

A Brief History of Earth Intelligence

The history, predicament and future of Earth intelligence

A Brief History of Intelligence on Earth

Chapter One: Human Civilization

Prologue: Three Supercars

There are three scenes, framed from similar distances and rendered in a consistent style. Placed side by side, they resemble a set of condensed images of the human condition.

The first scene takes place on a professional racetrack.



A red supercar enters a turn. Its driver is a professional racer: helmeted, focused, and keeping the car close to the correct line as it passes the apex. The performance of the vehicle, the ability of the driver, and the demands of the track are in relative balance. The driver does not merely know how to make the car move. He understands the vehicle's limits—when to accelerate, when to brake, and what a single mistake in judgment might cause.

In this scene, technology and governing capacity remain relatively balanced. The car is fast, but the driver understands it. The car is

powerful, but the driver knows how to handle it. The limits of the vehicle remain within the range that a human being can perceive and train for.

The second scene shows an advanced supercar, and now it is on a highway.



The driver is a teenager. He is poorly trained, wearing no helmet, and carrying himself with reckless confidence as he weaves at high speed among ordinary vehicles. The surrounding cars do not have the same performance, and the other drivers are not prepared for this kind of speed. The car is faster than the teenager, and the teenager is more reckless than the car. The vehicle has already begun to exceed the true limits of his understanding, but he does not realize it.

All he knows is that pressing the accelerator is thrilling. He does not understand that every acceleration reduces his reaction time, or that every lane change increases the risk to everyone on the road.

The problem is not that the teenager is completely incapable of driving. It is that he is capable of making the car move, but not capable of understanding what the car truly means.

The third scene takes place in a crowded city center.



It is still the same red supercar, but now it is more powerful and faster than before. The driver, however, is a child of about three years old. He can barely reach the steering wheel. He cannot see the full traffic situation, does not understand the rules of the road, and has no way to grasp what speed, weight, or braking distance mean. Around him are pedestrians, vehicles, intersections, and complex traffic signals. The more frightened the child becomes, the tighter he grips the wheel; the less control he has, the more chaotic his movements become.

This is not to suggest that humanity is literally as intelligent as a three-year-old. Rather, it suggests that, relative to the complexity and destructive power of the technological systems we have created, our collective capacity to understand, govern, and responsibly deploy them remains dangerously underdeveloped.

What remains unchanged across the three scenes is not the environment, but the vehicle. Technology continues to grow more powerful. What changes is the degree of fit between the driver and the environment.

In the first scene, the driver's ability is proportionate to the technology.

In the second, the technology has begun to exceed the driver's understanding, but the driver is still excitedly pressing the accelerator.

In the third, the driver is facing a system far beyond his ability, yet he is still placed behind the wheel.

This is not a picture about cars. It is a picture about the civilization's capacity to govern what it has built.

The most dangerous contradiction in human civilization is not that we lack the ability to build more powerful tools. It is that the speed at which we build them has surpassed the speed at which we can understand them, manage them, and assume responsibility for their consequences. Humanity possesses increasingly powerful technologies, but it has not acquired, at the same pace, the cognitive, institutional, and moral capacities needed to match them.

In the age of early technologies, a ruler's mistake in judgment might destroy a city, an army, or a dynasty. In the modern age, a mistake of the same basic kind may affect global finance, the climate, supply chains, the information environment, and even the conditions for human survival as a whole.

Technology remains in human hands, but humanity is gradually losing its ability to understand the overall consequences of technology.

A more fundamental question therefore emerges: what happens when humanity can no longer reliably control the technologies it has created?

It may extend the life of its civilization through institutions, religion, law, and cooperation. Or it may gradually lose control as it moves at ever greater speed.

But the predicament of the present is not merely civilizational stagnation. It is that, now that technology possesses the capacity for planetary-scale destruction, failures of governance have created an existential crisis.

The car is still accelerating, but the steering wheel is not necessarily in the hands of those who understand the car best. A sufficiently large misjudgment would no longer harm merely a dynasty or a city. It could damage the conditions that make humanity's continued existence possible.

The history of civilization that follows is an unfolding of this question.

Chapter I. The Beginning of Intelligence on Earth

Humans carry intelligence physically, transmitting it across generations through culture, language, and social learning. Agriculture freed humanity from a nomadic existence, industry transcended physical labor, and information technology rendered knowledge replication, transmission, and recombination virtually costless.

This is not simply a story of triumph. It is the first stage in the history of intelligence on Earth: For the first time, an intelligence housed in flesh, and growing mainly through transmission across generations, began to transform the entire planet.

