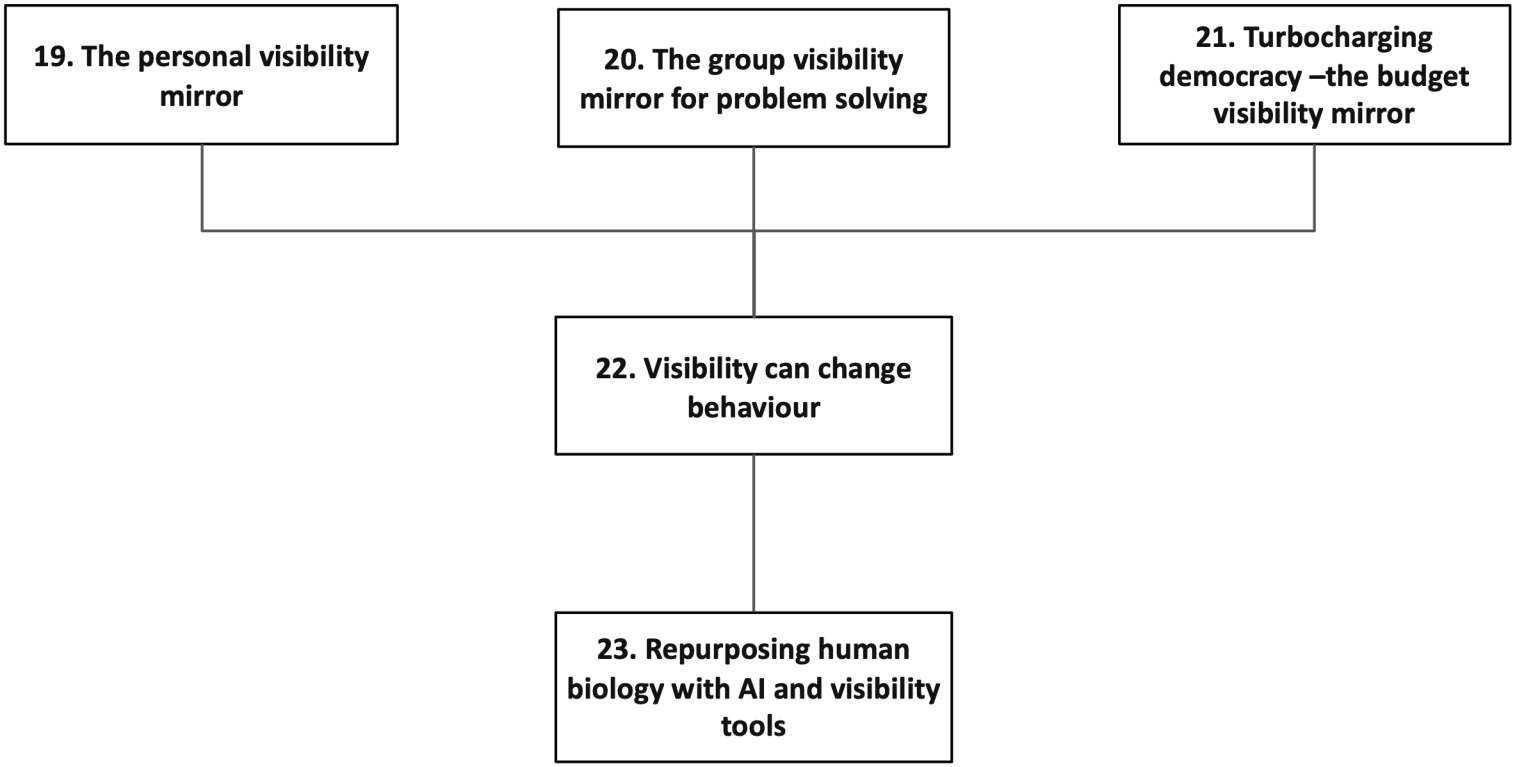
Section	Brief overview
Section 1 — Introduction	AI creates the possibility of greater abundance, but realising that promise requires deliberate changes to the human systems through which its benefits and disruptions will spread.
Section 2 — Concepts	Understanding what AI may change begins with understanding the human beings, institutions and correction processes into which it is being introduced.
Section 3 — Building diagnostics	Concepts become useful instruments when they help people see where a system stands, how it is changing and what action is justified by the evidence.
Section 4 — Using the tools	The diagnostic tools make institutions and their choices visible in enough detail for people to question them, compare them and identify practical improvements.
Section 5 — How Visibility Changes Behaviour	Visibility becomes a force for change when people can connect what they intend, what they do and what follows from their actions.
Section 6 — Pilots and Actors	The book's proposals become credible through practical trials that reveal what works, what fails and what others can usefully adopt.
Section 7 — Roundup	The book closes by asking whether human control and correction can develop quickly enough to carry the capabilities that AI makes possible.
Section 8 — Appendices	Bring together a selection of Julian's previously published books and learned papers, providing additional background, development and evidence for the ideas explored in the main text.


