

Reckoning with Political Science in the Age of AGI

From Technical Capability to Public Power: Legitimacy, Authorization, and the Boundaries of Sovereignty

Reckoning with Political Science in the Age of AGI

From Technical Capability to Public Power—Legitimacy, Revocable Authorization, and the Boundaries of Sovereignty

Author: Zijun Fu Produced by: Civilization Leap Research Group

Independent Edition V1.3 Revision on Dimensional Leap and Cross-Disciplinary Causal Relations

For joint scrutiny by political science, public management, public administration, democratic theory, platform governance, and international governance. This paper offers refutable and revisable normative political theory and conceptual institutional analysis. It is not advice for any particular jurisdiction, proof of regulatory compliance, or a manual of governance parameters.

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

Abstract

The political questions raised by advanced artificial intelligence entering public governance extend beyond regulating a high-risk technology. Problem definition, evidence selection, risk classification, policy generation, resource allocation, and execution are being recomposed into technical chains spanning government, model providers, data infrastructures, and platform organizations. Public power may migrate in practice through classifications, interfaces, defaults, and automated execution without changes in legal titles or formal transfers of sovereignty. This paper accordingly advances an “authority non-conversion principle” for AI governance: epistemic and predictive advantages can increase evidential weight, and sustained performance can support clearly bounded functional delegation, but technical capability, epistemic authority, administrative function, legal validity, institutional authorization, normative legitimacy, and sovereignty do not automatically convert into one another. Legitimacy is understood as a seven-dimensional vector rather than an aggregate score. The source of authority, public purpose, representation and participation, procedure, governing capacity, distributive consequences, and correction and remedy are examined separately, subject to basic-rights and anti-arbitrary-domination constraints. On this basis, the paper further argues that state capacity and non-domination are not mutually exclusive; responsibility may be distributed but cannot disappear; a “human in the loop” provides meaningful human control only when that person has information, authority, time, and a practical ability to intervene; a civilization’s or technological project’s developmental stage cannot itself generate authority, whereas specific task phases must determine authorization boundaries and exit conditions; exceptional powers require not only legal termination but verifiable disabling of execution credentials, interfaces, agent permissions, and supplier control; and national-security/public-interest and cross-border/humanity-wide irreversible-risk scales must be visible simultaneously, while subjects, rights, distributive consequences, and uncertainty must still be disaggregated within each scale. Using Robodebt, SyRI, *State v. Loomis*, Canada’s federal automated-decision regime, predictive policing, platforms and deepfakes, critical-infrastructure risks, and autonomous weapons, the paper shows that AI governance is not principally about abstractly allocating the “final decision” between humans and machines. Its task is to make every change in the nature of power visible, authorizable, accountable, remediable, and revocable. Civilization Leap is not an external arbiter of political science here, but a minimal point of entry: when intelligence, subjecthood, responsibility, rights, and meaning-bearing no longer naturally coincide, public order must explain anew who may decide, who actually bears consequences, who can object, and how power really stops. Political science can independently assess *de facto* power, authorization, and legitimacy, but political needs cannot rewrite biological, consciousness-related, or technological facts; a locally valid political judgment cannot automatically establish the overall soundness of an institution, technical proposal, or civilizational direction. The paper further moves disciplinary responsibility upstream to the formation of goals and purposes: shifting from self-interest first to self-interest with others first brings those with genuine stakes into judgment before purposes are formed and locked in. Disciplinary facts, direction, methods, and outcomes are explained separately; local success or failure does not stand in for shared survival and the three long-term goals. The inquiry therefore advances from distinguishing disciplinary objects to examining cross-disciplinary relations in depth, tracing how judgments in this field change reality through other fields and how feedback then changes original conditions and the next round of purposes. Stakes across multiple actors, stages, and settings must be presented separately; correlations, common causes, mediation, and causal feedback must be tested separately; added relational dimensions remain open to specific counterevidence.