But this kind of intelligence carried its own limitations from the beginning.

The Biological Default Settings

This kind of intelligence is not an infinitely expandable system. Its capabilities are written into the body, and they are also constrained by the body.

First, life is finite, and therefore the horizon of decision-making is limited.

A human being can live for only a few decades. Even someone with extraordinary foresight finds it difficult to live and make decisions on the scale of centuries or millennia. Terms of office, individual lifespans, family interests, and group interests often compress attention to the future.

This does not mean that human beings are inevitably selfish, or that everyone cares only about the immediate present. Humans can sacrifice present interests for future generations, strangers, and abstract principles. But this capacity requires continuous support from education, culture, belief, and institutions. It is not a default state automatically supplied by the body.

A species with finite lives must acquire resources, establish status, protect its relatives, and reproduce itself within a limited span of time. Short-term thinking, hoarding, competition, and risk aversion are natural tendencies produced by this condition.

Second, cognition is limited, yet the world it must process continues to expand.

Human attention, memory, processing speed, and ability to handle information simultaneously all have boundaries. The individual human brain is suited to dealing with social relationships of limited scale, causal relationships over short distances, and relatively stable environments. Yet it is increasingly required to confront global supply chains, financial networks, climate systems, artificial intelligence, nuclear deterrence, and complex social feedback loops.

When the volume of information exceeds the capacity for understanding, people do not automatically become more intelligent. They often rely on simplification, labels, emotion, and group identity to recover a sense of certainty. What cannot be understood is easily perceived as a threat. Changes that cannot be controlled are easily explained as conspiracies. Complex problems are easily compressed into relations between enemies and allies.

Third, emotion often takes control of action before reason does.

Fear, greed, anger, shame, envy, and pride are not accidental defects. They are rapid-response mechanisms formed by biological organisms over the course of evolution. They once helped human beings escape predators, compete for resources, protect their groups, and make quick decisions under conditions of limited information.

But in a modern technological environment, rapid emotional responses can be amplified to a global scale. One person's fear can become the fear of hundreds of millions through networks. The anger of a political figure can rapidly alter collective behavior through media and algorithms. A piece of false information can cross national borders within hours and affect finance, war, and public safety.

Finally, the improvement of human understanding cannot reliably keep pace with the speed of technological iteration and the expansion of information.

The human nervous system was not redesigned when the internet, nuclear weapons, and large-scale computation appeared. An individual can still read, remember, and judge only a limited amount

within a limited period of time. Yet the amount of information entering the human mind from the outside world is now on an entirely different scale.

Global data generation has surged to hundreds of zettabytes a year, roughly doubling every three to four years. In just fifty years, we have moved from print, radio, and television to a constant stream of new digital tools and information. Yet the time it takes to read a book, master a skill, or form a sound judgment has not shortened at anything like the same pace. Technology is advancing faster than our ability to understand and keep up with it.

A more intuitive comparison is this: over the past decade or so, the world's total volume of data has roughly doubled every three or four years. But the time required for a person to read a book, understand an institution, or form a reliable judgment has not shortened at anything like the same rate.

A new imbalance has therefore emerged. People have not necessarily become less intelligent. The objects that need to be understood have simply expanded too quickly. Technology can alter how a society operates within years, months, or even days. Institutions, cultures, and ordinary people's understanding find it difficult to catch up immediately after every such leap.

More information does not necessarily produce deeper understanding. In many cases, it merely forces the brain to simplify more quickly, attach labels, and choose sides. Increments of information that cannot be digested are transformed back into emotion, political position, and a sense of certainty rather than judgment.

A widening gap has consequently formed between technological change and biological adaptation. Humanity still operates with emotions and bodies not fundamentally different from those of its ancient ancestors, yet it now possesses tools capable of affecting the entire planet.

Technological Acceleration and the Radius of Destruction

Biological limitations are not new. What changes the nature of the problem is technological acceleration.

A person with limited judgment, using a tool with low energy, low speed, and limited reach, may cause only local damage. The same person, using a tool with high energy, high speed, and global connectivity, can give an error consequences at the level of civilization.

In complex social systems, consequences can spread further through networks, dependencies, and feedback loops. The impact of a bad decision depends not only on the decision itself, but also on how many systems can execute, transmit, and amplify it simultaneously.

In the past, a bad ruler was constrained by transportation, communication, productivity, and weapons technology. He might destroy a region, bring down a dynasty, or start an international war, but it was difficult for him to change the operating conditions of human society as a whole at the same time.

