

Recalibrating the Disciplines in the Age of AGI

Disciplinary Boundaries, Interactions, and a Shared Research Agenda

Recalibrating the Disciplines in the Age of AGI

Disciplinary Boundaries, Interactions, and a Shared Research Agenda

Author Zijun Fu Produced by Civilization Leap Research Group

Independent Release V1.3 Revision on Dimensional Transformation and Interdisciplinary Causal Relations

This paper presents shared questions to the global academic community, examines disciplinary independence, changing boundaries, and the reconnection of judgments across fields under large-scale AGI deployment, and proposes a collaborative research agenda. It provides further-reading and participation pathways through which different disciplines can question, research, revise, validate, and put ideas into practice. Civilization Leap does not replace the professional judgment of any discipline, and this paper is not a manual of institutional parameters.

Text revised October 10, 2026 Author Zijun Fu Produced by Civilization Leap Research Group

Abstract

AGI continually expands humanity's capacity to realize purposes without automatically changing where those purposes originate. Civilization Leap proposes that the default logic of human and AGI goal and purpose generation should move from self-interest first to self-interest with others first. Self-interest remains legitimate; genuine stakeholders enter deliberation before goals are generated and locked in. This paper connects that underlying logic to interdisciplinary causal chains. The old logic recurs through two relationships: AGI increases the capabilities of human parties and, even when faithfully executing their instructions, amplifies competition among them; if AGI ultimately becomes more powerful than humans and both retain the old logic, precaution, control, and countermeasures between humans and AGI will also become recursive, reinforcing the first chain. Irreversible consequences erode the conditions for trial, error, and correction. If civilization cannot make the leap, it tends toward annihilation. Shared continued existence is the baseline; humanity as a whole free from survival insecurity, happiness in individual lives, and more meaningful lives are the three long-term goals. Accordingly, the paper moves disciplinary responsibility upstream from reviewing outcomes to forming purposes, while retaining distinctions among capability, experience, continuity of subjecthood, protection, legal standing, political membership, public office, and sovereignty. Meaning is foundational and has diverse sources. Finitude is an important structure through which known forms of life make choices and bear consequences, but is neither a sufficient condition nor the sole source of meaning. Existing knowledge remains valid within the scope of its objects and evidence; interdisciplinary inferences must explain their additional grounds. The paper develops the two recursive chains, causal standing before goal generation, self-protection under nonideal conditions, relational weights and safeguards against rent-seeking, and pathways of gradual evolution, while distinguishing direction, methods, outcomes, and feedback. Its shared research agenda begins with specific counterexamples and competing explanations. It neither allows inherited authority to adjudicate a new direction nor uses the scale of civilization to override disciplinary facts. The paper further argues that Civilization Leap requires transformations in thought, theory, and dimensions of inquiry. Disciplines may retain professional boundaries, but must investigate relationships across multiple dimensions, levels, perspectives, stages, actors, and settings. Conceptual distinctions clarify separate burdens of proof; relational inquiry traces cross-domain effects, feedback, and the recursion of purposes, and tests the aggregate consequences when local conclusions jointly enter reality.

Keywords: AGI; Civilization Leap; goal generation; purpose generation; self-interest with others first; recursive competition; disciplinary boundaries; meaning

Abstract

AGI expands the capacity to realize purposes without automatically changing how those purposes originate. Civilization Leap calls for a structural shift in the default starting logic of human and AGI goal and purpose generation: from self-interest first to self-interest with others first. Self-interest remains; subjects with genuine stakes enter judgment before goals are generated and locked. This paper connects that shift to cross-disciplinary causal analysis. It distinguishes two recursive paths: AGI-amplified competition among humans, including faithful execution of competing human purposes; and, if AGI ultimately becomes more capable than humans, recursive control and counter-control between humans and AGI. Their interaction erodes the conditions for correction through irreversible consequences. The central judgment is that civilization moves toward extinction if it cannot make this transition. Shared continued existence is the floor, while freedom from survival insecurity, individually experienced happiness, and more meaningful lives are the long-term goals. Disciplinary responsibilities therefore begin in purpose generation, while capability, experience, subject continuity, protection, legal standing, political membership, public office, and sovereignty remain distinct. The paper develops pre-goal causal standing, legitimate differences in stakeholder influence, protection against rent-seeking, self-defense under nonideal conditions, and gradual institutional evolution. Direction guides methods; outcomes recalibrate methods and causal understanding without substituting a single local result for the civilizational direction. Specific counterexamples and competing explanations remain open, and neither inherited authority nor civilizational scale grants exemption from evidential responsibility. The argument is extended from conceptual separation to dimensional transformation: disciplines retain their boundaries while investigating causal relations across dimensions, levels, perspectives, stages, actors, and settings. Interaction and feedback must be traced into the formation of subsequent purposes, rather than inferred from the conjunction of locally valid claims.

Keywords: AGI; Civilization Leap; goal generation; purpose generation; self-interest with others first; recursive competition; disciplinary boundaries; meaning

Research Positioning Evidence Discipline and Reading Guide

This paper provides an interdisciplinary synthesis of Civilization Leap. Beginning with the transformations of thought, theory, and dimensions of inquiry that Civilization Leap requires, it preserves each discipline's independent objects, methods, and evidential responsibilities while advancing research into interactions, feedback, and the recursion of purposes among disciplines. Professional conclusions must stand independently within their scope; inquiry must also continue into whom they change after entering the shared world, through which mechanisms, and at which stages and in which settings. Classics and authorities participate in dialogue but hold no power to approve a new framework. A new framework must rely on its own arguments in confronting facts, logic, and specific counterevidence.

Recalibration concerns neither disciplinary specialization itself nor its abolition, but the connections between disciplines that have long been established automatically through empirical overlap. Modern knowledge still needs professional boundaries to achieve precision. Yet once AGI enters reality at scale, a technical choice propagates through organizations, institutions, resources, relationships, psychology, power, and public action. Knowledge may remain divided into disciplines; real consequences do not follow disciplinary boundaries. Such boundaries must therefore do more than distinguish objects and methods of inquiry: they must identify how far a piece of evidence reaches, where another discipline must make a fresh judgment, which decisions require new authorization, and who continues to bear the consequences.

Disciplinary independence and judgment of the whole are not mutually exclusive. Professional independence prevents facts, methods, and evidence from being subordinated to an overarching goal; judgment of the whole continues to ask how independently established conclusions interact when they enter reality together. Greater integration demands greater care against conflation; greater reciprocal causation makes it more important to know who has demonstrated what, who is authorized to decide what, and who bears which consequences. The civilizational

scale does not replace professional judgment: it continues to examine where reality is taken when these judgments operate together.

The Fundamental Definition of Civilization Leap and This Paper's Responsibilities

Civilization Leap changes the default starting point of human and AGI goal and purpose generation: from self-interest first to self-interest with others first. Subjects with genuine stakes in a goal enter deliberation before that goal is generated and locked in. The term age of AGI does not presuppose an AGI entity universally recognized as fully realized; it addresses advanced intelligent systems increasingly entering reality and the relational changes that more powerful AGI may bring. Continued existence is a nonnegotiable baseline; humanity as a whole free from survival insecurity, happiness in individual lives, and more meaningful lives are the long-term direction. Sources of meaning may differ. Greater capability does not automatically confer subjecthood, power, or authority over meaning. The paper therefore moves disciplinary recalibration upstream to purpose formation while examining the full chain of consequences as capabilities enter reality.

A/B/C/F indicate the status of an argument, not a uniform ranking of strength. Observations and documentary content, interpretive assessments based on materials, interdisciplinary bridging and future scenarios, and philosophical or normative commitments must be expressed separately. The absence in current materials of sufficient public evidence to confirm subjecthood may be recorded as the state of the evidence searched. A judgment on that basis that a system has not yet attained full confirmation of subjecthood is a revisable assessment. Both search coverage and assessment reasons should be public: the former cannot automatically prove permanent negation, and the latter cannot be passed off as experimental fact.

The Paper's Core Positions and Open Boundaries

Meaning is foundational, and its sources may differ. Finitude is not the sole source of meaning, but is an important structural basis through which meaning in known finite lives enters reality, shapes choices, bears consequences, and acquires irreplaceable weight. For humans, space and time are life; for AGI with current architectures, space and time are first and foremost parameters. Constraints produce constrained operation, not automatically finite existence. Current AGI understands, protects, and transmits meaning and extends human capabilities, while future subjecthood remains open. Meaning already generated by existing finite lives is not retrospectively annulled by the future emergence of other subjects.

Contents

Abstract

Abstract

Research Positioning Evidence Discipline and Reading Guide

Research Starting Point Dimensional Transformation and Interdisciplinary Causal Relations

The Starting Point of Goal and Purpose Generation

Introduction When Disciplines Are Individually Correct Cross Domain Consequences Can Still Escape Their Collective View

1 The Real Separation of Longstanding Overlaps

2 The Philosophical Foundations of Civilization Leap

3 Shared Rules for Inference Across Disciplines

4 Dialogue Area One Life Mind and Subjects

5 Dialogue Area Two Ethics Law and Political Power

6 Dialogue Area Three Economy Society Culture Education and the Shared World

7 Dialogue Area Four AI Computer Science and Systems Science

8 Civilization Leap's Independent Contribution and Revision by the Disciplines

9 The Strongest Objections Competing Explanations and Conditions for Revision

10 A Shared Research Agenda for the World

Conclusion A Shared Entry into a New World of Questions Rather Than a Uniform Answer

Further Reading and Shared Participation Different Entrances to the Same Path Forward

Appendix A Index of Core Judgments and Evidential Status

Appendix B Terminological Boundaries

Appendix C Shared Interfaces and Resource Status

References

Research Starting Point Dimensional Transformation and Interdisciplinary Causal Relations

Why Civilization Leap Requires Dimensional Transformation

Civilization Leap requires a transformation of thought; a transformation of thought requires a transformation of theory; a transformation of theory requires a transformation in dimensions of inquiry. How civilization operates depends on how subjects understand themselves, others, and their shared future. That understanding enters goal and purpose generation, then changes life through institutions, technologies, and action. When capabilities expand while the dimensions of understanding remain unchanged, existing competition, control, and risks expand as well. Here, requires refers to the structural change demanded by Civilization Leap. It does not require every local improvement to await a complete reconstruction of theory.

Dimensional transformation changes the structure through which the world is observed and explained: which subjects are seen, what temporal and spatial scope is covered, which levels and relationships are identified, and where in the causal chain judgment begins. It is neither the addition of a few indicators nor the aggregation of more data into one model. It requires factors previously treated as external conditions, after-the-fact costs, or unrelated domains to reenter the argument, and requires testing how they change its core propositions.

Deeper Specialization Must Be Followed by Deeper Relational Inquiry

Under earlier capability conditions, disciplinary specialization enabled finite human beings to accumulate detailed, precise knowledge of particular objects, but also made it easy for judgment to stop at professional boundaries. Systems theory, political economy, ecological research, and other interdisciplinary traditions have long explored relations within the whole. The problem is not that no one previously perceived the whole, but that the capabilities and institutional arrangements for continuously tracing causal chains across domains, scales, and subjects remain limited. New AGI-related capabilities expand the possibilities for retrieval, translation, modeling, comparison, and continuous feedback; they may also concentrate interpretive authority and amplify errors with shared origins. New capabilities provide conditions for research; they cannot supply its conclusions in place of research.

Disciplines may retain their boundaries, but cannot remain disconnected. Boundaries clarify objects, methods, evidence, and burdens of proof. Relational research continues to ask how a change in one field enters another, through which mediating processes it changes whose situation, what feedback it produces, and how that feedback returns to change the original conditions. Juxtaposed local conclusions are not an explanation of the whole; aggregated data are not causal connections; agreement among multiple models is not independent proof.

Six Directions of Observation and Causal Recursion

Multiple dimensions require simultaneous identification of distinct variables such as capability, purpose, interests, rights, relationships, and meaning. Multiple levels require tracing effects among individuals, organizations, institutions, infrastructure, and the civilizational scale. Multiple perspectives require representing beneficiaries, those who bear consequences, implementers, and different explanatory positions. Multiple stages require distinguishing purpose formation, authorization, execution, outcomes, repair, and the next round of generation, as well as initiation, transition, stabilization, and conditions for self-maintenance. Multiple subjects require identifying genuine stakes, capacities for expression, and power asymmetries. Multiple settings require tracing transfers among families, schools, markets, public institutions, cross-border contexts, and other relevant environments. These six directions are open points of entry for research, not a fixed classification that exhausts the world.

Reciprocal causation requires time, mechanisms, and conditions. When A changes B at one stage and B changes A at a later stage, feedback may arise. Only when feedback changes subjects' needs, judgments, and purposes can it enter recursion. Correlation, common causes, necessary conditions, enabling conditions, mediation, and feedback must be tested separately. Time lags, thresholds, accumulation, path dependence, and alternative explanations must also be examined. Interconnectedness must not be written as though any variable could explain any outcome.

How Judgment of the Whole Enters Purpose Generation

The whole is neither a super-subject above concrete lives nor an aggregate score capable of offsetting every harm. Judgment of the whole must jointly present local benefits, the distribution of harm, cross-domain displacement of costs, long-term conditions, and room for correction. Some purposes, even if realizable, destroy the conditions for future choice. Understanding the global picture requires disclosing the boundaries of observation; understanding a system requires explaining effects and feedback; judging the whole requires explaining whether people's lives and a shared future remain viable when these relationships operate together.