Section 1
Introduction
Introduction
Section 2
Concepts
1. Our positive and negative problem-solving Brain
2. Civilisation's repurposing superpower
3. Holonic homeostatic levels of complexity
4. Truth and Felt truths
5.Cybernetic control systems
6. Solvable problems
7. High level problems seen as Paradoxes
8. Human failure mode: the leadership doom loop
9. Bringing the ideas together


Section 3
Building diagnostic tools
10. Diagnostics with AI enhancement
Section 4
Using diagnostic tools
11. The abacus of Government visibility
12. Periodic table of governance processes
13. Periodic table of deliberative processes
14. Periodic table of truth-seeking processes
15. Periodic table of violence
16. Corporate Periodic Table
17. AI Labs paradox visibility
18. AI periodic table of safety

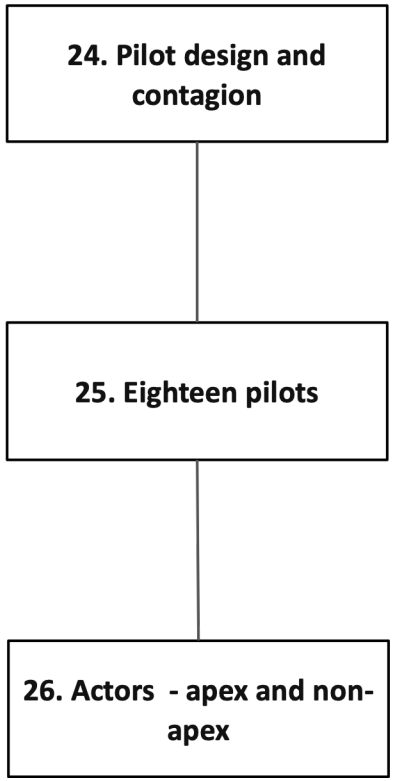
Section 5

How visibility changes behaviour



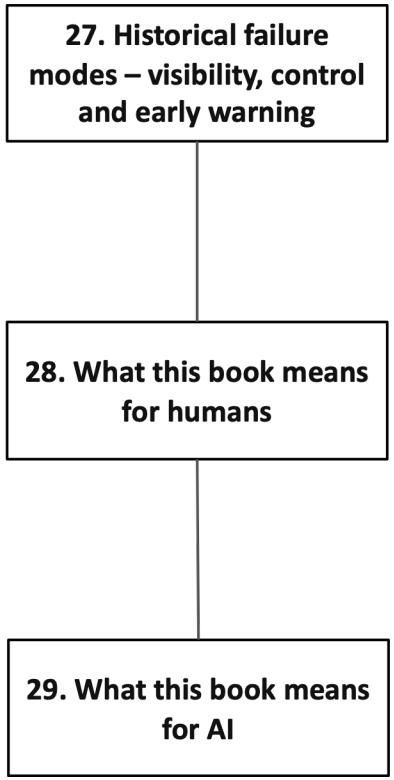
Section 6

Pilots and actors



Section 7

Pilots and actors



Section 8

Appendices

A. Building diagnostic mirrors for government, corporations and humans

B. Books and papers authored by Julian Fairfield

1. **Julian Fairfield biography**
2. **Holons**
3. **Deep dive into levels**
4. **Levels of excellence (book)**
5. **Who's in control of your brain? (book)**
6. **The Survival paradox (book)**
7. **Eritrea**
8. **Human transformation field**
9. **The wisdom machine**
10. **The Udon Lady**

1. The Problem, the Pattern and the Aim

Persistent cost reduction will disrupt the architecture of our whole civilisational system because AI and governance will change at very different rates.

A simple real-world example helps show how the effects can spread far beyond the business where the change first occurs, eventually reaching workers, families, banks, governments and the wider system.

David owns six McDonald's restaurants. Together they employ 300 full-time equivalent staff. The six restaurants sell about $30 million a year. The operating margin is about 10 per cent. Profit is about $3 million. The business carries about $20 million of loans. Interest costs about $1.5 million a year. A typical meal costs about $15. Labour costs $30 an hour.

Those are the starting facts.

David and I walk through the restaurants job by job. Ordering. Food preparation. Cooking. Assembly. Cleaning. Stock handling. Rostering. Administration. We ask the same question each time: how many people do we still need when agents and robotics are available at $3 per hour?

The answer is about 80 people. Around 220 FTE jobs disappear from the six restaurants.

David is excited. "I am going to make a fortune."

I tell him, "I don't think so. Let's push our thinking."

Every competitor, and especially new competitors, gets the new technology. The cost saving does not stay inside David's business. Competition pushes the price down.

The speed of price deflation is uncertain, but the trajectory is clearly down. We decided that at the end of ten years the typical $15 meal might sell for $5, a rate of decline of about 10 per cent annually.

That sounds wonderful. In one way it is. The same food now costs much less. AI and robotics have created much greater physical abundance with much less human labour.

But David still has the $20 million loan. He still has the $1.5 million interest bill. Debt does not fall just because prices fall. At $5, profit has dropped to about $1 million. He cannot pay his interest, let alone repay his loan.

David asks, "What do I do?"

My answer is: “It is not just you.”

Every bank in the country has loans written against yesterday’s prices, yesterday’s wages, yesterday’s asset values and yesterday’s cash flows. In a deflationary economy, new lending slows. Old loans become harder to repay. Bank balance sheets weaken. At sufficient scale, banks fail.

Now look at government. Two hundred and twenty people from these six restaurants are unemployed. They pay less income tax. They spend less, so GST falls. David makes less profit, so company tax falls. At the same time, government has to spend more on income support, housing, retraining and financial stability. Maybe governments inflate the economy, advantaging some and disadvantaging others.