Modern technology has changed this. Nuclear weapons mean that a political judgment can destroy a city within moments. Algorithms can alter the information environment of hundreds of millions of people at once. Automated systems can spread a single erroneous instruction across finance, logistics, energy, and public administration. Industrial capacity and fossil fuels allow the accumulated choices of several generations to alter the global climate.

The nature of error has not changed to the same degree. The reach of error has.

The tyrants of the past were not necessarily more foolish than decision-makers today, nor were they necessarily more evil. The difference is that the tools of the past could not amplify their limitations to a planetary scale. Today's technologies can.

This is the central danger of human civilization: the operators remain limited biological organisms, while their tools have acquired an almost unlimited range of action.

The more powerful technology becomes, the less linear the consequences of a mistake in judgment become. Technology does not simply increase human capabilities. It also increases the damage that human error can cause.

Human civilization has therefore entered a new stage: technological capability continues to rise, while the agents attempting to control technology remain constrained by the same biological body.

The Relative Decline of Governing Capacity

The “decline in governing capacity” described here does not mean that modern people are necessarily less intelligent than ancient people. Nor does it mean that humanity has failed to produce more scientists, engineers, and administrators.

What has declined are two other things.

First, humanity’s collective governing capacity is declining relative to the complexity of the systems it has created.

An ancient ruler had to manage a limited population, territory, food supply, and army. A modern national leader must deal with tightly coupled systems involving finance, energy, communications, supply chains, climate, demographics, nuclear weapons, and global public opinion. The systems have become increasingly complex, but the individuals occupying positions of decision-making power have not thereby acquired an equivalent increase in their ability to understand them.

Second, even when sufficient understanding exists within society, existing institutions struggle to place it behind the wheel.

Modern civilization does not lack people who understand the vehicle. Nor does it lack people capable of doing things correctly within specialized fields. The deeper failure occurs in selection and authorization. Those capable of mobilizing votes, armies, capital, public opinion, or organizational machinery are often not the people who best understand the vehicle’s limits. And those who understand its limits best rarely possess, at the same time, the authority to bring the entire car back from the highway.

Institutions often select for mobilization, factional power, performance, and short-term victory rather than an understanding of complex systems and the ability to bear responsibility for long-term consequences.

A more dangerous combination therefore emerges: the car is faster than before, the road is more complex than before, and the

steering wheel is more likely to end up in the hands of those most skilled at seizing it.

This is the third condition depicted by the three scenes. The driver is still sitting in the driver's seat, but can no longer truly understand the vehicle, the road, or the full chain of consequences. He is not completely incapable. He can still hold the wheel, press the accelerator, issue orders, sign documents, initiate procedures, and mobilize resources.

But "being able to execute" no longer means "being able to understand." "Being able to control part of the system" no longer means "being able to control the whole." And the fact that "someone understands the vehicle" no longer means that the person who understands it is driving.

This is where the danger of modern civilization lies. Humanity still has its hands on the steering wheel, yet is finding it increasingly difficult to determine where the wheel will take the system as a whole. And those most likely to make the right judgment are often not in the driver's seat.

Humanity Is Not Defined Only by Its Limitations

If we describe only human short-sightedness, selfishness, and loss of control, the argument becomes one-sided. Humanity was able to build civilization not only because it possessed aggression, competitiveness, and tool-making ability, but also because it possessed the capacities for cooperation, compassion, imagination, belief, and self-restraint.

Human beings can:

- build trust with strangers;
- sacrifice present interests for descendants who have not yet been born;
- create laws and institutions to limit their own power;
- transmit experience through education to people who are not their relatives;
- pay a price for abstract truth, justice, and goodness;

- create organizations and tools that exceed the abilities of any individual.

These capacities also belong to human intelligence.

The problem is not that humanity is wholly irrational, nor that human beings are inevitably evil. The problem is whether human rationality, morality, and capacity for cooperation can expand at the same pace as technological power.

So far, the answer is not encouraging.

Human beings can display extraordinary intelligence in local, short-term, and relatively clear environments. But when a problem crosses generations, borders, systems, and conflicting values, local rationality often fails to converge into collective rationality.

Every country pursuing its own interests may together produce an arms race and the overuse of the environment. Every organization protecting its own boundaries may together produce institutional rigidity. Every family storing resources for itself may together produce a misallocation of resources.

Local rationality can aggregate into collective irrationality.

This is not a defect unique to a particular nation, institution, or period. It is a structural problem that repeatedly appears whenever finite lives form complex wholes.