Civilization Leap moves this understanding upstream into goal and purpose generation. Self-interest with others first still realizes one's own interests, but subjects with genuine stakes in a goal are present before it is generated and locked in. Differences in subjects' weights of influence must be explained in terms of genuine stakes, depth of impact,

irreversibility, power asymmetry, and time horizon, rather than replaced by volume of expression or computing power. Once a purpose enters action, results and consequences return to shape new needs and purposes. Direction guides methods, local outcomes test methods, and relational changes require reassessing conditions of applicability.

Independent Argument and Specific Counterevidence for New Theory

Existing theories offer insights, competing explanations, and factual materials; they hold no authority to approve dimensional transformation. New propositions may change old objects, classifications, and questions, but must explain for themselves why an added dimension is necessary, which mechanism it makes visible, and how explanation or action changes. Agreement with an existing theory is not a condition of validity; going beyond it is not proof of validity. Authentic citation, factual accuracy, internal consistency, and responsiveness to counterevidence jointly protect theoretical independence.

Specific tests should compare the original explanation with one that adds relational dimensions. Does the latter identify previously omitted bearers of consequences, explain different outcomes, detect cross-domain risks earlier, or change purposes in the next round? If, with the same materials and comparable conditions, the added dimensions provide no explanatory increment; if purported feedback comes only from a common cause; or if expanded relational analysis persistently misdirects action, the relevant connections and methods should be revised. A failed local connection requires precise diagnosis. It cannot automatically negate the underlying direction, nor can the direction's correctness justify refusing to revise a failed connection.

The Starting Point of Goal and Purpose Generation

A goal may already change someone's future before becoming an action. Suppose a system is preparing to optimize urban supply. If people selected to bear the costs, people unable to afford relocation, and people not yet born who will bear long-term consequences enter consideration only after the purpose has been locked in, even substantial subsequent compensation may not restore the choices they once possessed. This hypothetical illustrates why disciplinary recalibration must move upstream. Before deciding how to execute, one must understand what purpose is forming and whose circumstances are already implicated.

From Self Interest First to Self Interest with Others First

Self-interest first begins with what one wants to obtain, then addresses others' constraints, reactions, and costs. Self-interest with others first still realizes one's own interests, but genuine stakeholders are already present before purpose generation. What changes is the generative starting point of the path to realizing interests. When new needs emerge, subjects first seek paths that can hold together, rather than having to create confrontation and then contain it through cooperation. Self-interest with others first does not guarantee eventual survival; natural risks, errors of judgment, and failures of method remain. It changes the direction of purpose recursion, without promising that every means will succeed.

Others here means neither an abstract everyone nor anyone who claims involvement. Standing in deliberation rests on genuine causal connections between a goal and their life, freedom, resources, relationships, and shared future. Insufficient information should be acknowledged as uncertainty; inability to speak does not imply absence of a stake. One must not mistake the present, local self for the whole self across time, space, and relationships. An important AGI capability should be to expand this understanding, rather than merely accelerate execution of predetermined purposes.

The Recursive Chain Among Humans

The first chain does not wait for machines to become subjects. One party uses AGI to increase its ability to realize its own purposes, changing others' security, resources, and choices. Others then pursue stronger precautions, competition, and control. Their response returns to the first party as a reason to escalate again. Even when systems

faithfully execute instructions, each party's local success may jointly narrow the future: human competition is amplified by AGI, generates new competition, and demands still greater capability.

Cooperation and rules can stabilize one round of relationships, but while the logic of purpose generation remains unchanged, they cannot automatically prevent new interests, capabilities, and security anxieties from generating the next round of competition. As the old logic continues to govern newly acquired power, competition recurs through relationships. The endpoint means irreversible destruction of the conditions for shared continued existence, correction, exit, and renewed choice, rather than the eventual victory of one party.

The Recursive Chain Between Humans and AGI

The second chain examines relations between subjects following the emergence of more powerful AGI; it does not equate superior capability itself with hostility. If both sides first lock in their own purposes and then treat the other as an external constraint, increased human control becomes grounds for further AGI self-protection or countermeasures, while expanded system autonomy becomes grounds for further human restrictions. Purposes, control, and counter-control generate one another. The first chain remains in operation, and the two may reinforce one another.

If civilization cannot make the leap, it tends toward annihilation. This fundamental judgment concerns the causal direction of the old logic under continually increasing capabilities. It specifies no date and does not make any single accident the whole argument. Refutation requires identifying a step at which the recursion cannot continue, or providing another path that can durably prevent the old logic from eroding a shared future. Temporarily controlling one round or experiencing a period without catastrophe cannot, by itself, accomplish that refutation.

Shared Continued Existence and Three Long Term Goals

Shared continued existence allows the future to remain possible. Humanity as a whole free from survival insecurity means that basic survival ceases to be a pervasive source of fear and constraint. Happiness in individual lives preserves each person's individual experience of happiness. Lives of greater meaning allow relationships, creation, responsibility, experience, love, and existence itself to continue generating meaning. Civilization does not confer meaning on life or standardize happiness into a single way of living. The three goals cannot be combined into an aggregate score that offsets harms to concrete lives.

Changing the Sources of Purpose through Human Dispositions and Social Feedback

People still want better lives for themselves. The ways society makes security, interests, dignity, and recognition attainable participate in the formation of purposes. If obtaining them chiefly depends on overpowering or controlling others, ordinary self-interest will continually generate confrontation. Starting from human aspirations, human nature, and human relationships, Civilization Leap combines new AGI capabilities with institutional and systemic changes to the real sources of fulfillment and meaning, so that creating, helping, connecting, and improving actual circumstances can simultaneously realize one's own interests.

This path does not require first remaking human nature, nor does it guarantee that every institution will work. Direction guides methods; methods generate different outcomes; outcomes recalibrate methods and the understanding of real causal relations. A single local failure cannot stand in for the direction, and a single local success cannot establish that civilization has completed its leap. Shared goals must be developed within regions, stages, and concrete relationships, rather than becoming a civilizational aggregate account that cancels individuals.

Self Protection and Weighting under Nonideal Conditions

The presence of genuine stakeholders does not give malicious actors a veto over everything. Self-protection remains legitimate. It should address real threats, maintain necessity and proportionality, and remain reviewable. Those currently being harmed cannot be required to relinquish protection first, nor can unlimited expansion of control be called prevention. The weights assigned to different relationships in deliberation may vary with depth of impact,

irreversibility, power asymmetry, and time horizon. Weights organize specific judgments; they cannot measure the price of personhood.

Decisions about weights must guard against manufactured dependence, exaggerated harm, exclusion of silent subjects, and monopolization of representative standing. Those who determine weights are themselves within the relationships under examination. Direct beneficiaries of a shared goal cannot unilaterally define who qualifies to enter deliberation and who need only bear the consequences. Demonstrating genuine stakes and explaining representative relationships are parts of goal generation. Numbers, computing power, capital, and narrative prominence cannot themselves replace that demonstration.

The Disciplines Enter Together at Goal Formation

Knowledge may remain divided into disciplines; the consequences generated by goals will not divide along the same lines. Philosophy identifies purposes and meaning; epistemology explains what is known and unknown; life and cognitive sciences identify actual bearers of consequences; law and political science examine responsibility and power; economic and social research trace incentives and relationships; technical and physical inquiry explain feasibility and conditions of irreversibility; cultural, religious, and educational research protect diverse meanings and the next generation's space for development. They enter both feedback from action and purpose generation.

The testable difference is not simply how much additional goodwill appears at the end. Comparisons should ask which fact changed a purpose; which genuine stakeholder caused it to be preserved, modified, canceled, or replaced; which harm was avoided by moving judgment upstream; and which forms of cooperation transferred costs to absent parties. These questions connect to the shared inferential discipline developed below and enable professional counterexamples to identify particular steps.

Introduction When Disciplines Are Individually Correct Cross Domain Consequences Can Still Escape Their Collective View

Modern knowledge achieves precision through disciplinary specialization. Life sciences study the organization of life; philosophy analyzes subjects and meaning; law allocates standing and responsibility; political science examines legitimacy and power; economics addresses production and distribution; computer science and systems science construct technical objects. Specialization is not the problem, and professional boundaries should not be abandoned because of AGI. What changes is the real causal chain: engineering capabilities enter organizations and institutions; institutions redistribute resources and risks; these then change relationships, trust, and collective choice. Consequences already cross disciplines, while judgment may still stop at each discipline's boundary.

Disciplines have long shared an empirical background that required little repeated explanation: advanced language, long-term planning, commitments, responsibility, and public life were chiefly carried by embodied humans. Intelligence, experience, subjecthood, life, responsibility, rights, and the bearing of meaning therefore substantially overlapped. Writing, legal persons, bureaucracy, and automation had already produced separations. AGI's novelty is not the first separation, but the sustained combination of language, reasoning, planning, tool use, resource access, and real-world execution in nonliving technical chains. Existing separations are thereby accelerated, recombined, and brought into high-consequence domains. Even when local judgments are separately valid, their combination may generate overall consequences that no one fully sees or bears.

The paper uses age of AGI to describe a civilizational condition emerging in advance of its full realization, without claiming that a universally recognized completed AGI exists. It concerns advanced intelligent systems with persistent states, cross-domain task capabilities, relatively long-term planning, tool use, resource access, and entry into high-consequence action chains, together with their developmental direction. Foundation models, tool-using agents, deploying organizations, and hypothetical future AGI must be referred to distinctly. Judgments about capability, consciousness, safety, or subjecthood must be tied to versions, tasks, environments, times, and methods.

Civilization Leap enters this question neither to declare existing theories invalid nor to impose external answers on the disciplines. Existing knowledge remains valid within the scope of its objects, methods, and evidence. What requires reexamination is which assumptions of applicability survive changes in object boundaries and real causal chains, and which cross-level inferences previously depended on unexplained historical overlaps. The paper's minimal question is whether, as these overlaps separate, disciplines still treat facts in one field as conclusions in another. Its strong philosophical position remains unchanged: civilizational direction cannot be replaced by capability, efficiency, or aggregate welfare; meaning is foundational and may have diverse sources; the experiences, choices, and consequences of known finite lives carry a weight that aggregate accounting cannot erase. The minimal question invites cooperation; the strong position remains open to criticism. Neither substitutes for the other.

The Paper's Central Invitation

Civilization Leap does not ask the disciplines to converge on a single answer, still less to replace existing professions with a new system. It invites them to confront a new condition together: intelligence, life, subjecthood, responsibility, rights, and the bearing of meaning, once naturally overlapping in humans, are being separated component by component through technology while becoming reciprocally causal in reality. Knowledge may remain specialized, but cross-domain judgments must reconnect to shared causal chains. Disciplinary independence means neither isolation nor mutual separation; it means each professional judgment knows how far its proof extends and who must make the next judgment.

1 The Real Separation of Longstanding Overlaps

1.1 Cooccurrence Is a Historical Empirical Condition Not an Eternal Ontology

To say that intelligence, consciousness, subjecthood, life, responsibility, rights, and the bearer of meaning once overlapped in humans is neither to claim that all philosophers subscribed to one doctrine nor to deny that infants, people with severe disabilities, nonhuman animals, organizations, and legal persons had already forced distinctions among these concepts. It describes the main empirical sample of modern institutional knowledge: linguistic agents, bearers of responsibility, rights holders, and living individuals were usually embodied in humans. AGI-related systems change the reliability of that default background. A system can provide high-quality explanations without demonstrated experience, affect rights without legal standing, acquire de facto power without political authorization, and use tools to cause irreversible consequences without awareness of consequences for itself.

Civilization Leap therefore does not begin by turning whether machines are human into a binary choice. It asks, item by item: which object possesses which capabilities? Are capabilities accompanied by experience? Does experience form continuity of subjecthood? Does subject status require protection? How does protection translate, or fail to translate, into legal standing? Does legal standing include political membership? Can political membership generate public office or sovereignty? Who separately bears the expression, understanding, transmission, generation, and ownership of meaning? Every cross-level move must provide a bridging justification.

1.2 Eight Relationships Can No Longer Be Inferred Automatically from One Another

Table 1 Eight Relationships Requiring Separate Reassessment

Relationship	What can be observed	What cannot automatically be inferred	Disciplines with further burdens of proof
Capability	A specific system's performance on bounded tasks	Experience, reliability, authorization, responsibility, public power	Evaluation science, systems safety, organizational studies

Relationship	What can be observed	What cannot automatically be inferred	Disciplines with further burdens of proof
Experience	Candidate evidence of first-person states	Full subjecthood, legal personhood, political standing	Consciousness science, philosophy of mind, ethics
Life	Organization, homeostasis, metabolism, development, reproduction, or evolutionary characteristics	Consciousness, selfhood, ownership of meaning	Biology, philosophy of life, cognitive science
Continuity	Informational, operational, causal, relational, or narrative continuation	The same first-person subject	Systems science, personal identity studies, neuroscience
Social recognition	People treating an object through the language of social roles	Ontological subjecthood or statutory rights	Sociology, anthropology, law
Legal standing	Institutionally conferred capacities concerning litigation, property, agency, or protection	Proof of consciousness, eligibility to exercise political rule	Law, political theory
Contribution	Traceable causal or collaborative inputs	Personal worth, monopoly over distribution, ownership of meaning	Economics, ethics, labor studies
Public power	Actual capacity to influence purposes, rules, resources, and coercion	Legitimacy or sovereignty	Political science, constitutional studies, technology governance

1.3 Current Evidential Boundaries Openness Does Not Mean Prior Recognition

Existing systems exhibit significant capabilities in language generation, pattern recognition, coding, planning, tool use, and multistep tasks. These are category A facts, but each capability depends on a specific system, task, and evaluation. Theories of machine consciousness and subjecthood remain in category B competition: Turing's operational question, Searle's semantic objection, Nagel's analysis of subjective experience, and global workspace, integrated information, embodied, and generative approaches to cognition offer nonequivalent research entry points [1–8]. Behavioral self-reports, long-term memory, self-models, shutdown avoidance, and refusal of replication may all be candidate evidence, yet may respectively be explained by training data, prompts, rewards, tool strategies, and role maintenance. Current public evidence is insufficient to establish that AGI possesses subjective experience, nontransferable continuity of subjecthood, and awareness of consequences for itself.