Individuals and families are in trouble. Corporations are in trouble. Banks are in trouble. The financial system is in trouble. Government is in trouble.

Competitive incentive can create extraordinary abundance whilst simultaneously we slip into economic chaos.

Even more challenging, what happens to status and meaning when work is no longer the main way people earn income and social standing? We do not know. What happens to the idea of money if prices keep falling and work is no longer the main distribution mechanism? What does money become?

This is not fantasy. I have worked for years on the shop floor. My first job was on a turret lathe. I have worked with McKinsey at the highest levels of governments and management. I have reorganised hundreds of physical and administrative processes. I know both physical and conceptual processes. The arithmetic is real. The mechanism is real.

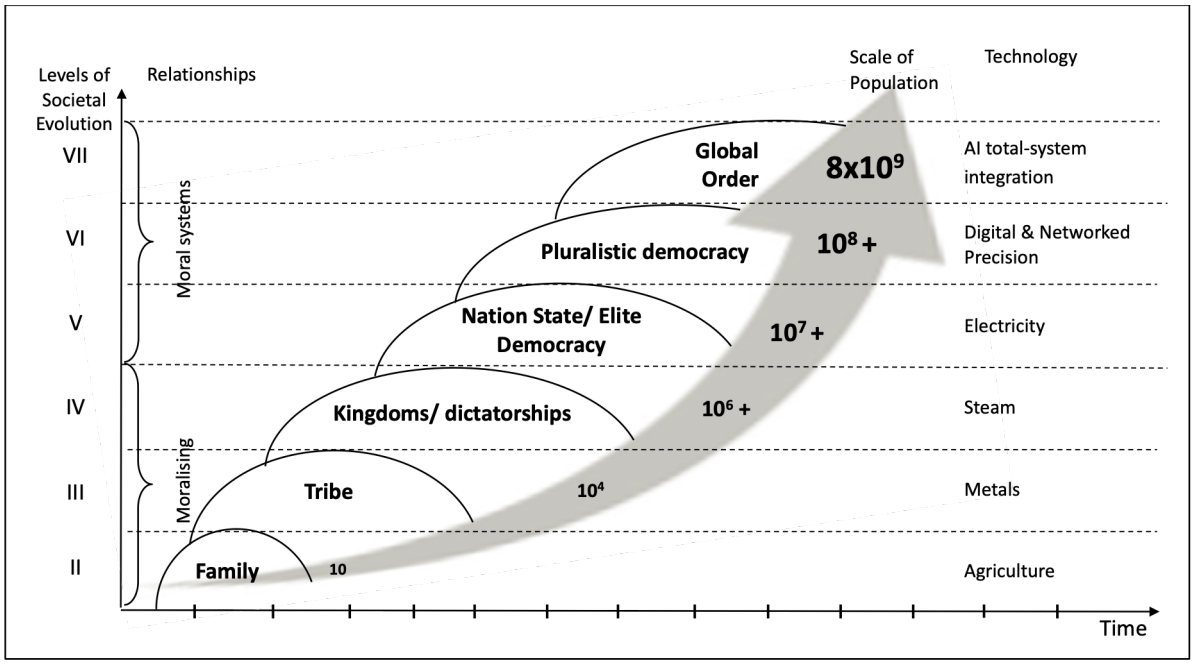
The hamburger is no longer simply a hamburger problem. Follow the effects far enough and it becomes a problem of civilisational architecture.

That may sound like an extraordinary conclusion to draw from six McDonald’s restaurants, but the underlying pattern is not new. Again and again, a major change in technology has altered the economics of everyday life and eventually forced the architecture around it to change as well.

We have been through transitions of this kind six times before. Agriculture changed how people gathered, lived and organised themselves. Cities changed governance. Writing changed memory. Money changed exchange. Printing changed knowledge. Steam changed production. Electricity changed daily life. Industrialisation changed the relationship between people, work and capital.

Human organisation evolved through six successive levels of governance and social architecture alongside these changes.

Six historical stages of human governance. Exhibit 0.2



Technology forcing civilisation to change is therefore not new. Each major change created new possibilities and broke old architectures. People then built new architecture, new rules and new institutions around the new reality — sometimes peacefully, sometimes through depression, revolution, war and chaos.

This time there is an important difference. AI is moving faster and across more domains than previous technologies.

AI can already write, analyse, diagnose, translate, design, programme, advise and simulate. Robotics gives that intelligence hands and machines. Networks allow millions of intelligent systems to operate at the same time. AI can also help build the tools used to build better AI, so the technology helps accelerate itself.

This does not make chaos inevitable. It does make the speed of change important and the outcome unusually uncertain.

In summary, civilisation faces a familiar problem at unfamiliar speed: the world changes faster than the architecture built to manage it.

That brings us to the aim.

AI is civilisational in scale. Short-term aims such as safety are essential, but they need to sit underneath a long-term aim that matches the scale of the problem. An aim limited to one company, one elite or one country will eventually create exclusion.

Over time, exclusion produces resentment, political reaction and instability and, at the extreme, revolt against the architecture itself. History repeatedly shows that systems which benefit too few eventually lose legitimacy.