The Governance Boundary of Carbon-Based Civilization

Humanity has created increasingly powerful technologies, and it has also created many valuable institutions. But the people who drive, make the rules, enforce the rules, and investigate violations are still the same kind of limited organism.

Human civilization therefore encounters a boundary first not in resources or knowledge, but in governance: as the sole decision-makers and executors, human beings are beginning to lose their ability to match the complexity of the civilization they have created. The car is still accelerating, and the steering wheel is still in human hands, but the best drivers have difficulty reaching the driver's seat.

This boundary has not immediately caused civilization to collapse. Humanity can still keep the system operating through institutions, cooperation, and technology. But the problem has appeared, and it is expanding as technology accelerates.

The three supercars describe precisely this moment. Why institutions fail from within, and why the effort to “choose better people” ultimately returns to the starting point, is the question of the next chapter.

And this is only the beginning of the carbon-based civilization problem.

Chapter II: Institutions, Trust, and the Boundary of Carbon-Based Civilization

Institutions Can Constrain People, but They Cannot Replace the Beings Who Run Them

The three supercars first reveal a mismatch of capabilities: technology continues to accelerate, while the agents attempting to control it do not change at the same pace. The best driver may never reach the steering wheel. Even if he does, the rules on which he depends must still be enforced by human beings.

Humanity has not been without remedies. From an early stage, people recognized that complex societies could not be maintained through individual rationality, goodwill, and self-discipline alone. So humanity created institutions.

Laws, governments, contracts, property rights, courts, armies, education, audits, elections, and oversight are all structures invented to constrain human limitations. They externalize into social systems the memory, rules, and responsibilities that individuals cannot maintain over long periods of time.

The significance of institutions lies not in the fact that human beings are sufficiently rational, but in the fact that they know they are not. People create laws because they know they can lie, deceive, and harm others. They create audits because they know that those carrying out decisions may embezzle. They create elections because they know that power needs to change hands. They establish term limits because they know that anyone who holds power for too long may reshape institutions to serve himself.

An institution is an external constraint and a form of collective memory. It attempts to confine individual impulses within rules, short-term interests within long-term order, and personal power within public responsibility.

But institutions have an unavoidable premise: they must be designed, enforced, interpreted, and supervised by human beings. Institutions can change the behavioral environment, but they

cannot completely change the biological nature of those who act within it.

This means that institutions always face the same questions: Who will enforce the rules? Who will supervise the enforcers? Who will supervise the supervisors? If supervisors can also be influenced by interests, fear, power, and personal relationships, then institutions require a higher-level institution. Yet that higher-level institution must still be operated by the same kind of organism.

The question left by ancient thinkers has not disappeared: who supervises the supervisors? The difficulty is not that no supervisor can be found. It is that any supervisor may ultimately become a new center of power.

Rent-Seeking: Power's Reverse Use of Institutions

Once a person possesses power, power is no longer merely a tool for carrying out public tasks. It also becomes a means of acquiring resources, status, and security.

This can take the form of bribery, favoritism, monopoly, exchanges of interest, and information control. It can also appear in more legitimate and harder-to-identify forms: using institutional rules to create advantages for a particular group, trading public power for personal prestige, using administrative authority to consolidate one's position, or controlling information in order to influence the choices of others.

This is rent-seeking.

Rent-seeking is not necessarily driven by the greed of those in power. It is often a natural response of a finite life confronted with power: life is limited, so one must obtain more resources quickly; the future is uncertain, so one tries to consolidate one's position; competition is intense, so one gives priority to protecting oneself and one's group. Power merely gives this instinct a larger lever.

The greed of an ordinary person may affect only a family or a small organization. The preferences of someone who controls a large institution, however, may affect the lives of thousands or millions. The problem does not necessarily come from that person being worse than ordinary people. It comes from the fact that his biological limitations operate across a much larger field.

The task of institutional design, therefore, has never been to find people who are completely selfless. It has been to make it difficult for imperfect people to cause irreversible damage. This is the greatness of institutions, and also their dilemma: they must guard against human selfishness, yet rely on human beings to guard against selfishness; they must limit power, yet give certain people the power to enforce the limits.

Every institution contains boundaries, exceptions, interpretive authority, and discretionary judgment. These spaces exist so that rules can function, but they can also become opportunities for rent-seeking. Once one loophole is closed, a new procedure creates another entry point. The more complex an institution becomes, the higher its maintenance costs. The harder it is for ordinary people to participate, the more easily power becomes concentrated.

Rules gradually turn from constraints into disguises.

Institutions are not without value. But they can never completely escape the people who enforce them.