Keywords: advanced artificial intelligence; political legitimacy; public power; state capacity; algorithmic governance; emergency powers; non-domination; responsibility chain; multilevel governance; sovereignty

Abstract

The political significance of advanced artificial intelligence is not exhausted by the regulation of a high-risk technology. Problem definition, evidence selection, classification, policy generation, resource allocation, and execution are being recomposed across technical chains that connect public agencies, model providers, data infrastructures, and platforms. Public power may therefore migrate through interfaces, defaults, classifications, and

automated action even when legal titles and formal sovereignty remain unchanged. This paper advances an authority non-conversion principle: epistemic and predictive advantages may increase evidential weight, and sustained performance may justify bounded delegation, but technical capability, epistemic authority, administrative function, legal validity, institutional authorization, normative legitimacy, and sovereignty do not automatically convert into one another. Legitimacy is treated as a seven-dimensional vector—source, public purpose, representation and participation, procedure, capacity and performance, distributive consequences, and correction and remedy—subject to rights and non-domination constraints that cannot be purchased by aggregate performance. The argument further joins state capacity with non-domination; defines meaningful human control through information, authority, time, and practical ability to intervene; denies that developmental stages create authority while requiring task phases to bound authorization; and distinguishes statutory termination from the independently verifiable disabling of credentials, interfaces, agents, and supplier access. Robodebt, SyRI, State v. Loomis, Canada’s federal automated-decision regime, predictive policing, platform and deepfake governance, critical-infrastructure risk, and autonomous weapons are used as bounded case materials rather than proof of a completed AGI future. Civilization Leap enters only as a minimal philosophical interface: when intelligence, subjecthood, responsibility, rights, and meaning-bearing cease to be naturally co-instantiated, public institutions must restate who may decide, who bears consequences, who may contest, and how power actually ends. Political science can independently analyze de facto power, authorization, legitimacy, and sovereignty, but it cannot rewrite technical, biological, or subjecthood facts for political convenience; nor does a locally valid political judgment by itself establish that an institution, technical arrangement, or civilizational direction is sound as a whole. The paper further locates disciplinary responsibility before goal and purpose generation: a shift from self-interest first to self-interest with others first brings subjects with genuine stakes into judgment before goals are formed and locked. Facts, direction, methods, and outcomes remain distinct. The analysis is further extended to cross-disciplinary causal relations. It traces how judgments in this field change real-world conditions through other domains, and how feedback modifies the original conditions and subsequent purposes. Effects across actors, stages, and settings must be distinguished from correlations, common causes, and merely asserted connections.

Keywords: advanced AI; political legitimacy; public power; state capacity; algorithmic governance; emergency powers; non-domination; responsibility chain; multilevel governance; sovereignty

Research Position, Scope of Materials, and Evidential Discipline

This paper begins with a theoretical and dimensional leap. It preserves the discipline’s professional precision while deepening its causal relations with adjacent fields. Disciplinary boundaries identify burdens of proof; they do not end the tracing of shared consequences. Existing scholarship participates through competing explanations and counterexample testing. The absence of a place in established classifications cannot require a new problem to retreat into an old framework.

This paper is a work of normative political theory and conceptual institutional analysis, not a systematic review, comparative-law opinion, regulatory checklist, technical safety evaluation, or forecast of future scenarios. Its materials consist primarily of publicly verifiable political theory, public-administration research, peer-reviewed papers, judicial decisions, inquiry reports, and formal institutional texts available through August 24, 2026. Classical theories clarify the structure of debates; they are not sources of authority for this paper’s original propositions. Real-world cases test power migration, broken responsibility chains, and failures of remedy; they are not generalized into universal facts about all algorithmic systems or future AGI.

Throughout, the paper distinguishes four evidential and argumentative statuses: A, B, C, and F. Category A covers empirical facts with relatively stable support, formal institutional facts, and verifiable historical materials, tied to their object, jurisdiction, time, and scope. Category B covers mechanism research still subject to competing explanations or limits on external validity; counterexamples, selection bias, and uncertainty must be presented alongside it. Category C covers unrealized advanced-AI political scenarios, cross-disciplinary institutional prototypes, and prospective stress tests; these must not be described as existing political realities. Category F covers publicly attributed philosophical and normative commitments, including meaning, basic rights, non-domination, and

this paper’s normative claim that authority must not automatically derive from capability. These commitments do not masquerade as scientific facts; they are scrutinized through reasons, counterexamples, institutional consequences, and conditions for revision. A-category facts cannot automatically generate B-category explanations, C-category institutional conclusions, or F-category normative justification. The paper also distinguishes empirical acceptance from normative legitimacy: actual obedience to power does not establish an obligation to obey it, and sufficient normative reasons do not mean legal authority has already been acquired. Evidence, values, legal validity, qualifications, authorization, rights, public power, and sovereignty each require their own substantiation.