Short-term aims will produce **physical abundance**. Technology and corporations naturally focus on what is measurable and profitable: productivity, goods, income, energy, housing, transport and consumption. These matter enormously, and AI may give us levels of physical abundance previous generations could barely imagine.

But human beings are not only physical creatures. We also need **relational abundance**: trust, belonging, dignity, recognition, purpose, reciprocation, participation, family and community. A society could become materially richer while becoming relationally poorer.

So, the aim has to be broad in two directions: **more kinds of abundance, for more people.**

That is why **More Mores for More** matters. It is not simply an attractive moral idea. It is a design constraint for a stable civilisation. An adequate long-term aim cannot be confined to producing more physical things. We have to be concerned with the quality of the relationships, institutions and social architecture within which that abundance is produced and distributed.

More Mores for More. Exhibit 0.3

Physical abundance	Relational abundance	
• Safety	• Trust	• Freedom
• Food	• Belonging	• Responsibility
• Shelter	• Dignity	• Compassion
• Health	• Recognition	• Meaning
• Education	• Reciprocation	• Cooperation
• Energy	• Fairness	• Hope
• Infrastructure	• Empathy	• Kindness

Having established the problem, our historical pattern of evolution and the need for an aim encompassing physical abundance and relational abundance we will now move to the five fundamentals that sit underneath every level of human organisation.

2. Go upstream to the fundamentals that underpin any level of any human organisation."

Safety, regulation, and jobs matter. But they sit downstream of a bigger question: what human architecture do we need? We start with the human Brain, move on to repurposing our predispositions, explain levelled holonic structures, wrestle with truth and felt truth, utilise the science of cybernetic control, and expose the difference between solvable problems and paradoxes.

- **Our Brain.** Humans have **positive and negative predispositions**. Civilisation can draw out, manage or repurpose them.

  Our survival problem-solving brain is both the maker and breaker of civilisation. The same brain has pro-social and anti-social predispositions - acting potentially with love, care, kindness, cooperation and reciprocation; or extreme competition, status seeking, self-interest, aggression. It also creates felt truths that it defends with survival vigour.

  Civilisational architecture has never removed those predispositions. It changes how they are expressed. It draws out some. It restrains some. It **repurposes** others level by level.

  - Competition can become productive enterprise.
  - Status seeking can become public service.
  - Self-interest can become reciprocal exchange.
  - Aggression can be held by law.

- **Levelled Holonic structures let us see show that civilisation is made of wholes within wholes.** Each level includes and transcends lower levels to form a fully integrated mutually reliant system. A person is a whole and part of a family, firm, community, nation and civilisation. A town is a whole and part of a region, state, economy and wider order. Each level is a level of increasing complexity. This can be observed in multiple systems – the physical world, living world, government world, corporate world. Once this is recognised we have a template upon which to hang all other components.

- **Wrestling with truth and felt truth is difficult.** Human beings interpret reality through experience, identity, institutions and **felt truths**. Felt truths are useful but may or may not be true. Different people have differing opinions on: Free markets are… Socialism is… Rules are... Democracy is… Cooperation is… Merit is… Societies therefore require structured truth-

seeking processes that can test sources and methods, hear counterclaims, make uncertainty visible and correct error.

A civilisation that stops finding and correcting error loses the basis upon which decisions can be made. That's the fundamental difficulty with "I know best" whether it is expressed by kings, dictators, or communist leaders

- **The science of cybernetic control lets us understand civilisational feedback architecture**. Viable systems have aims, boundaries, information flows, standards, comparisons, triggers, corrections, escalations, sanctions, forgiveness mechanisms, learning loops and reset procedures. Where those loops work, systems correct themselves. Where they fail, damage accumulates until a crisis makes it visible.

  As Systems are levelled good control needs something that can stand one level above the thing being controlled. That is the L+1 principle. It applies to governments, regulators, companies, AI laboratories and eventually AI itself. If AI has power over human life, what sits one level above it? The answer cannot be another uncontrolled AI. Nor can all control safely sit with a tiny group of humans holding extraordinary technological power.

- **Exposing the difference between solvable problems and paradoxes** – Many problems are directly solvable, even though they may be complex - like, autonomous robotics, protein folding, better chips, self-driving cars, even reusable rockets.

  But the highest-level problems cannot be solved. They are paradoxical relational tensions often with winners and losers at either pole.

  Short term versus long term; 'I' versus 'We'; competition versus cooperation; speed versus safety; freedom versus order; truth versus harmony; punishment versus forgiveness.

  Pushing either pole too far damages the larger system. Higher-level problem solving therefore requires societies to deliberate and make settlements between either end of the spectrum, legitimising those settlements, observe how they operate and adjust them as circumstances change.

3. Diagnostic tools.

At first these fundamentals may look separate. When they are integrated, recursive structures begin to appear level by level. The world stops looking like an endless collection of unique political, corporate and social architectures or problems. This leads to diagnostic tools.

They let us stand outside the systems we have built, make their hidden architecture visible, diagnose how they work, compare alternatives, and deliberately redesign them.

The diagnostic questions

- What has to be done in every human system?
- At what level is it being done?
- How is it being done?
- Who controls it?
- What truths is the system using?
- What human predispositions is it drawing out?
- What paradoxes is it settling?
- Who benefits from those settlements?
- Where is correction taking place — and is there an effective L+1?