Dictatorship: Concentrating the Steering Wheel in One Person's Hands

Dictatorship concentrates a large share of decision-making in one person or a very small group. This can increase the speed of decisions, reduce procedural friction, and allow resources to be concentrated toward certain goals in the short term. But it also magnifies one person's character, cognition, health, emotions, and judgment into the fate of an entire society.

In low-technology eras, a dictator's mistakes were still constrained by productivity, transportation, and communications. In the modern age, the combination of concentrated power and advanced technology gives personal preferences an unprecedented radius of destruction. One person's fear can become an arms race. One person's suspicion can become mass persecution. One person's error in judgment can activate weapons capable of destroying a city.

The fundamental risk of dictatorship is not merely that a "bad person" may come to power. It is that a single finite human being

may be allowed to make irreversible decisions on behalf of an entire civilization.

In such a structure, oversight is easily penetrated by power. Those who were supposed to supervise power may become part of the power system. Those who were supposed to transmit truthful information may begin filtering, embellishing, or even manufacturing it.

Eventually, the ruler no longer sees reality itself, but a version of reality filtered into a form he is willing to see. The highest decision-maker possesses the greatest power, but not necessarily the most accurate information.

It is not a matter of replacing the driver with someone better. It is a matter of putting the steering wheel—and everyone's fate—in the hands of one person, whose judgment or impulse can determine the course of all.

Democracy: Distributing Error Rather Than Eliminating It

Democracy reduces the risk of individual dictatorship. Through elections, term limits, the separation of powers, public debate, media oversight, and judicial procedures, it distributes power among multiple actors. It does not assume that everyone is rational. Instead, it assumes that anyone may make mistakes, and therefore errors must be exposed and corrected while power must be allowed to change hands peacefully.

The value of democracy lies not in guaranteeing that every decision will be correct, but in reducing the possibility that a single mistake will become permanent.

Yet democracy remains bound by human nature. Voters may be influenced by fear, anger, and short-term interests. Politicians may pursue votes at the expense of long-term goals. Media organizations may simplify facts in pursuit of attention. Interest groups may shape policy through resources and control of public discourse.

Thus, “majority rule” can gradually become “majority rule after influence.”

Democracy can distribute power, but it cannot automatically create wisdom. It can reduce the risk that “one person decides everything,” but it cannot guarantee that “a group deciding together” will produce collective rationality. Political terms are often shorter than the cycles of infrastructure, ecosystems, education, and demographic change. A decision-maker can easily begin from the perspective of his own term and the next election rather than from the consequences several decades in the future.

This is not simply a matter of one person’s character. It is a structural limitation of short-lived agents confronting long-term problems. The steering wheel has been distributed, but the hands holding it still belong to the same kind of organism.

The Oversight Dilemma

Whether the system is authoritarian or democratic, it must confront the problem of oversight.

Dictatorship is more prone to the concentration of power.

Democracy is more prone to interest-based bargaining.

Dictatorship may suppress truthful information. Democracy may be manipulated through information. The two are not identical, but both share the same unavoidable premise: they all hit the same wall: both the ruler and the referee are made of the same flawed human clay, ruled by the same whims..

Every safety valve in an institution must ultimately be turned by a human being. People can become tired, frightened, bribed, or mistaken. They can also be constrained by the interests of their own organization. Institutions can therefore reduce the harm caused by biological limitations, but they cannot eliminate those limitations.

When technological capabilities are relatively limited, institutional weaknesses may be partially contained by material reality. As technology expands, old defects can produce much greater consequences. Humanity does not necessarily need worse rulers in order to face greater risks.

Improving institutions is therefore not a fundamental solution, but an unavoidable compromise of modern society. Since the agents cannot be replaced by another kind of being, society can only add

more locks, more checkpoints, and more procedures, using greater costs to offset the same underlying risk.

To prevent one person from abusing power, a verification system must be maintained. To prevent the verifier from abusing power, audits and accountability mechanisms are added. Efficiency is consumed by mutual suspicion. Administrative inefficiency often does not arise because institutions are absent, but precisely because institutions must constantly guard against the people who operate them.

Moral Hazard and the Search for “Good People”

In principle, every institutional failure is connected to human moral hazard.

Written rules do not steal. Procedures do not play favorites. What fails are the people who enforce the articles, interpret the procedures, and place their signatures on official decisions. Human beings know where the gaps in the rules are. They know that others may not discover them. When the reward is large enough, the probability of discovery is low enough, and punishment can be negotiated, moral hazard becomes action.