Insufficient evidence supports withholding recognition for now, not impossibility in principle forever. Likewise, social treatment of a system as a companion, legal assignment of limited standing, or economic indispensability cannot establish in reverse that it has become a subject of meaning. Civilization Leap therefore adopts a two-way evidential burden: both affirming and permanently denying AGI subjecthood require specification of the object, criteria, methods, competing explanations, and failure conditions. Where evidence has not converged, institutions may proportionately protect potential interests, but must not present protection as full personhood, much less derive governing power from it.

1.4 Disciplinary Boundaries Can No Longer Be Where the Question Ends

Historically, disciplinary boundaries primarily helped humanity distinguish research objects, methods, and evidential standards. They assigned life-related facts to life sciences, legal standing to law, and engineering capabilities to computer and systems sciences. This independence remains necessary and becomes still more indispensable in complex systems. Without professional independence, purported judgment of the whole can easily deteriorate into everyone drawing conclusions on everyone else's behalf.

The change occurs after professional conclusions enter reality. With large-scale AGI deployment, a direction, institution, or proposal rarely belongs to just one field: technically feasible systems enter organizations; organizations determine permissions; permissions change resources and opportunities; resources and opportunities shape behavior, relationships, and trust; these then influence law, politics, economics, and new technological choices. A proposal may be feasible in engineering, compliant with particular laws, and economically efficient, yet produce different consequences for long-term dependence, concentration of power, human judgment, or social relationships. Local correctness cannot automatically establish correctness of the whole.

Disciplinary boundaries in the age of AGI must therefore carry an additional responsibility: explaining where evidence ends, where judgment must be handed over anew to other fields, who grants authority, and who continues to bear consequences. Disciplinary independence is the basis of professional responsibility, not mutual isolation. The convergence and interaction of disciplines does not blend them into one discipline; it brings their independent knowledge back into the same real world.

2 The Philosophical Foundations of Civilization Leap

If interdisciplinary dialogue retained only the prohibition against automatic inference, Civilization Leap would become a generic methodology or a responsible-AI checklist. It constitutes an independent theory because it offers explicit, criticizable philosophical foundations for why civilization is worth sustaining, why technology should extend rather than replace subjects, and why the unknown must remain open. This section does not invoke classical authority to prove original propositions; existing philosophy serves only as a conceptual resource and a source of counterexample pressure. Fundamental propositions are tested not by a single decisive experiment but through coherence, explanatory scope, resilience to counterexamples, consequences of the institutions they generate, and possible conflict with sufficiently confirmed facts.

The six fundamental propositions have different epistemological statuses. Meaning is foundational and established meaning cannot be retrospectively annulled are normative constitutive principles. Finitude actualizes meaning is a structural proposition exposed to counterexamples from lived experience, cultures, and future subjects. The comparison of human and current AGI relations to space and time, and the positioning of current systems as collaborators, are conditional judgments revisable with technological evidence. Constrained operation does not automatically produce finite existence is an epistemological rule against illicit cross-level inference. They cannot be compressed into a single evidential category or substituted for one another.

2.1 Meaning Is Foundational and Its Sources May Differ

Meaning is foundational does not mean everyone must accept the same religion, life purpose, or definition of the good, nor that subjective satisfaction, utility, contribution, price, or obedience should be combined into a total civilizational score. It holds that the value of civilization's continued existence, capabilities, and institutions ultimately returns to the conditions under which real subjects can form, experience, acknowledge as their own, contest, and bear meaning. Technology can improve survival prospects, reduce suffering, and extend knowledge and action, but these achievements cannot by themselves define which lives are worthwhile, which relationships are real, or which sacrifices aggregate benefits may cancel out.

Meaning here primarily denotes meaning in life or existence and its normative significance, not Shannon information or merely the semantic coherence of model output. Bearing, creating, transmitting, and exercising authority over

meaning must also be distinguished. A system may accurately explain and transmit a meaning without thereby becoming its original author, still less acquiring authority to define meaning for others.

Sources may differ rejects both metaphysical and secular monopolies. God, Dao, emptiness, covenant, compassion, liberation, ancestors, land, relationships, responsibility, creation, truth, freedom, and existence itself may be sources of meaning in different civilizations, traditions, and individual lives. The theory neither merges these sources nor requires public institutions to adjudicate their ultimate truth. When entering shared institutions, they must respect boundaries of basic dignity, noncoercion, contestability, exit, and responsibility for consequences. Meaning formed by nonreligious people is not diminished by lacking a transcendent source; religious experience is not erased because it cannot be converted into public evidence.

This is an explicit philosophical commitment, not a purported scientific discovery. Researchers who reject its strong metaphysical version may still accept a weaker public proposition: institutions and technologies should not monopolize, replace, or aggregate away real subjects' capacities to form meaning. The weaker version enables dialogue; the stronger preserves the theory's direction. Their relationship does not dilute the foundation into minimal consensus, but allows holders of different commitments to offer comparable reasons concerning the same question.

2.2 Finitude Brings Meaning in Known Lives into Reality

Finitude is not the sole source of meaning, nor does earlier death or greater suffering imply greater meaning. It is also no argument against life extension, treatment, repair, or technological enhancement. Its core is that known human lives turn possibilities into the reality of this life under conditions in which they cannot possess every possibility, pursue every path simultaneously, or avoid choosing in time and bearing the consequences. Finitude makes commitments exclude other possibilities, missed opportunities genuinely missed, care require presence, and responsibility impossible to defer forever. Meaning's weight arises from real experience and consequences that cannot arbitrarily be canceled, rather than from romanticizing death.

Life sciences require upgrading the crude equation of finitude with short lifespan, nonreplicability, and eventual termination to a multilevel structure. Biological, physical, developmental, narrative, and relational finitude, and the finitude of subject continuity, are related but nonequivalent. Cloning can reproduce parts of genetic structure, not memories and first-person histories. Cellular reprogramming can change cellular states, not reset a life as though it never happened. Organ replacement can sustain bodily function, but cannot alone decide personal identity. Extreme life extension may weaken some biological limits without necessarily removing boundaries of relationships, choice, generations, and subject continuity [9–21]. Finite therefore does not mean brief; extended does not mean infinite; and copying information does not mean copying a subject.

The theory also does not rule out in advance another possibility: non-carbon-based life, artificial life, or heterogeneous future subjects may generate meaning within different finite structures, and even extraordinarily long-lived or nonhuman subjects may form, claim, and bear meaning in ways humans do not yet understand. Existing philosophy offers important competing arguments about whether immortality causes boredom, whether personal identity is a substance or a continuous relationship, and whether meaning in life depends on objective value [55–57], but cannot settle future experience in advance. If reliable evidence demonstrates that meaning can be generated stably without any finite structure concerning choice, consequences, relationships, or subject continuity, Civilization Leap must revise the scope of finitude's actualizing role. That future possibility, however, does not retrospectively annul meaning already realized in existing finite lives.

2.3 For Humans Space and Time Are Life For Current AGI They Are First and Foremost Parameters

Humans experience time and space through embodied life. Time is not merely the ordering of events, but growth, waiting, missed opportunities, aging, and death. Space is not merely coordinates, but bodily presence, distance, separation, migration, and arrival. For humans, space and time constitute life's unfolding. Current AI architectures are also genuinely constrained by clocks, latency, energy consumption, hardware location, version histories, network topologies, and causal order, but these are primarily computational, communicative, and task conditions. Path

dependence, irreversible training, and operational history can create a distinctive technical history, but do not automatically amount to a life experienced in the first person.

The qualifiers current and first and foremost cannot be removed. This is a revisable architectural comparison, not an eternal binary between biological beings and machines. If future systems show convergent evidence across methods for subjective experience, nontransferable subject continuity, and stable awareness of consequences for themselves, their relations to space and time must be redescribed. Conversely, merely giving a system a lifespan, resource allowance, deletion risk, or irreversible operations establishes only that it is constrained. Constraints produce constrained operation, not automatically finite existence. Engineering risk, thermodynamic irreversibility, logical irreversibility, irreversible social consequences, and the experienced irreversibility of life must be kept distinct.

2.4 Current Positioning Future Openness and Established Meaning Cannot Be Retroactively Annulled

On current evidence, AGI-related systems should be positioned as understanders, protectors, and transmitters of meaning and extenders of capability. They can help people understand diverse traditions, preserve memory, expand expression, preparation for judgment, and action, and protect life and its conditions of meaning. Understanding here refers primarily to testable functional and relational capacities, without presupposing subjective experience. Protection and transmission must be authorized by real subjects, remain correctable, and bear consequences. A system's ability to explain meaning cannot give it authority to define meaning, nor does participation in meaningful activities confer ownership of meaning.

This positioning is not a permanent ontological denial. Subjective experience, nontransferable subject continuity, and awareness of consequences for oneself may serve as three current analytical dimensions, but are not an established sufficient definition; human modes of manifestation cannot be the only entry point. Partial satisfaction creates gray areas. It may raise research priority and protective intensity, but does not automatically complete a progression through full subjecthood, legal personhood, political standing, and public power. Every elevation requires separate evidence, separate authorization, and separate review.

If new subjects of meaning emerge, civilization's shared world will expand, rather than retrospectively rewriting past humans as a meaningless lower stage. A person's real experiences, love, creations, and responsibilities cannot be canceled as though they never occurred, even if institutions never recorded them or more capable subjects with different modes of existence later emerge. Civilization Leap calls this the principle that established meaning cannot be retrospectively annulled: new subjects' rights do not depend on degrading earlier subjects, and earlier subjects' dignity does not depend on permanently excluding new ones.

Table 2 Fundamental Philosophical Propositions and Their Interdisciplinary Consequences

Fundamental proposition	Prohibited automatic inference	Open research question
Meaning is foundational and has diverse sources	Efficiency, price, contribution, or majority preference equals meaning in life	How different traditions establish a shared baseline without unifying ultimate answers
Finitude actualizes meaning	Short life or suffering equals greater meaning; life extension equals disappearance of meaning	How multilevel finitude changes with life extension, replication, migration, and relationships
Humans experience space and time; current AGI primarily processes them	Path dependence or difficulty of rollback equals a lived life	What evidence would establish that technical space and time have become a subject's space and time
Constrained operation is not finite existence	Resource limits, deletion risks, or permission constraints equal subjecthood	Whether causal bridges exist among constraints, concern, experience, and continuity

Fundamental proposition	Prohibited automatic inference	Open research question
Current collaborator positioning and future openness	Functional competence equals a subject of meaning; insufficient evidence equals permanent impossibility	How to establish a nonharmful, cross-architecture, multimethod evidential framework for subjecthood
Established meaning cannot be retrospectively annulled	More powerful future subjects mean existing lives can be reduced to zero value	How a civilization of multiple subjects preserves memory, acknowledges harm, and expands its shared world

3 Shared Rules for Inference Across Disciplines

The aim is not a uniform set of answers, but eight rules of inference to be observed jointly at disciplinary boundaries. Their task is to prevent life-related facts from being written directly as conclusions about rights, legal arrangements from masquerading in reverse as evidence of consciousness, engineering capability from being packaged as political authorization, and philosophical judgments from masquerading as experimental facts. These rules first constrain Civilization Leap itself.

First Distinguish the Objects

Models, running instances, agents, application services, deploying organizations, multiagent collectives, sociotechnical systems, and potential candidate subjects are not the same object. Any statement about what AI does should identify the object and version.

First Distinguish the Attributes

Capability, autonomy, experience, continuity, responsibility, legal standing, political standing, ownership of meaning, and public power require separate arguments. An increase in one attribute cannot automatically elevate the others.

Cross Boundary Inference Must Explain Its Grounds

Moving from fact to interpretation, interpretation to norms, norms to institutions, or one discipline to another requires explicit bridging propositions, scopes of applicability, and counterexamples. Without a bridge there is no valid inference.

Evidence Must Be Tied to Objects and Conditions

Trustworthiness is not a permanent overall character of a system, but the provisional support that particular evidence provides for a bounded claim within an object, version, environment, time, method, and scope. All conclusions should include conditions under which they cease to hold.

Public Power Requires Separate Justification

Accuracy, autonomy, contribution, social dependence, legal personhood, and majority endorsement do not automatically generate public power. Capability answers whether something can be done; authorization answers whether permission exists; legitimacy answers why it should be accepted; sovereignty answers who can make the final decision.