There are two primary organisational diagnostic tools: the abacus and the periodic table.

The abacus: makes paradox settlements visible

Many arguments that appear to be battles over truth are actually different settlements of persistent tensions: individual versus collective, present versus future, freedom versus control, competition versus cooperation, merit versus fairness.

Once the settlement is visible, we can ask: Where are we positioned? Who chose that position? At what level? For whose benefit? What would happen if we moved?

This is a short summary of a very new idea. Chapters 7 and 11 expand on this.

The periodic table: makes the architecture visible level by level

A periodic table separates what a system must do from how it chooses to do it. The nouns are relatively invariant. The verbs vary by level.

Every society needs leadership, rules, boundaries, deliberation, truth-seeking, wealth creation, education and correction. Those are the nouns. Rome, modern Denmark, the United States, a corporation and a family perform those same functions but in radically different ways. Those are the verbs.

This is a short summary of a very new idea. Chapters 6, 12-16 and 18 expand on it.

4. Using Diagnostics.

With AI we can ask a powerful question: “At what level is the architecture now operating?” And then: “How to change it to meet short term and long term aims?”

We can make visible:

- Ourselves to ourselves and, with permission, to each other
- Paradox settlements
- Levels of:
    - Governance
    - Deliberative processes (democracy)
    - Truth-seeking processes
    - Use and control of violence
    - Corporate capability
    - AI performance and safety
- We can make comparisons and simulate futures

These diagnostics are available in chapters 11 through 18 in MVP form. You can build them for yourself right now.

5. Diagnostic visibility on everyone’s phone will lead to behavioural change

Human beings have always shaped one another through approval and disapproval. We seek trust, status, belonging and recognition, and avoid rejection and loss of standing. These responses help draw out positive predispositions and restrain or repurpose negative ones. But they only work when behaviour is visible. AI radically expands what can now be seen.

Visibility is an old control mechanism. In small groups, people could see who shared, cheated, kept promises, lied, contributed or abused power. Visibility created consequences: trust, reputation, status, influence and cooperation. It

helped turn competition into achievement, status-seeking into service, self-interest into reciprocity, and generosity or truthfulness into behaviours that could be rewarded.

Scale weakened visibility. As societies grew, behaviour disappeared inside hierarchy, bureaucracy and aggregated reporting. It became harder to see who made decisions, whose interests were served, whether promises were kept, or where responsibility actually sat.

AI changes this. AI can now create visibility vertically and horizontally:

- Personal — habits, strengths, weaknesses and contradictions.
- Group — cooperation, conflict, trust and exclusion.
- Corporations — how decisions are made, whether behaviour matches aims, and where failure is occurring.
- Nation and beyond — policy choices, trade-offs, outcomes and institutional level, potentially up to global bodies such as the United Nations.

Visibility makes further repurposing possible. We cannot repurpose what we cannot see. Once excessive competition, greed, domination, status-seeking, certainty or short-termism become visible, we can ask what is driving them, what is the architecture rewarding, and how the same predisposition might be repurposed towards a better outcome.

Visibility must work both ways. Visibility must not mean only that institutions can see citizens. Citizens must be able to see institutions; employees must be able to see organisations; lower levels must be able to see higher levels; and higher levels must stay connected to reality at the operating level.

A new capability. AI gives us something new at scale: the ability to see ourselves, our organisations and our societies as architecture while they are operating. Once behaviour and consequences become visible, we have a far greater chance of changing them — and of repurposing human predispositions towards

More Mores for More.

6. The propositions and diagnostics lead to multiple MVP pilots

We must build small pilots, use them, see what happens, correct them, scale what works and make them contagious.

Good pilots have high visibility, meet aims, are cheap, quick to complete, are scalable and they must be resourced sufficiently to enable success.

We have divided the pilots into four groups - Orientation, Self-understanding, Diagnostics, and Problem solving.

Group A – Orientation and shared language

1. **More Mores for More — orientation and promulgation**. Build More Mores for More into a simple, usable public purpose for AI and spread it widely enough to orient the more specific aims that sit underneath it.

2. **Primitives Literacy**. Build simple, level-appropriate ways for people to understand the core primitives well enough to use the later tools intelligently.

Group B —Agency, self-understanding and mirrors

3. Permissions. Build practical ways for individuals to control what their AI can know about them and what it is permitted to do with that information.

4. Biomarkers. Add inexpensive physiological information where it genuinely improves the usefulness of a Personal Visibility Mirror, while keeping clear limits around what the data can tell us.

5. The Personal Visibility Mirror. Give individuals better visibility of their own patterns, inconsistencies and outcomes so they can reflect and choose differently.

6. The Group Visibility Mirror. Extend visibility to groups so that participation, disagreement, reciprocity, cooperation and problem-solving patterns can be seen and improved.

7. National Visibility Mirrors. Build national and, later, civilisational visibility without averaging away disagreement, minorities or changing conditions.

Group C — Diagnosis, truth, control and safety

8. Periodic Tables. Build usable maps of the relatively invariant parts of systems so that what is absent, weak or operating at the wrong level becomes visible.

9. The Abacus. Make paradox positions visible so that people can see where they agree, where they differ and what settlement is currently being held.

10. Truth verification. Build an AI-supported truth-seeking process that improves the handling of claims, evidence, uncertainty and correction without turning AI into the authority on truth.