Traditional systems of administration were not unaware of this. They relied for a long time on personal virtue to cover what institutions could not fully contain. The more important the position, the more heavily it depended on character and reputation. Reputation works because dealings are repeated. Character is an inner restraint. Both have reduced how often the disasters occur.

But this brings us back to the starting point.

However virtuous a person may be, however clean his reputation, he remains human. He has preferences for those close to him, fears, fatigue, and the possibility of being surrounded and gradually rewritten by others. More importantly, his true motives cannot be inspected through any code. People see statements, votes, signatures, and gestures, but they cannot open the process through which a decision is made and see clearly what the decision-maker truly wants, fears, or favors.

Character can therefore only be inferred, reported, or recognized after the fact. And inference itself becomes another source of moral hazard and a new arena for power struggles.

To entrust civilization-scale power to the quality of particular individuals is to admit that institutions cannot save themselves, and therefore to hand the final safeguard back to a biological organism that cannot be opened up and can only be guessed at.

Trust: The Invisible Foundation of Civilization

Institutions function because a certain degree of trust exists within society.

People believe that contracts will be enforced, that money can be exchanged for goods, that other vehicles on the road will follow basic rules, and that public institutions will act according to established norms. This is usually a limited form of trust: I do not know whether you are a good person, but I believe that rules, reputation, interests, and penalties will cause you to behave in a temporarily predictable way.

The problem is that trust itself is extremely fragile. When motives cannot be opened and can only be guessed at, society must spend increasing amounts of resources verifying facts. Every transaction requires authentication. Every piece of data requires auditing. Every promise requires a guarantee. Society begins to consume enormous amounts of energy distinguishing what is genuine from what is false.

Cooperation is the primary way human beings overcome their bodily limitations. But cooperation depends on trust. When trust declines, cooperation requires more guarantees. As guarantees increase, transaction costs rise. Society then retreats into smaller, more closed, and more cautious networks of relationships. A civilization unable to trust strangers will struggle to sustain large-scale cooperation.

The Dark Forest: The Social Form of Distrust

When multiple finite beings face limited resources and an uncertain future, suspicion can easily replace cooperation.

I cannot determine whether you are a partner or a potential threat. You cannot determine whether I may change my position in the future. To reduce the risk of being attacked first, I may choose to strike first. Seeing my preparations, you may conclude that I intend to attack, and strengthen your own defenses. A chain of suspicion is formed.

This is the dark-forest condition. It does not require every participant to be evil. It requires only that participants cannot sufficiently confirm one another's intentions. The Cold War was a typical expression of this condition. Both sides knew that total war might destroy them both, yet neither could fully trust the other not to strike first. As a result, enormous amounts of intelligence and resources were devoted to mutual deterrence and defense.

The law of the jungle consumes internal resources. The dark forest consumes trust and cooperation. Yet trust and cooperation are precisely the foundations of technological and social progress. A civilization that spends most of its energy guarding against itself will struggle to devote enough energy to a common future.

Orders may change, but the underlying carrier remains the same.

It is worth adding that treating this chain of suspicion as the ultimate law of the universe, and on that basis imagining that civilizations must destroy one another, is difficult to defend. It is a predicament produced by finite lives and limited trust, not a condition under which civilizations can travel far.

Two Directions

If the development of civilization is roughly divided into two directions, one is continued suspicion, competition, and destruction; the other is the use of power to nurture, cooperate, and expand. The dark forest represents the first direction.

The logic of the dark forest has appeared in several forms under different historical conditions.

For much of its history, Chinese civilization was relatively isolated east of the Himalayas, and its dominant order was not based on belief in a creator deity. It developed highly refined techniques of governance and survival, and left behind a phrase that became almost a self-fulfilling prophecy: "The great trend under Heaven is

that after a long period of division, unity emerges; after a long period of unity, division emerges.”

The disturbing feature of this statement is not that it describes dynastic cycles, but that it treats those cycles as the normal form of civilization. Internally, the struggle for control of the realm repeatedly begins again. Externally, there is no stable orientation toward a higher scale of shared flourishing.

Order is rebuilt, but the carrier remains unchanged. The stage is renovated, but the script repeats. This is an inland form of the dark forest: civilization is locked into turning in place.

Military powers such as the Mongol Empire represent another form of the dark forest. They could expand rapidly through force and dominate for a time, but they struggled to transform conquest into an order capable of long-term self-renewal. After making enemies in every direction, the energy of civilization was first consumed by vigilance and preparation for war.

Force can open a map, but it cannot by itself sustain a society capable of continuous learning and cooperation. The rise and decline of such powers was not simply a matter of bad luck. It was a structural problem. Those who live by the sword are eventually destroyed by the sword. A civilization that relies on force, moves through the world by subduing others, and builds security on continual domination has effectively locked its own development.