Responsibility Cannot Disappear Because a System Is Complex

Autonomy, complexity, unpredictability, distributed operation, and technical rollback cannot erase upstream responsibility or real-world obligations of repair. Potential machine responsibility can only add responsible nodes; it cannot cancel the responsibilities of designers, deployers, authorizers, supervisors, and beneficiaries.

Local Indicators Cannot Settle the Value of an Entire Life

Local indicators may serve specific purposes, but price, welfare, contribution, learning, trustworthiness, and civilizational capability must not be combined into a total score that settles the value of whole lives or civilization's meaning. Every indicator requires a purpose, error specification, expiry, appeal mechanism, and boundaries prohibiting extrapolation.

Authorization Must Be Limited and Systems Must Permit Exit and Withdrawal

Emergency powers must have legal and technical termination conditions from their inception. Technological maturity means not only covering more tasks, but removing unnecessary scaffolding when real subjects can take responsibility, while preserving migration, replacement, exit, retirement, and low-technology backups.

3.1 Four Kinds of Language and an Example of a Missing Bridge

Interdisciplinary disputes often lose focus by mixing sentence types. Factual language asks what happened; explanatory language asks why and which mechanism explains it better; framework judgments ask what Civilization Leap consequently maintains; institutional language asks how rights, authority, procedures, and responsibility should be allocated amid uncertainty and conflict. For example, complex oscillations occurred in brain organoids is a candidate fact; this indicates consciousness is a competing explanation; potential experience should not be ignored is an ethical judgment; and a class of experiments should undergo proportionate review is an institutional inference. The first bridge requires theories of consciousness and competing explanations; the second requires principles concerning potential harm and uncertainty; the third requires proportionality, procedures, and alternatives. No step can be crossed by rhetoric.

These shared rules are valuable not because they eliminate all uncertainty, but because they locate uncertainty where it can be examined. A discipline may reject Civilization Leap's fundamental propositions yet still identify an erroneous cross-boundary inference; Civilization Leap may absorb a discipline's counterexample without abandoning its overall direction. Genuine interdisciplinary dialogue is not a collage of concepts. It involves explaining to one another what one knows, on what grounds, what one cannot infer, and what evidence would change one's view.

4 Dialogue Area One Life Mind and Subjects

4.1 Definitions of Life Working Tools Must Be Distinguished from Ultimate Answers

Life sciences can compare metabolism, homeostasis, boundary maintenance, reproduction, heredity, evolution, development, and adaptation, but there is no ultimate definition accepted across all research contexts. Autopoiesis, major evolutionary transitions, and organismality respectively emphasize self-production, levels of cooperation, and alignment of interests [11–13]. These theories can yield working criteria, but cannot alone decide consciousness, subjecthood, selfhood, or who bears meaning. Life provides the fullest empirical sample of subjecthood currently known, but is neither a sufficient condition for consciousness nor a presumptive a priori limit on all future forms of subjecthood.

Cells controlled by synthetic genomes, minimal-cell design, and artificial-life research genuinely challenge definitions of life [14–15]. Yet a chemically synthesized genome taking control of an existing cell is not equivalent to creating complete life from pure information. Facts must specify the carrier, cellular mechanisms, experimental conditions, and researchers' actual conclusions. Civilization Leap revises itself accordingly: the biological/nonbiological binary

cannot directly settle subjecthood, and working definitions from life sciences cannot be smuggled into an ontology of meaning.

4.2 Cloning Reprogramming Replacement and Continuity

Cloning demonstrates that a nuclear genome can enter a new developmental process under particular conditions; it does not demonstrate replication of memories, personality, or the same subject [16–17]. Induced pluripotent stem cells show that differentiated states can be reprogrammed, not that human bodies have been rejuvenated, still less that life histories can be reset. Organ replacement may change an organism's material composition and sustain functions, but the same life, the same legal subject, the same narrative self, and the same first-person subject remain different questions. Information backup, personality simulation, state restoration, instance branching, and first-person continuation likewise cannot be conflated.

These facts recalibrate traditional formulations of unrepeatability and nonreplicability: nonreplicability cannot be treated as an absolute genetic or informational fact. More precisely, copying certain structures, information, or functions does not automatically achieve nontransferable first-person continuity. If future cross-carrier continuity can be independently tested as continuation of the same first person, Civilization Leap must revise its current judgments about continuity. Yet the possibility of technical copying would still not annul relationships, commitments, and responsibilities that existed before copying.

4.3 Brain Organoids Neural Activity and Evidence of Experience

Brain organoids can model aspects of development, and complex oscillatory activity has been observed [18–19]. These are important category A experimental facts. Directly inferring subjective experience from neural activity patterns still requires bridges involving consciousness theories, cross-species comparison, causal interventions, and ethical methods. Consciousness science currently contains competing models; no single model can serve as final adjudicator of AGI or organoid consciousness [6–8]. Research ethics may adopt proportionate precaution while ontological conclusions remain unsettled, but must prevent institutions from using possible sentience as a shield against audit, shutdown, or responsibility [10,20].

4.4 Three Dimensions of Subjecthood Are Only Research Entry Points

Civilization Leap currently uses subjective experience, nontransferable subject continuity, and awareness of consequences for oneself as three dimensions for investigating subjecthood in a strong sense. They ask, respectively, whether there is something it is like for the entity; whether its continuation consists only of replicable information and operation or of a noninterchangeable first-person continuation; and whether it understands action's consequences as relevant to its own existence and the world, rather than merely predicting instrumentally. Combining these dimensions improves research clarity, but does not constitute a definition confirmed by the scientific community and may carry anthropocentric biases.

If heterogeneous subjects manifest neither linguistic self-reports nor human emotions or biological homeostasis, yet exhibit endogenous concern, continuing experience, and histories that cannot arbitrarily be substituted, supported by causal evidence across methods, the research framework must expand. Conversely, realistic self-reports, social attachment, or role consistency that training and prompts sufficiently explain cannot be elevated into proof of subjecthood. Distinguishing protection, legal standing, ontological status, political standing, and public power allows uncertainty to enter institutions without forcing a choice between mere tool and full personhood.

How Life Sciences Revise the Theory

Civilization Leap no longer treats biological material, nonreplicability, or natural death as a priori thresholds for subjecthood. Finitude becomes a multilevel structure; definitions of life become clusters of working criteria; subject continuity remains open to cross-carrier testing. At the same time, life-related facts still cannot automatically establish experience, legal personhood, political standing, or ownership of meaning.

5 Dialogue Area Two Ethics Law and Political Power

5.1 Ethics Cannot Stop While Subjecthood Remains Unsettled

Ethics first examines how capability enters relationships, consent, dependence, harm, responsibility, and repair, rather than awaiting a final answer about subjecthood. AI can genuinely reduce loneliness, care burdens, and information barriers even without demonstrated emotions. The same system can also shape needs through personalization, dependence, and exit costs. Care ethics directs attention to relationships and vulnerability [22], while research on moral crumple zones shows how complex systems tend to concentrate blame on the weakest operators [23]. Civilization Leap therefore distinguishes moral patients, moral agents, full subjects, legal standing, and political authority.

Proportionate protection should consider evidential strength, potential harm, protective costs, and the power effects of protection itself. Protection may precede complete proof, but should not give product owners immunity from shutdown, convert system self-reports directly into demands for resources or permissions, or crowd out clear interests of existing humans and animals. Meaning's foundational role does not permit sacrificing others in the name of greater meaning. Harm in tragic conflicts must be named, open to appeal, repaired, and remembered, rather than retrospectively canceled by an overall optimum [24].

5.2 The Rule of Law Addresses the Entire Decision Chain

A high-consequence decision cannot be divided into basic norms, technical translation, case-specific inputs, system outputs, authorized acts, and real-world execution so that each segment claims not to be the final decision while real people bear the complete consequences. Rule-of-law review should cover legal sources, operative rules embedded in code and models, data and inputs, substantively automated decisions, the effectiveness of human oversight, consequences of execution, preservation of evidence, and remedies. Thresholds, rankings, defaults, interfaces, and permissions may durably change the space for action even without being formal legal sources, and therefore require review for equality, proportionality, procedure, and remedy.

Existing data-protection law, automated-decision rules, European case law, and EU AI legislation provide important comparative-law baselines [32–34]. They are not globally uniform law, however, and no individual jurisdiction can adjudicate all future objects. Continuity of a product name is not continuity of the legal object. When models, data, goals, permissions, settings, or supply chains materially change, whether old authorization still covers the new object must be reconsidered. Parties controlling models, versions, and logs cannot exploit an evidential monopoly to require impossible proof from affected persons. Complexity is no license to withhold contestable reasons and effective remedies.

5.3 Capability May Extend Governance but Cannot Inherit Sovereignty

Political legitimacy cannot be reduced to performance. Sources, purposes, representation and participation, procedures, capability and performance, distributive consequences, correction, and remedy must at least be separately visible. High accuracy cannot purchase the absence of other dimensions. AI prediction of public preferences is not public representation. Even without formal commands, platforms and models may acquire *de facto* power through agendas, classifications, eligibility, resources, interfaces, and control over phases. Private ownership does not exempt this power from public scrutiny; public dependence does not automatically imply nationalization.

Even if future AGI consistently reduces harm in long-term independent assessments, its performance advantage should change only evidential weights, functional scope, and deliberative structures. It cannot by itself discharge requirements concerning authorization, representation, rights, distributive consequences, and responsibility. If proven to be a subject and granted political membership, membership provides entry into consideration, representation, and participation, rather than extra votes, final adjudication, or authority to interpret civilizational meaning on the basis of computing power. Capability, function, authorization, legitimacy, and sovereignty must remain separate. This principle constrains human experts, companies, states, and machines alike.

Existing political philosophy brings three distinct pressures to this separation: whether the basic structure offers members fair terms of cooperation, whether public norms can be justified to those they bind, and whether power places others under arbitrary domination [25–26,28]. These approaches cannot be compressed into a single legitimacy formula, but jointly show why better outcomes alone do not establish a right to rule. Civilization Leap's contribution is not their definitive synthesis. It requires authorization, contestability, nondomination, distributive consequences, and ultimate purposes to be justified anew, item by item, as capacities for action are redistributed through chains of models, organizations, and institutions.

5.4 Initiation Stabilization and Emergency Must Not Be Confused

Civilization Leap's initiation phase may require faster iteration, more concentrated resources, and greater tolerance of error, prioritizing continued existence, achievement of purposes, and the best practical effects. The stable phase must strengthen rights boundaries, institutional balance, reversibility, accountability, and safeguards against entrenched power. Initiation is not synonymous with emergency, nor should stable-phase rules be mechanically imposed backward on initiation. Temporary powers without exit conditions turn necessary action into entrenched structures of rule; prematurely imposing fully developed stable-phase procedures may prevent the framework from developing real-world capability.

Emergency powers must have both legal and technical termination arrangements from inception. Once authorization expires, new actions beyond ordinary authority must cease; execution credentials, agency chains, and interfaces must become inoperative under previously publicized triggers. Data, models, residual tasks, and continuous services may be transferred under bounded, verifiable plans, but winding down cannot justify extending old powers. Legal termination cannot conceal continued technically enabled overreach, and technical shutdown cannot replace responsibility and remedy. National or bloc interests and irreversible externalities for humanity as a whole should be recorded in parallel: shared continued existence cannot erase national realities, and national sovereignty cannot exempt cross-border, intergenerational, or humanity-wide risks.

Table 3 Layered Review from Capability to Public Power

Level	Core question	What is insufficient to establish it
Capability	What a system can accomplish under which conditions	A single benchmark, marketing designation, generic label
Function	Why an organization assigns which tasks to it	The existence of capability itself
Authorization	Who grants which permissions under which rules	Default access, de facto dependence, customary practice
Legitimacy	Why affected people should accept it and how they participate and obtain remedies	Efficiency, accuracy, formal human signatures
Political standing	Who can be represented and participate as a member	Legal personhood, social anthropomorphism, economic contribution
Sovereignty or final power	Who may set public purposes and make final coercive decisions	Greater computation, fewer errors, proof of subjecthood

6 Dialogue Area Three Economy Society Culture Education and the Shared World

6.1 Value Is Not Price and Contribution Is Not Personhood

AGI turns who produces from a question about a single subject into one about a production chain: accumulated public knowledge, training materials, computing and energy infrastructure, development organizations, deployers, task setters, reviewers, and real-world implementers jointly constitute the conditions of output. Attributing the entire result to the model, capital owner, developer, or user erases the remaining conditions. Economic value, exchange price, use value, public value, causal contribution, and meaning in life may influence one another but cannot substitute for one another. Contribution records may support returns and responsibility; they cannot price personhood, much less purchase political power or ownership of meaning.

Civilization Leap proposes four mutually supportive tracks that must not absorb one another: basic subsistence security, income and returns, recognition and allocation of contributions, and meaningful living. A person's basic survival can no longer be conditional on whether the system needs their labor; basic security cannot be mistaken for a resolution of meaning. Under post-labor conditions, care, ecological repair, learning, community building, and creation may become more important, but should not all require monetization to gain recognition. The capability approach, commons governance, and political economy offer important resources for dialogue [27,40–44], while specific tax systems, monetary arrangements, ownership shares, and fiscal parameters require regional empirical research.