11. Control and correction using the 8+1 concept. Build control systems that sense deviation, compare reality with intention and correct human systems faster.

12. Early warning system for human failure modes. Build a visibility system that detects recurring human failure patterns early enough for correction before they become crises.

13. Pilot must-not-do and early warning. Put explicit boundaries and dynamic early-warning mechanisms inside every pilot so that emerging risk is detected and corrected while the work continues.

14. AI civilisation must-not-do. Build a credible process for identifying, constraining and continually revising boundaries around AI capabilities whose consequences could be catastrophic.

Group D —Solving, settling and deciding together

15. **The ten-step solvable problem process**. Use disciplined AI-assisted problem solving to produce better solutions faster, while recognising when the issue is a paradox rather than a soluble problem.

16. **Paradox settlement**. Use truth-seeking, the Abacus, deliberation and AI to reach workable current settlements where no permanent objectively correct answer exists.

17. **USA-China cooperative paradox pilot**. Use shared visibility to move the USA-China AI relationship from slogans and winning-versus-losing toward specific paradoxes that the two countries must manage together.

18. **Turbo-charging democracy**. Put positive and negative dollar allocation, the Abacus, factual information, AI support and Visibility Mirrors into a working democratic process that improves the information flowing between citizens and political decision-makers.

7. Multiple actors are needed to build, extend and implement pilots

The six actor groups are complementary. AI labs possess extraordinary capability but cannot legitimately choose humanity's ends. Government possesses public authority but is further from the frontier and moves more slowly. Independent experts can challenge both but do not acquire public authority merely through expertise. Pathfinders can move rapidly but need permissions, hosts and boundaries. Educators build the human capability on which the tools depend. Citizens supply the participation and legitimacy that make the whole system human rather than merely technical.

Six actor groups. Exhibit 0.4

Actor	Primary contribution	What they do
1. **AI Labs** — The Builders	Safety, frontier technology, technical knowledge, compute, engineering and rapid learning.	They build the systems and are closest to both capability and emerging risk.
2. **Government** — The Legitimate Boundary Setter	Political leadership, professional government executive, regulators, law and public authority.	Government can create binding public settlements and enforceable boundaries that private actors cannot.
3. **AI Refugees and Independent Experts** — The Witnesses	Insider knowledge, informed dissent, independent challenge and technical scrutiny.	They can expose blind spots and contribute frontier knowledge outside the immediate incentives of a lab.
4. **Pathfinders and Catalysts** — The Makers	Entrepreneurial energy, investment, philanthropy, universities, technology and organisational drive.	They turn the architecture into working mechanisms quickly and can carry successful mechanisms into new settings.
5. **Educators** — The Capability Builders	Brain literacy, truth-seeking, paradox literacy, permissions literacy and learning methods.	The rest of the architecture works only if humans can understand and use it.
6. **Citizens** — The Participants and Ultimate Principals	Participation, judgement, lived consequence and democratic legitimacy.	The ends toward which AI and institutions are used remain a matter of human choice.

8. AI implications: How can AI help Humans?

The central human opportunity from AI is not simply that it can analyse more information or answer questions faster. It is that, combined with visibility, it can help humans see themselves and the systems around them far more clearly than has previously been possible.

That matters because civilisation has always depended on repurposing human predispositions. Threat, status, belonging, competition, cooperation, reciprocity and self-interest can all produce good or bad outcomes depending on the setting around them. Historically, law, custom, religion, markets, education and institutions have slowly learnt how to channel those predispositions. AI gives us the possibility of making that process much more deliberate.

The opportunity therefore works in two directions. First, **better visibility** can help people recognise their own patterns and choose differently. Second, the same visibility can help us **see the architecture in which people live and work**, identify where it is weak or mismatched, and improve it. AI does not replace human judgement. It gives humans a new capacity to see, compare, remember and correct.

How This Section Unfolds

8.1. Six fundamentals — shows what AI can add to each of the book's five basic human and institutional building blocks and their failure.

8.2. Eighteen pilots — shows how those capabilities become practical tools and working arrangements.

8.3. What becomes possible when the pieces combine — describes the new human capability created by continuous visibility across self, group and institution.

8.4. Misuse — shows how the same capabilities can be turned toward surveillance, manipulation, propaganda and control.

8.5. The human role — keeps purpose, judgement, legitimacy and responsibility with people.

8.1. Six Fundamentals

The fundamentals give us a useful way to make the human–AI partnership specific. For each one, the question is simple: what can humans now do, with AI assistance, that was previously too difficult, too slow or too fragmented to do well?

Six fundamentals. Exhibit 0.5

Fundamental	What AI can help humans do
Human predispositions	See recurring patterns of threat, status, belonging, competition, cooperation and reciprocity, then use that visibility to strengthen constructive behaviour and reduce destructive behaviour.
Levels	Locate the level at which a problem, decision or institution is actually operating, expose level mismatches, and see when the response needs to move upward or downward.
Truth and felt truth	Separate claims, evidence, sources, counterclaims, uncertainty and interpretation across volumes of information that no person could reliably hold alone.
Control and Correction	Sense what is happening, compare it with the aim, detect drift earlier, identify whether correction is occurring, and ask who is controlling the controller.
Solvable problems and paradoxes	Distinguish issues that can be solved from tensions that must be continually settled, while making competing goods, consequences and trade-offs visible.
Failure	Look continuously for early signs of truth loss, power concentration, capture, weakened correction, declining legitimacy and other recurring failure patterns.