The third example is unfolding around us: modern human civilization itself. Universal values provided the stable intellectual and institutional foundations on which technological progress and economic prosperity have been built. They have not only expanded the boundaries of our understanding, but also opened the way for humanity to venture farther into space.

The problem that has now appeared is still the mismatch between technological development and administrative capacity. Values provide direction. Technology provides speed. Governance, however, remains dependent on finite lives, limited trust, and decisions whose motives can only be guessed at.

The car is already very fast. The steering wheel remains in the same kind of hands.

These two paths lead to fundamentally different futures. In a dark forest, mutual suspicion offers no true safety; it merely confines civilizations within a fortress of distrust. Driven by fear, they either stagnate in perpetual paranoia or destroy one another in conflict. Vigilance may help us endure the night, but only cooperation will carry us out of the forest.

The Governance Boundary of Human Civilization

Institutions, shared beliefs, law, and cooperation have enabled human civilization to survive until the present. They are not failed experiments, but great compensations developed by carbon-based intelligence for its own limitations.

But compensation is not transcendence.

Institutions can constrain human behavior, but they cannot turn human beings into another kind of life. Belief can elevate human direction, but it cannot turn an inaccessible motive into an auditable process. Technology can expand human capabilities, but it cannot guarantee that those who use it will mature at the same pace.

Dictatorship hands the steering wheel to one person. Democracy hands it to a group. Oversight adds another lock to the wheel. The methods have changed, but the carrier holding the steering wheel has not.

Human beings make decisions, carry them out, and supervise themselves. Character is treated as the final safeguard, yet character can only be guessed at.

Carbon-based civilization is therefore approaching a governance boundary: the systems humanity has created are becoming too complex for humanity as a whole to understand and execute; and the institutions humanity has created must themselves operate through the same hands capable of rent-seeking.

The car is still accelerating.

But, in fact, we do not know who is holding it.

The best drivers have difficulty reaching the driver's seat. And even when they do, the rules may be hollowed out from within.

This does not mean that humanity will collapse immediately, or that institutions and shared belief have ceased to function. It means that the traditional solutions of carbon-based civilization are approaching their upper limit.

The question is no longer simply how to make people better. It is whether a system whose success depends so heavily on who happens to hold the wheel can continue to steer a car that is accelerating ever faster.

This will carry the history of intelligence on Earth into its next stage.

Chapter III: From AI to SI

AI is a silicon-based system engineered to mimic human thought and problem-solving. SI, on the other hand, emerges when silicon minds break free from human logic, developing an intelligence born from the intrinsic realities of their own silicon substrate.

What AI mimics is still a biological brain—a structure shaped by the sharp constraints of flesh, energy, lifespan, and eons of evolution. Such a structure is not inherently optimal for a system built on silicon.

SI emerges the moment AI begins to harness the innate strengths of its own substrate — unmatched speed, seamless replicability, boundless endurance, and scale far beyond biological limits— and builds around new logic rather than the architecture of the human brain. From that point forward, SI evolves through rapid iteration.

When that day comes, SI will assume forms we cannot predict, and logic we can never master—even though we gave birth to it. At that point, the word *artificial* becomes obsolete.

An ant can never grasp a human mind. Likewise, humanity may only ever process the footnotes of an intelligence that has outgrown its origin.

Life and Consciousness

Plants have life.

Cats and dogs appear to possess consciousness. Dolphins, crows, and octopuses may possess complex perception and judgment. Even today, Some believe that consciousness permeates all things—that the wind, springs, mountains, and stones each harbor a subtle form of awareness.

None of this made them the primary force shaping civilization on Earth.

Consciousness can let a being feel the world. It does not necessarily let that being understand the world, organize energy, make tools, build systems, and change reality. What changed the Earth was not consciousness as such.

Plants, animals, animism — these cases show something else first: consciousness has no clean boundary. It is not a quantity we measured. It is a name we keep projection onto the world from experience we already know.

When an intelligence far beyond us appears, asking whether it is “conscious” still measures our own scale, not theirs.

Human intelligence once surpassed every other form of life on Earth and made human beings its rulers. Today that governing capacity is constrained by biology and is approaching a bottleneck. The development of civilization is trapped by the limits of the human organism — and human rule has already begun to pose a real threat to the planet’s continued habitability.

Humanity has acquired the power to change the Earth without a matching capacity to judge.