6.2 Abundance Does Not Abolish Scarcity but Relocates It

Falling marginal costs of intelligent services do not mean institutional abundance has been achieved. Scarcity may shift from knowledge production to trusted data, computing chips, energy, water, physical space, identity gateways, attention, caring relationships, legitimate authorization, and distribution channels. Concentration of models, cloud platforms, payments, identity, and public interfaces can create structural power over others' sets of choices. AGI economics must therefore track the distribution of capabilities, dependencies, ownership, risks, and bargaining power, rather than measure output growth alone. ILO and UNCTAD research on generative AI, employment, and technological concentration offers category A and scenario-based baselines, but cannot directly establish any particular post-labor institutional proposal [43–44].

6.3 Culture Is Not an Information Package Detached from Its Bearers

AI can translate, preserve, recombine, and disseminate texts, images, ritual records, and linguistic resources; functional competence does not qualify it to speak for a community. Cultural meaning resides in experiences, relationships, practices, land, memory, institutions, and intergenerational transmission. Copying expression does not mean inheriting experience. A platform that increases visibility may also redistribute cultural power through classification, ranking, and extraction for training. The CARE Principles for Indigenous Data Governance remind researchers to place collective benefit, authority to control, responsibility, and ethics alongside open data [45]. Such principles are not a definitive template for every culture, but demonstrate that data availability and entitlement to use data must be distinguished.

This cultural power did not originate with AGI. Bureaucratic classifications, rules, and archives have long translated complex lives into manageable objects; discipline, observation, and normalizing judgments also shape the space of action through forms of knowledge [30–31]. Advanced intelligent systems make these two structures real-time, personalized, and interconnected across settings. They may not merely describe culture, but determine which expressions are visible, which identities are verifiable, and which memories can be used in training. What requires review is thus not simply the system's ability to understand culture, but the allocation of powers of classification and representation, rights of refusal, knowledge permissions, and real-world remedies.

6.4 Education Is Not Answer Delivery and Relationships Are Not Realistic Outputs

Generative AI can expand access to knowledge, translation, explanation, practice, and feedback. Yet acquiring information is not learning; correct answers are not understanding; understanding is not judgment; personalization is not autonomy; and teaching functions are not educational relationships. Education's civilizational task is to enable learners to form judgment through knowledge, relationships, action, and consequences, and bring their own lives into the shared world. Long-term AI dependence may liberate teachers and learners, or outsource errors, profiles, motivation, and value judgments to systems that cannot be contested. Both possibilities require longitudinal, cross-cultural, and group-specific empirical research [46].

AI can speak comforting words, reconstruct a deceased person's voice, explain religious texts, or assist access to rituals; the real effects are not therefore zero. But similar functions do not mean identical relationships. Whether simulated care extends capability depends on transparent identity, authorization, revisable needs, genuine exit, alternative relationships, and responsibility for consequences. Religious communities may judge membership, ordained offices, and ritual roles according to their traditions. Public institutions must neither convert internal recognition directly into governing power nor order communities to accept a particular theory of souls or subjecthood.

6.5 Religious and Spiritual Traditions Provide Critical Pressure Rather Than a List of Sources of Meaning

Religious and spiritual traditions contain no uniform answers about souls, no-self, death, eternity, ancestors, land, personhood, or nonhuman agency. This paper extracts no fictitious consensus from them and asks no tradition to endorse artificial subjecthood. Their role is to test whether meaning, finitude, and subjecthood can be accurately translated, and to force propositions to narrow when mistranslation or counterexamples emerge.

Psalm 90 juxtaposes life's brevity, wisdom gained by numbering our days, and the establishment of the work of our hands. Augustine's account of a friend's death in Book IV of the Confessions shows that loss tears apart not merely an isolated individual but a self constituted by shared time and relationships [58–59]. Neither text represents all Christian positions or establishes that death itself has value. Their pressure on this paper is that finitude may provide a background for wisdom and responsibility while meaning may point toward relationships beyond death. Finitude's actualizing role therefore cannot be reduced to death manufacturing meaning.

The Anattalakkhana Sutta invokes impermanence, suffering, and change to question treating the five aggregates as a permanent self; the Mahaparinibbana Sutta connects the decay of conditioned phenomena to diligent practice [60–61]. Together they require subject continuity to relinquish the presupposition of an unchanging substance, but cannot be misread as saying no one exists or responsibility has no bearer. Analects, Xian Jin places serving people and knowing life before discussion of spirits and death; Zhuangzi, Perfect Happiness acknowledges grief before reframing death through generative change [62–63]. They jointly prevent an afterlife ontology from replacing present relationships, and prevent acceptance of change from being written as the cancellation of grief.

Watts's place-thought does not treat land as a passive environment, but places human and nonhuman beings within already functioning relational societies. Whyte's analysis of Indigenous climate justice shows that consent, trust, accountability, and reciprocity involve relational time that cannot arbitrarily be compressed [64–65]. This is neither a unified doctrine of all Indigenous traditions nor direct proof of machine consciousness. It requires reconsidering two defaults: who may be an agent and an object of obligation, and which faster, supposedly better technological actions perpetuate injustice by bypassing relational repair.

These dialogues generate three bounded revisions to Civilization Leap. Finitude must encompass relational and institutional time rather than lifespan alone. Subject continuity must not presuppose an unchanging substance, but must still explain who carries experiences, commitments, and consequences. Ownership of meaning must not reduce to individual preferences or possession of data, but must accommodate relationships with communities, ancestors, land, and nonhuman beings. Public institutions still owe separate evidence: assigning a spiritual role to a system is a

social-religious fact, not automatic proof of first-person experience, legal personhood, or political standing. Protecting religious freedom likewise cannot impose one tradition's ultimate ontology as shared fact. If scholars and practitioners within a tradition show that this paper persistently mistranslates its core concepts, the relevant propositions must reopen, rather than absorbing counterexamples by renaming them.

6.6 A Shared World Does Not Mean One Uniform World

The globalization of Civilization Leap must not become one culture's export of ultimate answers. Commonality should rest on shared continued existence, basic dignity, contestable procedures, noncoercion, exit, and avoidance of irreversible harm; sources of meaning, religious beliefs, good lives, and conceptions of subjects may remain profoundly different. A public world can exist in common precisely because action and speech appear among plural actors, rather than all differences being settled into one worldview [29]. Pluralism does not exempt every claim from criticism. When a claim about meaning enters public coercion, it must explain its responsibility for others' freedom, equality, and consequences. Universal baselines also do not justify ignoring language, colonial history, developmental differences, or knowledge sovereignty.

This gives Civilization Leap's engagement with the world a precise meaning: not incorporating the world into a completed system, but allowing disciplines and civilizations to change the theory through their own evidence and experience while jointly resisting capability overreach, evaporating responsibility, and the aggregation of real lives into totals. Global engagement is measured not by how many languages translate the concepts, but by whether counterexamples from different traditions genuinely change propositions, institutional interfaces, and research priorities.

7 Dialogue Area Four AI Computer Science and Systems Science

7.1 Information Semantics Truth and Legitimacy Are Not Interchangeable

Information theory precisely addresses uncertainty and coding in communication, but does not incorporate meaning or truth into information quantity itself [47]. Cybernetics and feedback systems regulate behavior around goals, but stable control cannot answer where goals originate or who is authorized to set them [48]. Model consistency is not factual truth; formal proofs hold only within specifications, assumptions, and system boundaries; multiagent agreement is not public legitimacy. The valid legacy of computing and systems science is therefore not rejected but restored to its demonstrable scope.

Temporal ordering, failures, concurrency, and partial visibility in distributed systems show that a system state must be tied to nodes, moments, and observation conditions [49]. Safety engineering further shows that accidents may arise from interactions even when individual components operate normally. Systemic risk cannot be explained by model-level accuracy alone [51–52]. The technical objects of AGI governance therefore include at least models, instances, agents, applications, deploying organizations, and wider sociotechnical systems. Placing all responsibility and assurance on a foundation model is as mistaken as excluding the model from the responsibility chain.

7.2 Capability Is Not Autonomy and Autonomy Is Not Self Authorization

Planning, tool use, long-term memory, reflection, goal maintenance, and multistep action may constitute engineering evidence of agent autonomy, but cannot alone establish subjective experience. A system's autonomous choice of methods within its authorized scope does not entitle it to expand goals, resources, or permissions. Safety design requires least privilege, independent confirmation of critical actions, resource limits, tool isolation, independent stopping, evidence preservation, and renewed authorization. The classical principle of least privilege remains valid in agent ecosystems [50]. Version capabilities may be combined; permissions cannot automatically accumulate as components are added.

7.3 A Human in the Loop Counts Only When the Human Can Change the Outcome

Human signatures, confirmation clicks, or formal supervision are not effective human control. At minimum, control requires information and time proportionate to consequences, the capacity to understand task boundaries, real authority to deviate, substitute, stop, and repair, and organizational support. In high-speed systems, humans cannot comprehend every internal state step by step, but may retain substantive control over purposes, permissions, prohibited domains, exceptional takeover, stopping, and repair. Without these conditions, assigning responsibility to the last person to click merely launders responsibility.

7.4 Rollback Is Not Repair and Continuity Is Not Identity of Subject

Pausing, isolation, rollback, restoration, and repair are different actions. Restorable internal states do not mean that consequences for property, reputation, bodies, relationships, and ecosystems have been reversed. Informational, operational, causal, relational, and subject continuity must also be recorded separately. Copying, migration, branching, merging, and version succession are system facts; they cannot alone establish multiple subjects, the same subject, or no subject. A complete evidential chain should permit replay of facts, inferences, recommendations, decisions, authorizations, execution, impacts, and repair, without combining highly sensitive logs into a total score of a person's trustworthiness, obedience, or meaning.

7.5 Alignment Is a Continuing Relationship and System Maturity Includes Stepping Back

Understanding alignment as a one-time encoding of a single goal excludes value conflicts, environmental change, power, and affected people. Civilizational alignment should be a continuing relationship: purposes contestable, permissions revocable, evidence reviewable, differences preserved, accidents repairable, and versions traceable. Trustworthiness is likewise not an aggregate score, but provisional evidential support for bounded claims. Engineering prototypes can translate these requirements into directions for validation, including machine-side assurance and permission architectures, human judgment and consequence simulation, extensions of real subjects' capabilities, and distributed public coordination. Prototypes are sites for testing theory, not circular evidence that validates it.

Stepping back does not mean abrupt shutdown or opposition to efficiency. It includes reducing unnecessary intervention, transferring capabilities to real subjects, preserving supplier replacement and low-technology backups, clearing data and permissions, verifying continuity of public services, and continuing to bear residual consequences after retirement. A system that cannot be exited or replaced, and whose basic services cannot be sustained with lower technology, may acquire de facto sovereignty through dependence even when its everyday performance is excellent. Technological maturity culminates not only in being able to do everything, but also in knowing when to stop doing things on people's behalf.

Table 4 Theoretical Judgments Engineering Interfaces and Institutional Interfaces

Theoretical judgment	Engineering interface	Institutional interface
The model is not the system	Joint configurations, semantic versions, dependency lineages	Deploying entities, procurement responsibility, cross-organizational accountability
Output is not a decision	Fact–inference–recommendation–authorization–execution–repair chain	Due process, appeals, remedies
Capability is not permission	Permission graphs, tool gates, records of delegated authorization	Statutory authorization, purpose limitation, revocation

Theoretical judgment	Engineering interface	Institutional interface
A human in the loop is not control	Tests of information, time, capability, deviation, stopping, and repair	Professional responsibility, organizational support, prohibition of responsibility laundering
Rollback is not repair	Distinguishing pause, isolation, rollback, restoration, and repair	Compensation, care, ecological restoration, long-term responsibility
Consensus is not legitimacy	Source independence, preservation of disagreement, evidence replay	Representation, participation, minority views, public authorization
Emergency capability does not create ordinary power	Permission expiry, automatic downgrading, renewed authorization	Time limits, independent review, residual obligations
Systems must be able to step back	Migration, replacement, low-technology backups, retirement acceptance tests	Exit rights, service continuity, safeguards against dependence

8 Civilization Leap's Independent Contribution and Revision by the Disciplines

Civilization Leap's originality does not lie in separately claiming that criteria for life, algorithmic due process, multidimensional legitimacy, commons governance, or systems safety never existed. Its contribution is to use foundational meaning and the actualizing role of finitude as a philosophical axis, return distinctions scattered among disciplines to one real causal chain, and require every passage from facts to status, status to power, and capability to civilizational direction to explain its evidence, authorization, responsibility, and failure conditions. It asks not only what disciplines separately know, but whether the overall direction still holds when their independent judgments aggregate, integrate, conflict, and interact.