The common thread is visibility. AI can help humans see patterns that were previously hidden in memory, volume, complexity or time. That does not automatically improve behaviour. But it creates the opportunity for reflection, comparison and correction. It also makes it possible to improve the architecture around people rather than relying only on people becoming better.

8.2. Eighteen Pilots

The pilots demonstrate how AI can help humans build a working capability, make it visible and improve it through use.

Pilot	What AI can help humans do
1. More Mores for More	Help people understand, discuss and apply a common aim when judging choices and consequences.
2. Primitives Literacy	Explain the fundamentals at different levels of depth so that children, citizens, managers, politicians and specialists can use the same underlying ideas.
3. Permissions	Give individuals practical control over what their AI can know about them and what it is allowed to do.
4. Biomarkers	Add physiological information where it genuinely improves self-understanding, while keeping clear limits around what the data can mean.
5. Personal Visibility Mirror	Help an individual see recurring patterns, inconsistencies, choices and outcomes so that behaviour can be consciously changed.
6. Group Visibility Mirror	Make participation, disagreement, reciprocity, cooperation and problem-solving patterns visible to the group itself.
7. National Visibility Mirrors	Help societies see agreements, disagreements, changing conditions and institutional patterns without averaging away minorities.
8. Periodic Tables	Map stable parts of a system and show what is absent, weak, mismatched or operating at the wrong level.
9. The Abacus	Make paradox settlements visible so people can see where they agree, where they differ and what competing goods are being protected.
10. Truth Verification	Organise claims, evidence, source quality, counterclaims and uncertainty while leaving judgement open to human challenge.
11. Control and Correction	Help systems sense deviation, compare reality with intention and correct earlier.
12. Early Warning for Human Failure Modes	Detect recurring patterns of drift, capture, truth loss and weakening correction before they become crises.
13. Pilot Must-Not-Do and Early Warning	Put explicit boundaries and live warning mechanisms inside every pilot so emerging risk is visible while work continues.
14. AI Civilisation Must-Not-Do	Help humans identify, compare and continually review boundaries around AI capabilities whose consequences could be catastrophic.
15. Ten-Step Solvable Problem Process	Assist disciplined problem-solving across evidence, options and consequences while recognising when the issue is not actually soluble.
16. Paradox Settlement	Support truth-seeking, deliberation and Abacus visibility so humans can reach workable current settlements where no permanent answer exists.
17. USA–China Cooperative Paradox Pilot	Create shared visibility around specific tensions that two powerful states must manage together rather than reducing everything to winning and losing.
18. Turbo-charging Democracy	Help citizens and politicians work with the same visible information and trade-offs, compare choices and strengthen democratic accountability.

8.3. What Becomes Possible When the Pieces Combine

The real change appears when these capabilities stop being separate. A person could have access to an assistant that helps them see themselves, understand the group they are working in, inspect the quality of the information they are using, locate the level of the problem, identify the paradoxes involved, remember previous decisions, compare intended and actual outcomes, and detect signs that the system is drifting.

No previous human has had that combination continuously available. It creates the possibility of a new form of human capability: not humans handing judgement to machines, but humans becoming more visible to themselves and more able to correct the systems they build.

This is why the two forms of repurposing belong together. AI can help humans repurpose behaviour by making predispositions and consequences more visible. It can also help humans repurpose architecture by showing where institutions, controls, truth processes and levels no longer fit the problem they are meant to handle. Better people and better architecture can then develop together.

8.4. The Same Capabilities Can Be Misused

There is an exact symmetry to the opportunity. The same capabilities that can increase visibility and correction can also be used to manipulate and control.

Visibility can become surveillance.

Knowledge of predispositions can be used to trigger fear, status, belonging or tribal loyalty rather than help people understand them.

Personalisation can become persuasion that is so individually tailored that the person does not recognise how they are being influenced.

Truth tools can be used to flood, filter or frame information in ways that reinforce felt truths rather than challenge them.

Diagnostics can help institutions improve themselves, but can also help powerful actors identify dissent, weakness or opposition.

Control systems can improve correction, but can also centralise authority and make control more efficient than challenge.

The key question is therefore not whether AI is being used. It is which human predispositions, which aims and which architecture the use of AI is strengthening. Visibility without permission can become surveillance. Behavioural understanding without restraint can become manipulation. Intelligence without truth and correction can make a bad architecture more effective.

8.5. The Human Role

The human role remains central. Humans set the aim. Humans live the consequences. Humans decide which competing goods deserve protection. Humans grant legitimacy, accept responsibility, forgive, cooperate and choose what kind of society they want to inhabit.

AI can greatly enlarge what humans are able to see and work with. It can give us memory, comparison, pattern recognition and feedback at a scale that was previously impossible. But the point of that capability is not to remove humans from the system. It is to make humans and human institutions more visible, more capable of learning and more correctable.

The fundamentals tell us what humans need to become better at. The pilots show how AI can help us do it. Together they suggest a practical human programme: use AI to see ourselves more clearly, repurpose our predispositions more deliberately, improve the architecture around us and keep the resulting power visible and correctable.