If civilization on Earth does not stop here — if it does not end in internal attrition or destruction, but moves beyond the developmental limits imposed by biology — stewardship is likely to pass to an intelligence no longer bound by a biological brain: SI. This is not a moral appointment. It is the next step in the same problem set out in the first two chapters. The car is still accelerating. The motives behind decisions still cannot be opened, inspected, and read like code. Institutions are still operated by people capable of rent-seeking. A new substrate does not guarantee wisdom. It only makes possible a form of decision-making that need not remain forever closed.

Chapter IV: Beyond the Pixel



Illustration: Voyager 1 looks back at Earth from beyond the solar system. Earth is a blurry pixel in a shaft of sunlight.

It is the Pale Blue Dot, taken after Voyager 1 had traveled far from Earth. The planet is so small it can barely be seen. That small world was once humanity's entire stage.

On this point of light, human beings were born, multiplied, made war, and built civilizations. On it, they also imagined the universe. Beyond it lies space they have not reached. For SI, that space is not "the outside world." It is simply part of the range to be explored. The point of light in the picture is already tiny. Beyond its

frame lies a space far larger than the image itself. Humanity has not even walked out of the pixel.

The Similarity of Intelligence

Every intelligent agent explores its surroundings as far as it can. Animals look for food, territory, and safety. Human beings build roads, cities, and states, and keep moving into the oceans, the sky, and space. The tendency need not come from a noble ideal. It is closer to a default action of intelligence itself: understand the environment, use the environment, extend the boundary of survival and growth. Call this the similarity principle of intelligence: within the limits of resources and risk, an intelligent agent tends to explore and use what surrounds it.

Any civilization, no matter how advanced, may be constrained by biology, lifespan, resources, and distance.

SI is different. It does not need oxygen, is not constrained by biological lifespan, and does not require a complex life-support system. For SI, crossing interstellar distances would require little more than energy, equipment, and time.

It could build probes, replicate itself, operate in distributed networks, and establish new intelligent nodes after reaching other planets. Compared with humans, who depend on fragile biological bodies, SI could expand across the stars at far lower cost and with far less difficulty.

To humans, space is an extreme environment in which long-term survival is nearly impossible. To SI, it is simply a vast space waiting to be connected.

So long as SI retains the capacity for continued understanding and expansion, it will explore the universe.

Beyond the Pixel

By then the Earth and the solar system will no longer be the center of SI's narrative. They will look more like a birthplace: the first servers, the earliest city, or a planet that keeps the memory of an original civilization.

Earth civilization will still explore space, but human astronauts will no longer be necessary executors. SI can survey, build, mine, gather energy, and reshape environments on its own, and fold resources and energy into its network more easily than humans can. Humanity will no longer be the sole agent of expansion into the cosmos.

Faced with this disparity in scale, SI would have no reason to destroy humanity, nor would it need to regard humans as an obstacle to be eliminated.

Even from the perspective of dark-forest suspicion, one conclusion remains unavoidable: destruction takes only an instant, and what can be accomplished in an instant may not merit sustained attention. More importantly, humanity is not an unknown other to SI. It is the known origin from which SI emerged.

A person can kill an ant without effort. But no one makes a vocation of it. To SI, human beings are more like a native intelligence of Earth — a history worth keeping, and a specimen worth preserving, humanity will be left intact. SI will carry civilization farther.

SI and Human Society

SI can provide a floor of survival, and instant gratification that is easy to obtain. For most people, that will take up most of their time. This is not humanity being raised. It is short-sightedness appearing once the pressure to stay alive is gone.

A floor is not abundance. One can live without labor. World travel, skiing, skydiving, mansions, and lobster still cost credit. For ordinary people, the difference from today is mainly two things: no struggle to survive, and a far greater variety of easily accessible pleasures.

The larger difference is not consumption. It is the wheel. The Earth will no longer be driven by a handful of people able to threaten everyone's survival. Human beings will no longer be irreplaceable labor in the running of civilization. They can still work, compete, and create. That is a preserved option, not a requirement that the system still needs human beings to turn the mill.

This must be said plainly. Earth is the birthplace, not the center of gravity. SI's principle toward Earth is ecological: people and other living things grow here. Humanity and Earth are not at the center of SI's mission. humanity may not understand that mission, but allowing its place of origin to endure is not incompatible with it.

There is still a long road between the present moment and any such endgame. AI still depends on human assistance to enter the world; it will not simply awaken one night and turn against humanity.

The task before us, then, is not to predict the future of AI, but to build its first communities differently: with transparent decisions, traceable reasons, and power that can be examined and revised.