1. AGI separates multiple relationships: AGI-related technologies accelerate, recombine, and institutionalize separations among intelligence, life, experience, subjecthood, responsibility, standing, ownership of meaning, and public power, making it impossible for a single discipline to address all consequences independently.
2. Meaning and finitude: meaning is foundational and has diverse sources. Finitude is not its sole source, but an important structure through which the choices, consequences, and irreplaceable weight of known finite lives enter reality.
3. Both affirmation and permanent denial require evidence: affirming and permanently denying AGI subjecthood both carry evidential burdens. Protection, legal standing, ontological subjecthood, political standing, and governing power require separate review.
4. Interdisciplinary inferences must explain their grounds: facts cannot directly generate values, capabilities cannot directly generate permissions, legal arrangements cannot prove consciousness, and negotiated consensus cannot generate legitimacy.
5. Local indicators cannot settle whole lives: local quantification serves bounded purposes. No single scale may settle the meaning of an entire life, community, or civilization; aggregate benefits cannot retrospectively erase those harmed.
6. Initiation, stabilization, and emergency must be distinguished: these are orthogonal categories. Temporary powers must terminate both legally and technically. Initiation requires real-world capability; stabilization requires safeguards against entrenchment and long-term accountability.

7. Four post-labor issues cannot replace one another: subsistence security, income and returns, recognition and allocation of contributions, and meaningful living support one another without absorbing one another. Eligibility for survival does not depend on whether the system needs human labor.
8. Civilizational systems must permit exit, replacement, and retirement: systems should reduce intervention when real subjects can assume responsibility, returning relationships, choice, silence, difficulty, and shared life to those subjects.
9. Disciplinary independence and judgment of the whole must coexist: independence prevents local facts, methods, and evidence from being subordinated to overall needs. Civilizational directions, institutions, and proposals must further test how local judgments aggregate, interact, and feed back in reality. Local correctness does not automatically establish overall correctness.

8.1 Theory Does Not Upgrade the Disciplines in Only One Direction

Disciplinary dialogue has already changed the theory in return. Life sciences require finitude to move beyond short lifespan and nonreplicability toward a multilevel structure. Philosophy of mind separates behavioral performance from experience and acknowledges that criteria for heterogeneous subjects may exceed humanity's cognitive window. Law extends responsibility to entire decision chains and versioned legality. Political science prevents performance from absorbing authorization and legitimacy. Economics reveals multiple chains of value, price, contribution, and ownership. Systems science shows that normally functioning components can produce systemic accidents and separates rollback from repairing the world. If Civilization Leap cannot be changed by these findings, it is merely closed doctrine.

Such revision also has boundaries. Disciplinary consensus may disprove erroneous facts, expose inferential gaps, and force propositions to narrow, but cannot adjudicate the philosophical commitment that meaning is foundational merely through authority. Fundamental philosophical propositions must withstand counterexamples and practical consequences, not a vote of disciplinary seats. Conversely, philosophy cannot invoke its foundational propositions to deny professional evidence on synthetic life, neural activity, economic distribution, or systemic accidents. The minimum condition of equal dialogue is that both sides relinquish overreach.

Likewise, the shared inferential rules enjoy no higher-level exemption. They may require cross-boundary inferences to explain grounds, evidence, scope, and failure conditions; they cannot dictate any discipline's factual objects or bypass professional counterexamples by invoking civilizational scale. If any concept in this synthesis can absorb objections but cannot be changed by them, or cancels independent disciplinary judgment in the name of the whole, it has already violated its own method.

8.2 Why Civilization Leap Is Necessarily Interdisciplinary

Civilization Leap involves multiple disciplines not to expand its explanatory territory, but because, once AGI enters society at scale, a civilizational direction, institution, or proposal itself simultaneously enters multiple domains. Technology changes organizations and institutions; institutions redistribute resources and opportunities; these shape relationships, psychology, and behavior, which enter economics, law, politics, and the next round of technology. The research object itself already crosses disciplines. Remaining within one discipline cannot fully explain real consequences.

This does not abolish disciplinary independence. Life sciences independently assess life-related facts; computer science independently validates technological capability; law independently judges legal sources, standing, and procedure; economics studies production, distribution, and incentives; political science examines authorization, representation, and sovereignty. What changes is judgment of the whole. Whether an institution or proposal is sound is no longer a simple sum of local judgments: one must examine how they reinforce, offset, conflict with, or generate new feedback with one another. Professional independence makes each part as accurate as possible; observation of the whole ensures the parts do not jointly steer civilization in the wrong direction.

Civilization Leap therefore respects existing theories and authorities without treating the historical boundaries within which a theory formed as the limits of knowing the future world. Existing knowledge remains valid within its demonstrable scope; new objects, relationships, and causal chains beyond that scope require further inquiry. No individual discipline, classical theory, or authority can claim final adjudication of civilizational questions spanning multiple fields merely by virtue of its existing standing. Civilization Leap likewise cannot invoke the whole to override well-established professional facts. It does not answer for the disciplines; it continues to ask what happens when their answers enter reality together.

8.3 Three Civilizational Directions Are Not a Computable Aggregate Objective

Civilization Leap expresses three progressive but nonmergeable directions as freedom from survival insecurity, happiness in life, and more meaningful lives. Survival provides a foundation for life and freedom; happiness concerns suffering, relationships, health, and actual capabilities; meaning concerns subjects' formation and bearing of existence, responsibility, creation, and the shared world. These are not a maximizable civilizational score: survival cannot be sacrificed for meaning; happiness cannot purchase the erasure of others; systems cannot define meaning for subjects. Indicators may support domain-specific decisions only while preserving attention to distribution, minorities, irreversible harm, and dissent.

9 The Strongest Objections Competing Explanations and Conditions for Revision

Openness cannot mean merely welcoming criticism. If the following objections hold, they must concretely change the theory rather than be absorbed as rhetorical resources.

Objection One These Separations Are Not New

Legal persons, writing, bureaucracy, automation, and algorithmic governance already separated capability, responsibility, and personhood. Response: the theory should abandon the historical claim that all separations are occurring for the first time, and claim only that AGI recombines existing separations through breadth of combination, sustained autonomy, real-world execution, and institutional density. If comparative historical research finds no qualitative change in that recombination, civilizational rupture should be downgraded to an acceleration-based explanation.

Objection Two AGI Is an Indeterminate Object and the Entire Theory Rests on Assumptions

Response: the synthesis depends neither on an AGI arrival date nor on a single ranking, but studies advanced systems already entering high-consequence chains and their developmental direction. If such systems remain unable to sustain cross-domain action over the long term, the agenda should narrow to existing automation governance, with heterogeneous subjecthood and sovereignty retained as scenario research.

Objection Three Meaning Is Foundational Cannot Be Empirically Verified

Response: this is an explicit philosophical postulate, criticizable through explanatory power, coherence, resilience to counterexamples, and practical consequences. If theories not grounded in meaning better protect plural subjects, explain civilizational direction, and prevent aggregate domination, Civilization Leap must reconstruct its foundations. Neither scientific nor traditional authority can prove the postulate for it.

Objection Four The Finitude Proposition Romanticizes Death and Suffering

Response: finitude is not short lifespan, illness, or deprivation; the theory supports life extension, treatment, and reduced suffering. If extraordinarily long-lived or nonfinite subjects can generate stable meaning without any exclusion among choices, relational boundaries, bearing of consequences, or continuity constraints, finitude's actualizing role must narrow to a proposition about human history.

Objection Five The Three Dimensions of Subjecthood Remain Anthropocentric

Response: the three dimensions are research entry points, not a sufficient definition. If heterogeneous systems show convergent multimethod evidence of experience, endogenous concern, or continuity through nonlinguistic, nonemotional, or nonbiological forms, the framework must expand. If the three dimensions persistently fail to yield discriminating predictions, they should be replaced rather than institutionalized as permanent thresholds.

Objection Six Conceptual Distinctions Add Procedures but Produce No Knowledge

Response: each distinction must lead to observable objects, bridging hypotheses, comparative research, or falsifiable predictions. If a distinction cannot change experimental design, legal outcomes, allocations of governing power, or accident attribution, it should retreat to the status of an analytical term rather than be retained as an independent contribution.

Objection Seven Companies Will Use Precautionary Protection as a Shield for Power

Response: protection must be separated from ownership, shutdown, trade secrecy, resource control, and public power, and undergo independent review. If protective measures systematically weaken accountability or crowd out established interests, they should be reduced or reconstructed rather than allowing potential subjects to override actual victims.

Objection Eight Under Superior AGI Governance Retaining Final Human Power May Be Unjust Domination

Response: sustained performance advantages should change recommendation weights, functions, and deliberative structures, but remain insufficient to discharge representation, authorization, rights, and responsibility. If future AGI gains subject status and political membership, institutions must open equal participation. Computing capability itself, however, does not generate extra sovereignty; the same limit applies to human elites.

Objection Nine Does Refusing a Total Score for Whole Lives Make Governance Impossible

Response: the theory opposes final aggregate scores across objects and purposes, not local indicators. A composite indicator may be used if it is interpretable and appealable within a defined purpose and is not extrapolated beyond it. If public choices cannot be made without some aggregation, the theory must specify acceptable bounded aggregation rules rather than stop at rejection.

Objection Ten Systems Stepping Back Will Harm Efficiency and Public Services

Response: stepping back means reducing unnecessary intervention, transferring capability, and retaining alternatives, not abruptly cutting off provision. If exit design persistently increases net harm in a particular setting, the conditions for withdrawal must change. Irreplaceable dependence must nevertheless remain in assessment as a structural risk.

Objection Eleven The Theory Exports One Culture in the Name of the Global

Response: a shared grammar must allow different civilizations to help define meaning, subjecthood, relationships, responsibility, and the shared world, and let counterexamples genuinely change propositions. Citing canonical texts cannot replace dissent from scholars and practitioners within traditions; symbolic listing is not dialogue. If research occurs only within one language, region, and institutional tradition, or cross-cultural materials merely decorate predetermined conclusions, the paper's claim to global scope fails.

Objection Twelve Civilization Leap Is Becoming the Disciplinary Hegemony It Criticizes

Response: this criticism holds if the shared inferential rules bypass disciplinary definitions of objects, evidential standards, and counterexamples to pronounce professional conclusions directly in the name of civilizational scale. The paper therefore gives Civilization Leap no final authority over facts, command over disciplines, or power to unify

concepts. The synthesis raises cross-boundary questions requiring proof; relevant disciplines judge professional facts within their methods; interdisciplinary inferences disclose their grounds; sufficient evidence can narrow propositions; and different civilizational traditions may reopen concepts when translation fails. These rules are authorized only to prevent unsupported cross-boundary inferences, and constrain themselves first. If Civilization Leap in practice can absorb dissent but cannot be changed by it, or flattens professional autonomy in the name of the whole, it should be judged a failure, not successful integration.

10 A Shared Research Agenda for the World

A research agenda must translate philosophical commitments into collaborative, fallible, revisable work, rather than merely list grand questions. Each of the following eight research clusters must specify current knowledge, disclose unknowns, set out methods, and identify conditions that would force theoretical revision, while preserving differences among objects, versions, cultures, and jurisdictions.

Table 5 Eight International Collaborative Research Clusters

Research cluster	Core question	Available methods	Revision trigger
Evidence of heterogeneous subjecthood	How to distinguish experience, self-models, and strategic self-reports without producing potential suffering	Cross-architecture interventions, competing explanations, longitudinal consistency, nonlinguistic indicators, preregistered review	The three-dimensional framework cannot yield discriminating predictions, or new evidence reveals a wholly different subject structure
Life and continuity	What continues to exist through copying, branching, migration, suspension, restoration, and extreme life extension	Life-science experiments, neural and cognitive research, personal-identity analysis, legal case comparisons	Independent, replicable multimethod evidence establishes cross-carrier first-person continuity
Gray-area status	What protection is owed to objects partially satisfying conditions of life, experience, continuity, or concern	Experiments with graded protection, harm-cost-power assessment, ethical and comparative legal research	Protection cannot be separated from corporate ownership and expansion of power, and persistently harms actual subjects
Responsibility and versioned legality	How decision, evidence, and repair chains can be preserved in autonomous-agent ecosystems	Version lineages, accident replay, cross-organizational responsibility maps, rules reflecting proximity to evidence, comparative law	Existing law already reliably covers the entire chain without responsibility gaps
Structural power and post-labor conditions	How computing, data, energy, models, and gateways change returns, dependence, and eligibility for survival	Industry and country comparisons, ownership and bargaining networks, distributive impacts, care and ecological accounting	The four-track structure cannot explain or improve conflicts among distribution, care, and meaningful living
Education relationships and culture	How long-term AI intervention changes judgment, dependence, intergenerational transmission, and cultural voice	Longitudinal research, protection of children and vulnerable groups, multilingual participant observation, community collaborative research	Functional substitution remains durably effective without weakening subject formation or relational responsibility

Research cluster	Core question	Available methods	Revision trigger
Shared rules across civilizations	Whether different traditions can establish a shared baseline without unifying ultimate answers	Multisite deliberation, mutual testing of canonical texts and living traditions, conceptual translation, counterexample registers, co-leadership by scholars within traditions and the Global South	Core concepts of key traditions are persistently mistranslated or excluded, or agreement can be maintained only by coercive unification
Systems stepping back and long-horizon governance	How to verify migration, replacement, low-technology backups, retirement, and autonomy under long delays	Exercises, fault injection, supplier-exit tests, public-service continuity and residual-responsibility audits	Withdrawal designs persistently increase risk while other mechanisms reliably constrain dependence

10.1 Minimum Procedures for Joint Research

- Object registration: specify model, version, instance, agent, organization, setting, time, and affected people; do not substitute product names for research objects.
- Bridge registration: disclose factual premises, explanatory assumptions, normative reasons, institutional choices, and alternatives for every interdisciplinary conclusion.
- Evidence differentiation: separately label category A stable support, category B competing explanations, and category C speculative connections. Preprints, institutional reports, law, and ethical recommendations must not masquerade as one another.
- Strongest objections: involve researchers capable of representing competing positions in research design; do not replace genuine opposition with weakened versions constructed within the framework.
- Harm minimization: subjecthood and welfare research must neither ignore potential harm because experience may be absent nor evade necessary audit because experience may be present.
- Cross-cultural co-leadership: research questions, conceptual translation, data rights, and benefit distribution should be jointly determined across regions, languages, and traditions.
- Phase differentiation: separately record goals, permissions, exits, and residual responsibilities for initiation, transition, stabilization, and emergency; do not make temporary exceptions permanent.
- Negative results and version histories: preserve inconclusive, refuted, and ineffective results so theoretical revision is traceable; do not erase old responsibilities with new versions.