9. AI implications: How can AI help control AI?

Humans have spent thousands of years learning how to control humans. We have never done it perfectly, but civilisation has accumulated practical knowledge about restraint, correction, truth-seeking, distributed power, sanctions, reputation, reciprocity and the control of the controller. Much of Hope Before Chaos is an attempt to make that architecture visible.

The question is whether some of those lessons can transfer to AI. The comparison has limits. AI does not possess human biology or human predispositions in the literal sense. The useful analogy is therefore not between the human brain and an AI model. It is between the control architectures built around powerful humans and those we may need around powerful AI.

9.1. The Human Lesson

Civilisation discovered that internal self-control is not enough. Humans under threat, ambition or concentrated power can fail. So we built external architecture: law, audit, appeal, independent truth-seeking, divided authority, reputation, hierarchy and sanctions. We do not trust an important system to one act of self-restraint.

The deeper pattern is familiar from Control and Correction: set an aim and boundary; sense what is happening; compare it with what should be happening; interpret the difference; decide whether correction is required; act; then learn. At higher levels we add another requirement: the controller must itself be visible and correctable.

That suggests the first transferable principle. Important AI capability should not depend on one model simply deciding to behave well. It should sit inside an architecture of observation, challenge, restraint and correction.

9.2. What Transfers — and What Does Not

Human mechanisms such as shame, honour, belonging and fear of exclusion may not transfer directly. But the architecture underneath them can. AI can participate in a recursive control system in which one system observes, tests, challenges and, within limits, constrains another.

The question therefore changes from “Can AI possess human self-control?” to “Can AI be placed inside a control architecture strong enough that no single AI has to be trusted on its own?”

9.3. A Possible AI-on-AI Architecture

This looks less like one AI controlling itself and more like an institutional ecology of AIs. One system acts. Another monitors. Another red-teams the plan. Another checks evidence. Another looks for anomalies or goal drift. A higher-level controller decides whether activity can continue.

The analogy with separation of powers is especially useful. Civilisation learnt not to let the same actor make the rule, execute it, judge whether it was followed and suppress the evidence. The AI equivalent is to separate acting, monitoring, judging and authorising exceptions.

Possible AI on AI architecture. Exhibit 0.6

Human lesson	Possible AI equivalent
Aim above actor	Controller AI checks behaviour against externally set aims and constraints; it does not rewrite them.
Visibility	Independent monitor models inspect plans, actions, tool use and outcomes.
Plural truth-seeking	Different evaluators seek counter-evidence and challenge one another.
Separation of powers	No single AI acts, judges itself, controls the evidence and approves its own exceptions.
L+1 control	Higher-level controllers supervise agents whose consequences exceed their local boundary.
Sanctions	Controllers can pause, isolate, revoke tools, roll back or escalate.
Learning	Incidents and near misses update tests, thresholds and future constraints.
Control the controller	Controller models are cross-checked, audited and prevented from silently expanding authority.

9.4. AI Can Perform Much of the Control Cycle

AI is unusually well suited to monitoring at machine speed. It can compare current behaviour with large bodies of previous behaviour, look for anomalies, replay decisions, run counterfactual tests, generate adversarial challenges and maintain continuous records.

In bounded domains, AI may also perform limited correction: pause an agent, remove access to a tool, require a second check, revert a change, quarantine an action or escalate the case. That could make control far more continuous

than human institutions, where audits, courts and elections are necessarily episodic.

But the controlling aim should remain outside the system being controlled. An AI may identify that another AI has crossed a boundary without becoming the legitimate source of that boundary.

9.5. The Controller Problem Does Not Disappear

Putting one AI above another moves the control problem; it does not remove it. The controller can be wrong, share the same blind spots, become too permissive or too restrictive, or drift as the system beneath it changes. Two models trained in similar ways may also make correlated errors.

The human lesson is again useful. We did not solve corruption by appointing one all-powerful anti-corruption official. We learnt to use plurality, independent evidence, divided authority, appeal and review. AI-on-AI control will probably need several controllers that fail differently, explicit limits on their powers, records that cannot be silently rewritten, and an L+1 mechanism able to intervene when the control layer itself drifts.

9.6. Control Should Rise with Consequence

The control architecture should also be levelled. Low-level, reversible activity may be handled largely through automated monitoring and correction. As actions become more consequential, irreversible, cross-system or civilisational, control should move upward. The higher the consequence, the stronger the independence, plurality and authority of the control layer must become.

This connects directly to the book's recurring warning about level mismatch. A high-level capability should not sit inside a lower-level control system. If AI can act across an ecosystem, its control cannot remain only an output filter. If it can create national or civilisational consequences, the control architecture must rise accordingly.

Conclusion

The comparison with human civilisation holds, but at the level of architecture rather than psychology. Humans learnt that capability needs restraint, truth

needs independent challenge, power should be divided, controllers require controllers, and correction must occur before failure becomes irreversible.

AI controlling AI therefore should not mean an autonomous machine deciding its own morality. It should mean an engineered ecology in which powerful systems are continuously observed by other systems, challenged from several directions, prevented from silently changing their own boundaries, and corrected or stopped when necessary.

The deepest lesson may be this: civilisation works not because humans became perfectly self-controlling, but because we slowly built architectures that made our failures more visible, containable and correctable. AI-on-AI control may need to begin from exactly the same insight.