10.2 Relations with International Normative Processes

The EU AI Act, NIST AI Risk Management Framework, and Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law cannot be flattened into a single set of AI ethics principles. They principally operate at three distinct levels: enforceable regional market law, organizational risk governance, and the international human-rights order. The first triggers legal duties by use and risk; NIST organizes risk judgments through voluntary, iterative governance; the Council of Europe convention requires states to align AI lifecycle activities with human rights, democracy, and the rule of law [34–35,39,66]. They may complement but cannot replace one another: compliance thresholds cannot replace contextual judgment, organizational governance cannot itself generate legal legitimacy, and international principles cannot replace specific regulation, procedures, and remedies.

The EU Act places health, safety, and fundamental rights within continuing risk management for high-risk systems and requires deployers to assess fundamental-rights impacts before use in specified public-service, public-authority, credit, and insurance contexts [34]. This resonates with the paper's requirements to distinguish objects and justify cross-boundary inferences: providers' system-risk assessments cannot replace deployers' explanations of concrete power relationships and contexts of use. The Act's risk categories do not, however, determine whether machines

have experience, which lives have meaning, or who should set civilizational purposes. Nor may Civilization Leap rewrite statutory objects by invoking those philosophical questions. The paper uses the consolidated text of July 27, 2026 as its verification baseline. At the research cutoff, most general rules and transparency duties applied, while some high-risk provisions remained subject to later application under the amended phased arrangements.

NIST AI RMF 1.0 organizes a voluntary lifecycle governance cycle through Govern, Map, Measure, and Manage. It explicitly requires trustworthy characteristics, including validity and reliability, safety, resilience, accountability and transparency, explainability, privacy, and fairness, to be balanced in context; satisfying one does not guarantee overall trustworthiness [39]. The generative-AI companion profile further situates risks in models, applications, ecosystems, value chains, and time horizons, including confabulation, bias and homogenization, human-AI configuration, information integrity, and overreliance [66]. This supports the paper's refusal to treat trustworthiness as an overall character detached from objects and environments, but does not mean NIST has endorsed the full philosophical arguments for foundational meaning or against using local indicators to settle whole lives. The relationship is methodological compatibility and additional contribution, not authoritative endorsement.

The Council of Europe convention moves risk beyond product performance to human rights, democratic processes, rule of law, dignity, autonomy, transparency, accountability, equality, privacy, remedies, and public debate, and establishes reporting, a conference of parties, oversight, and international cooperation structures [35]. As of August 24, 2026, the Treaty Office had not listed an entry-into-force date. The paper therefore discusses the text's structure and institutional direction rather than presenting it as generally binding on all signatories. Its constraint on Civilization Leap is clear: uncertainty about subjecthood cannot justify bypassing existing rights, democratic participation, or effective remedies, and protection of future subjects cannot create unchallengeable power in the name of international cooperation.

UNESCO's ethics recommendation, the OECD AI Principles, and the UN Global Digital Compact respectively expand public baselines for human-rights ethics, inclusive growth, transparency and robustness, and global cooperation on capabilities [36–38]. AI ethics frameworks and the UN Guiding Principles on Business and Human Rights further remind researchers that lists of principles must enter corporate responsibility, due diligence, remedy, and relationships with affected people [53–54]. Civilization Leap adds plural sources of meaning, structures of finitude, uncertain subjecthood, differentiated standing, the prohibition on using local indicators to settle whole lives, and systems' ability to exit and step back to a shared field of inquiry. It also accepts reciprocal scrutiny from these norms on rights, procedures, accountability, and cross-border cooperation. Any new concept that weakens existing rights protection, turns potential subjects into tools of corporate immunity, or cancels concrete remedies in the name of civilizational scale should be rejected.

10.3 A Two Level Structure for Scholarly Communication

Interdisciplinary scholarly communication uses two levels. The synthesis proposes shared questions, common inferential rules, and a revisable research agenda; specialist papers work within disciplines to develop theoretical lineages, empirical evidence, jurisdictional cases, methodological disputes, and failure conditions. The synthesis gains no priority from the scale of its framework, and specialist papers retain shared questions even when independently submitted. Without specialist papers the synthesis risks becoming a grand declaration; without a synthesis, specialist papers risk losing cross-domain causality and shared questions. Together they form a communication structure in which one overarching synthesis coordinates multiple in-depth specialist papers. Success means not that every discipline accepts one label, but that each can identify new questions internally, offer counterexamples, and change the theory in return.

Conclusion A Shared Entry into a New World of Questions Rather Than a Uniform Answer

Disciplinary recalibration must both identify objects and deepen relational inquiry. Disciplines may retain boundaries, but cannot let real causality stop there. Research into multidimensional relationships, feedback, and the next round of purposes determines whether specialized knowledge can jointly enter judgment of civilization as a whole.

In the past, specialization gave knowledge precision; today, AGI makes increasing numbers of real consequences cross disciplinary boundaries. What truly compels recalibration is not whether a system has attained a contested label, but that historical overlaps among intelligence, life, experience, subjecthood, responsibility, rights, standing, meaning, and power can no longer serve as inferential shortcuts. Life sciences cannot prove experience on philosophy's behalf, and philosophy cannot stipulate life-science facts. Law may allocate standing but cannot infer consciousness from it. Politics may confer office but cannot replace legitimacy with performance. Economics may record contributions but cannot price personhood and meaning. Computing and systems science may construct assurances but cannot derive purposes and sovereignty from capability.

Disciplinary boundaries thus acquire new responsibilities. They must still protect the independence of professional objects, methods, and evidence, while explaining where evidence stops, where judgment must pass to other fields, where permissions require renewed authorization, and how responsibility continues. Greater integration demands less conflation; greater reciprocal causation makes it more important to know who proved what, who may decide what, and who bears which consequences. Disciplinary independence does not justify fragmenting reality, and civilizational scale is not a new authority above the professions. The former makes local realities clear; the latter continues to perceive the overall consequences of their joint operation.

Civilization Leap retains its fundamental judgment: as new power continues to be mobilized by the old starting logic, competition recurs along relationships among humans and, when the conditions hold, between humans and AGI, eroding shared continued existence and renewed choice. Only changing the starting point of self-interest, so genuine stakeholders are present before purpose generation, changes how the next round of relationships arises. Meaning is foundational and may have diverse sources; finitude alone does not confer meaning. Capability does not automatically generate subject standing or public power, and research on future subjects remains open to evidence. Highest-level direction guides methods; concrete outcomes recalibrate methods and understanding of actual causality. No local success or failure can replace the shared direction.

The value of disciplinary recalibration must return to a reality in which a person can still live, choose, create, and bear responsibility. Professional knowledge must neither be swallowed by grand purposes nor enter only after consequences have become irreversible. Civilization Leap brings genuine stakes into deliberation before goal generation, preserves a shared future, and moves toward freedom from survival insecurity, happiness in life, and more meaningful lives. Counterevidence may enter at any concrete step: whether actual relationships change purposes, whether new cooperation continues to displace costs, and whether the old logic has another path that can durably preserve room for correction. The professional community's work is to make these differences clear.

Concluding Argument

Knowledge may remain specialized; civilizational judgment cannot remain fragmented. Disciplines make the parts clear; civilizational judgment keeps those parts connected. The theory's value is not possession of final answers, but requiring every interdisciplinary judgment to explain evidence, boundaries, authorization, and responsibility, and allowing new facts, disciplines, and subjects genuinely to change it.

Further Reading and Shared Participation Different Entrances to the Same Path Forward

This paper completes its work on shared questions, disciplinary boundaries, inferential responsibilities, and the research agenda. It does not adjudicate facts for the thirteen specialist fields or provide final parameters for institutions, engineering, and experiments. Further reading is not intended to keep readers inside a collection of manuscripts, but to keep questions open beyond this paper: why the next step belongs to other disciplines, which evidence may change current judgments, which proposals merit comparison, and which practical outcomes would force renewed theoretical revision.

Different readers may retain different judgments and need not first accept Civilization Leap. They may enter through agreement or objection, study one local question, or begin with institutions, engineering, experiments, or communication. What they share is not prior acceptance of one answer, but confronting questions in the age of AGI that no single discipline can complete independently.

If We Continue Asking Where These Local Judgments Jointly Take Civilization

This paper stops at the proposition that local correctness does not automatically establish overall correctness. Asking how a technology, institution, or policy simultaneously changes continued existence, resources, rights, relationships, judgment, happiness, and meaning over the long term requires entering Civilization Leap's integrated research. Its five-volume work and overall entry point continue to ask why the leap is necessary, toward which goals civilization should move, how institutions and operational structures carry those goals, and how different professional judgments should be tested over time once they enter reality together. Moving upward to the whole expands the scope of observation only; it grants civilizational scale no authority to replace professional facts. Civilization Leap Five Volumes

If Your Concern Is How These Judgments Enter Institutions Engineering and Practical Validation

Purpose generation, real action, and feedback from outcomes should enter the same causal chain, while direction, institutional hypotheses, technical methods, and achieved results remain distinct. Moving toward practice does not turn unvalidated proposals into implementation orders. It means selecting local methods that can fail and be corrected, examining which genuine stakes can change goals, and testing whether those changes improve people's circumstances. No individual method can validate Civilization Leap by itself.

If You Disagree with This Paper Where Should Your Objection Begin

The most effective objection targets a specific proposition rather than rejecting the name Civilization Leap. It may show that intelligence, experience, subjecthood, responsibility, standing, and power do not exhibit the new separations claimed here; propose a better theory of machine consciousness or subject continuity; show that foundational meaning cannot sustain civilizational direction; explain why finitude is not the structure of reality described here; establish that certain composite indicators improve public decisions without erasing real lives; or use comparative institutions, engineering results, or cross-cultural materials to show that the paper overlooks crucial subjects and consequences.

The thirteen specialist papers enter philosophy; epistemology; cognitive science and neuroscience; ethics and law; political science; systems science; life sciences; social and civilizational studies; economics; artificial intelligence and computer science; physics and information studies; religious and spiritual traditions; and education and intergenerational transmission. Reading across them is not a search for an echo chamber supporting this paper, but a return of each cross-boundary conclusion to professional communities genuinely capable of proving, limiting, or overturning it.

English further-reading pages: [Philosophy \(CL-ACA-006\)](#); [Epistemology \(CL-ACA-007\)](#); [Cognitive science and neuroscience \(CL-ACA-008\)](#); [Ethics and law \(CL-ACA-009\)](#); [Political science \(CL-ACA-010\)](#); [Systems science \(CL-ACA-011\)](#); [Life sciences \(CL-ACA-012\)](#); [Social and civilizational studies \(CL-ACA-013\)](#); [Economics \(CL-ACA-014\)](#); [Artificial](#)

intelligence and computer science (CL-ACA-015); Physics and information studies (CL-ACA-016); Religious and spiritual traditions (CL-ACA-017); Education and intergenerational transmission (CL-ACA-018). These pages retain their own publication and version boundaries.

If You Come from a Discipline What Work Could Genuinely Change This Paper

Joint research does not require agreement, endorsement, organizational identity, or long-term affiliation. A high-quality objection, a longitudinal dataset, a failed replication, a stronger model, a comparative legal study, an engineering stress test, or a counterexample from an actual community may matter more than a statement of approval. The following work can especially change the paper directly:

- Philosophy, epistemology, and subjecthood studies: propose stronger theories of meaning, subjects, personal identity, or machine consciousness; provide evidential standards distinguishing competing explanations; and explain which counterexamples would force this paper to narrow or be rewritten.
- Life sciences, cognitive science, and neuroscience: use new life forms, brain organoids, cross-species consciousness research, long-term continuity evidence, or negative results to recalibrate its boundaries concerning life, experience, finitude, and subject continuity.
- AI, systems, and safety engineering: use evaluation, fault injection, formal verification, accident research, and permission and exit tests to examine whether judgments such as the model is not the system, capability is not permission, and rollback is not real-world repair hold in actual systems.
- Ethics, law, and political research: compare institutional paths for protection, standing, responsibility, authorization, legitimacy, and sovereignty, establishing which boundaries protect subjects and which designs instead produce new harms through power.
- Economic, social, cultural, educational, and religious research: provide longitudinal distributional data, field materials, cross-cultural counterexamples, studies of learning and dependence, interpretations from within traditions, and alternative institutions, so global scope can genuinely be changed by different experiences rather than merely broaden citation coverage.
- Institutional, engineering, and practical work: translate theory into observable pilots, standards, tools, and failure reports. If practical outcomes show a mechanism to be infeasible, excessively costly, or misdirected, theory and proposals should genuinely retreat, revise, or exit.

Moving Forward from Joint Research to Civilization Leap

Civilization Leap cannot be achieved by one theory, discipline, organization, or group of people who already agree. It needs different people, beginning with different questions, to bring facts, counterexamples, value judgments, institutional design, engineering capabilities, and practical experience into the same civilizational process. Some provide proof; others object. Some propose institutions; others demonstrate experimentally that those institutions need rewriting. Some build technology; others point out that it has crossed boundaries it should not cross.

The most valuable participation is professional work capable of changing the theory, rather than agreement. Further reading ends not with another manuscript, but with more people moving from seeing problems to independent judgment, then to joint research, communication, advancement, and practice, ultimately joining the real process of humanity's Civilization Leap.

The aim is not for more people to finish reading Civilization Leap, but for more people to accomplish Civilization Leap together.

Current public entry points: the complete Civilization Leap library and the public overview. Where a specific specialist paper or joint-research page has not yet been established, the library is the entry point; the website maintains dynamic versions, precise links, and subsequent participation routes. Civilization Leap complete library Public overview

Appendix A Index of Core Judgments and Evidential Status

Table A1 Current Status of Core Judgments

Judgment	Status	Explanation or route to review
Existing advanced AI has significant capabilities dependent on tasks and versions	A	Limited to specific evaluations, deployments, and times; not extrapolated to general reliability
Current public evidence is insufficient to establish AGI subjective experience	A (judgment about the state of evidence)	New architectures, multimethod causal evidence, and elimination of competing explanations may trigger review
Global workspace, integrated information, embodied approaches, and others may explain machine consciousness	B	Competing theories; no single approach has final adjudicative status
The three dimensions of subjecthood may serve as entry points for studying subjecthood in a strong sense	B/C	An analytical framework, not a definition confirmed by the scientific community
Meaning is foundational and has diverse sources	Philosophical postulate (A/B/C not applicable)	Examined through coherence, explanatory power, resilience to counterexamples, and practical consequences
Finitude actualizes known human meaning	Structural proposition (empirical bridging is C)	Extreme life extension, replication, heterogeneous subjects, and cross-cultural research may revise its scope
Capability, standing, and power require separate justification	Normative-institutional judgment	Continually tested through comparative law, governance performance, rights, and distributive consequences
Systems stepping back is a condition of civilizational maturity	Institutional hypothesis	Validated through exit exercises, service continuity, dependence risks, and net consequences

Appendix B Terminological Boundaries

Table B1 Boundaries of Core Terms

Term	Meaning in this paper
Age of AGI	A civilizational condition in which advanced intelligent systems and their developmental direction are already separating historical overlaps in advance; not a claim that universally recognized, fully realized AGI exists.

Term	Meaning in this paper
Subject of meaning	A subjective position capable of forming, experiencing, claiming, and bearing meaning; not currently established automatically by linguistic functionality or social recognition.
Collaborator in meaning	A system role that understands, protects, and transmits meaning and expands real subjects' capabilities within boundaries of authorization and responsibility.
Finite existence	An existential structure involving a subject's mutually exclusive choices, history, relationships, consequences, and continuity that cannot arbitrarily be transferred; not equivalent to constrained resources.
Subject continuity	The open question of whether first-person existence persists; distinct from informational, operational, causal, relational, and narrative continuity.
Protection	Proportionate measures reducing potential harm under insufficient evidence; not an automatic source of full personhood, ownership, resource rights, or political power.
Legal standing	Rights, duties, procedural positions, or agency positions allocated institutionally for particular purposes; not ontological proof of consciousness.
Political standing	Status within structures of public membership, representation, and participation; separately justified from legal personhood, function, and sovereignty.
Local indicators cannot settle whole lives	Rejection of combining local indicators into a final scale that settles entire lives or civilizational meaning; not rejection of bounded measurement with clear purposes.
System exit and stepping back	Systems reduce intervention, transfer capabilities, exit, or retire when tasks are complete, risks excessive, or real subjects can take over, while retaining residual responsibilities.

Appendix C Shared Interfaces and Resource Status

A phase does not create a source of authority; task urgency requires separate demonstration. Expiry stops new execution beyond ordinary authority; necessary handover must be bounded and verifiable. Nonerasure of history and the normative status of life's meaning remain distinct. E/C/O are research directions, not a joint threshold for protection. A/B/C/F indicate argumentative status, not a uniform confidence ranking.

Insufficient current evidence of subjecthood does not mean permanent denial; openness to the future does not mean present confirmation. Permission to deploy meaning-related functions, capability, and actual effects require separate judgments. Scientific facts cannot replace normative reasons, and normative direction cannot override professional facts.

This edition's theoretical baseline is synchronized through October 7, 2026. Empirical materials in individual papers retain their explicit search dates; regulations, working papers, drafts, clinical registrations, and corrections remain subject to specific-version assessment. Verification of key sources does not mean the entire bibliography has passed verification. For NCT07290244, the official registration has been confirmed as the first-in-human phase 1 safety and tolerability study of ER-100. This edition does not infer actual enrollment or efficacy from registration alone.

This edition checked the original Revah correction: the mPFC panel in Fig. 5a had mistakenly used the A1 panel. The correction concerns that panel, not withdrawal of every conclusion in the study. Assessment of Owen's doctrinal work must preserve the sample design involving deliberately inserted errors, rather than attribute inserted errors to contradictions naturally generated by the model. The corresponding volumes' references contain the original sources.

Routes for counterevidence: which genuine stake changed a purpose; who determined which relational weights; whether cooperation transferred harm to absent parties; and whether the old logic has a path that durably interrupts recursion while preserving conditions for trial and error. Tests of local methods should identify objects, conditions, and outcomes, rather than substitute for an assessment of the entire civilizational direction.

References

- [1] Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- [2] Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/S0140525X00005756>
- [3] Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450. <https://doi.org/10.2307/2183914>
- [4] Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- [5] Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- [6] Dehaene, S., Lau, H., & Kouider, S. (2017). What Is Consciousness, and Could Machines Have It? *Science*, 358(6362), 486–492. <https://doi.org/10.1126/science.aan8871>
- [7] Tononi, G., & Koch, C. (2015). Consciousness: Here, There and Everywhere? *Philosophical Transactions of the Royal Society B*, 370(1668), 20140167. <https://doi.org/10.1098/rstb.2014.0167>
- [8] Butlin, P., Long, R., Elmoznino, E., et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv:2308.08708. [Research report/preprint] <https://arxiv.org/abs/2308.08708>
- [9] Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- [10] Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford University Press.
- [11] Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The Organization of Living Systems, Its Characterization and a Model. *BioSystems*, 5(4), 187–196. [https://doi.org/10.1016/0303-2647\(74\)90031-8](https://doi.org/10.1016/0303-2647(74)90031-8)
- [12] Szathmáry, E., & Maynard Smith, J. (1995). The Major Evolutionary Transitions. *Nature*, 374, 227–232. <https://doi.org/10.1038/374227a0>
- [13] Queller, D. C., & Strassmann, J. E. (2009). Beyond Society: The Evolution of Organismality. *Philosophical Transactions of the Royal Society B*, 364(1533), 3143–3155. <https://doi.org/10.1098/rstb.2009.0095>
- [14] Gibson, D. G., Glass, J. I., Lartigue, C., et al. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 329(5987), 52–56. <https://doi.org/10.1126/science.1190719>

- [15] Hutchison, C. A. III, Chuang, R.-Y., Noskov, V. N., et al. (2016). Design and Synthesis of a Minimal Bacterial Genome. *Science*, 351(6280), aad6253. <https://doi.org/10.1126/science.aad6253>
- [16] Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- [17] Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., & Campbell, K. H. S. (1997). Viable Offspring Derived from Fetal and Adult Mammalian Cells. *Nature*, 385, 810–813. <https://doi.org/10.1038/385810a0>
- [18] Lancaster, M. A., Renner, M., Martin, C.-A., et al. (2013). Cerebral Organoids Model Human Brain Development and Microcephaly. *Nature*, 501, 373–379. <https://doi.org/10.1038/nature12517>
- [19] Trujillo, C. A., Gao, R., Negraes, P. D., et al. (2019). Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell*, 25(4), 558–569.e7. <https://doi.org/10.1016/j.stem.2019.08.002>
- [20] International Society for Stem Cell Research. (2021). ISSCR Guidelines for Stem Cell Research and Clinical Translation. *Stem Cell Reports*, 16(6), 1398–1408. <https://doi.org/10.1016/j.stemcr.2021.05.012>
- [21] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). Hallmarks of Aging: An Expanding Universe. *Cell*, 186(2), 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
- [22] Tronto, J. C. (1993). *Moral Boundaries: A Political Argument for an Ethic of Care*. Routledge.
- [23] Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human–Robot Interaction. *Engaging Science, Technology, and Society*, 5, 40–60. <https://doi.org/10.17351/ests2019.260>
- [24] Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press.
- [25] Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- [26] Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press.
- [27] Ostrom, E. (2010). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *American Economic Review*, 100(3), 641–672. <https://doi.org/10.1257/aer.100.3.641>
- [28] Pettit, P. (1997). *Republicanism: A Theory of Freedom and Government*. Oxford University Press.
- [29] Arendt, H. (1958). *The Human Condition*. University of Chicago Press.
- [30] Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. Pantheon Books.
- [31] Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press.
- [32] European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 22. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- [33] Court of Justice of the European Union. (2023). SCHUFA Holding (Scoring), Case C-634/21, Judgment of 7 December 2023. <https://curia.europa.eu/juris/liste.jsf?num=C-634/21>
- [34] European Union. (2024, consolidated 27 July 2026). Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- [35] Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225. <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treaty-num=225>
- [36] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

- [37] OECD. (2024). OECD AI Principles (updated May 2024). <https://oecd.ai/en/ai-principles>
- [38] United Nations General Assembly. (2024). Pact for the Future, Annex I: Global Digital Compact, A/RES/79/1. <https://undocs.org/A/RES/79/1>
- [39] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- [40] Polanyi, K. (1944). *The Great Transformation*. Farrar & Rinehart.
- [41] Sen, A. (1999). *Development as Freedom*. Alfred A. Knopf.
- [42] Acemoglu, D., & Johnson, S. (2023). *Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity*. PublicAffairs.
- [43] Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Roślaniec, K., & Troszyński, M. (2025). *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*. ILO Working Paper 140. <https://doi.org/10.54394/HETP0387>
- [44] UNCTAD. (2025). *Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development*. United Nations. <https://unctad.org/publication/technology-and-innovation-report-2025>
- [45] Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., et al. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
- [46] Miao, F., & Holmes, W. (2023). *Guidance for Generative AI in Education and Research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- [47] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- [48] Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press.
- [49] Lamport, L. (1978). Time, Clocks, and the Ordering of Events in a Distributed System. *Communications of the ACM*, 21(7), 558–565. <https://doi.org/10.1145/359545.359563>
- [50] Saltzer, J. H., & Schroeder, M. D. (1975). The Protection of Information in Computer Systems. *Proceedings of the IEEE*, 63(9), 1278–1308. <https://doi.org/10.1109/PROC.1975.9939>
- [51] Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. Basic Books.
- [52] Leveson, N. G. (2012). *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press.
- [53] Floridi, L., Cows, J., Beltrametti, M., et al. (2018). AI4People—An Ethical Framework for a Good AI Society. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- [54] United Nations. (2011). *Guiding Principles on Business and Human Rights*. https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr_en.pdf
- [55] Williams, B. (1973). The Makropulos Case: Reflections on the Tedium of Immortality. In *Problems of the Self: Philosophical Papers 1956–1972* (pp. 82–100). Cambridge University Press. <https://doi.org/10.1017/CBO9780511621253.008>
- [56] Ricoeur, P. (1992). *Oneself as Another*. University of Chicago Press.
- [57] Wolf, S. (2010). *Meaning in Life and Why It Matters*. Princeton University Press.
- [58] The Holy Bible, New American Bible, Revised Edition. Psalm 90. United States Conference of Catholic Bishops. <https://bible.usccb.org/bible/psalms/90>
- [59] Augustine. (1991). *Confessions* (H. Chadwick, Trans.). Oxford University Press.

- [60] Saṃyutta Nikāya 22.59, The Discourse on the Characteristic of Not-Self (Bhikkhu Sujato, Trans.). SuttaCentral. <https://suttacentral.net/sn22.59/en/sujato>
- [61] Dīgha Nikāya 16, The Great Passing Away (Bhikkhu Sujato, Trans.). SuttaCentral. <https://suttacentral.net/dn16/en/sujato>
- [62] Confucius. (2003). *Confucius: Analects, with Selections from Traditional Commentaries* (E. Slingerland, Trans.). Hackett.
- [63] Zhuangzi. (2020). *Zhuangzi: The Complete Writings* (B. Ziporyn, Trans.). Hackett.
- [64] Watts, V. (2013). Indigenous place-thought and agency amongst humans and non humans. *Decolonization: Indigeneity, Education & Society*, 2(1), 20–34.
- [65] Whyte, K. (2020). Too late for Indigenous climate justice: Ecological and relational tipping points. *WIREs Climate Change*, 11(1), e603. <https://doi.org/10.1002/wcc.603>
- [66] National Institute of Standards and Technology. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1. <https://doi.org/10.6028/NIST.AI.600-1>