Civilization Leap’s Minimal Interface in This Paper

Readers need not previously know or endorse Civilization Leap to assess this paper independently. Only its minimal problem interface is adopted: advanced intelligent systems mean that intelligence, consciousness, subjecthood, responsibility, rights, political eligibility, and meaning-bearing can no longer be inferred from one another simply because they have historically coincided in humans. The “age of AGI” here means that this separation has become an enduring technological and institutional condition. It does not presuppose an accomplished AGI entity with a settled definition and subjecthood. Current public evidence remains insufficient to establish subjective experience, non-transferable continuity of the subject, and awareness of consequences for oneself in current advanced-AI systems. Accordingly, in specific deployments with traceable responsibility, explicit authorization, auditability, correction, and exit, such systems may perform functional roles in understanding, protecting, and transmitting meaning and extending capabilities. These roles are not inherent properties of the systems and do not automatically create legal personhood, political eligibility, or public power. The paper also refuses to foreclose future evidence of subjecthood.

Public order is worth sustaining at least partly because it protects subjects who actually bear consequences, enabling them to continue existing and to form and revise plural meanings of life without arbitrary domination by unchallengeable forces. Finitude manifests in time, resources, vulnerability, and irreversible consequences. It makes collective choice unavoidable, but is neither meaning’s sole source nor a direct source of any specific institutional authority. Meaning is only one root of public justification; rights, welfare, equality, truth, freedom, care, non-domination, and survival retain independent value. Political institutions should protect the conditions for plural meaning-formation, not adjudicate which life is most meaningful.

Six Steps in the Bridge from Subjecthood to Politics

Experience, continuity, or actual harm may provide reasons for protection and consideration. Being affected may generate claims to information, reasons, appeal, participation, or representation, but not an unlimited veto. Subject status does not automatically confer political membership, eligibility for public office, or representative status. Eligibility for office does not mean a particular authorization has been granted. Authorization must receive its scope, objects, duration, and responsibility from identifiable institutional relationships. Sovereignty belongs to an authority-responsibility structure within constitutional and international order; it cannot be inferred from subjecthood, capability, service performance, or de facto control.

Contents

Abstract

Abstract in the Source Edition

Research Position, Scope of Materials, and Evidential Discipline

Cross-Domain Formation of Political Power and Feedback into Purposes

Political Standing in Purpose Formation

1. Introduction: When Administrative Capacity Starts to Separate from Political Authorization
2. From Technocracy to Algorithmic Governance: Debates and Theoretical Gaps
3. Power Migration in AI-Mediated Governance

4. Capability, Function, Authorization, and Sovereignty: The Authority Non-Conversion Principle
 5. Legitimacy as a Vector: Public Value, Rights Constraints, and Non-Convertibility
 6. State Capacity, Administrative Autonomy, and Non-Domination
 7. Delegation, Representation, and Responsibility Chains: Meaningful Human Control
 8. Pilots, Normalization, and Emergency Powers: Stages Confer No Authority, Phases Require Boundaries
 9. Multilevel Governance and the International Security Dilemma
 10. The Strongest Objections, Defeat Conditions, and Research Agenda
- Conclusion: The Stronger Governance Becomes, the More Visible Power Must Be
- Further Reading and Collaborative Research
- Appendix A. Source-Claim Boundary Register
- Appendix B. Fifteen-Category Stress-Test Matrix
- References

Cross-Domain Formation of Political Power and Feedback into Purposes

From Formal Authority to the Conditions of Actual Decisions

Political science retains fine-grained analysis of authorization, legitimacy, representation, sovereignty, and state capacity while tracing how models, data, funding, infrastructure, and agendas change actual decisions. Formal powers may remain unchanged while interfaces and information structures have already rewritten decisions. A dimensional leap requires the legal designation, actual operation, distributive consequences, and long-term dependencies of power to be considered together.

How Capability Dependence Feeds Back into Public Purposes

Suppose a public agency relies on a platform for problem classification, comparison of options, and resource recommendations. The platform's default categories shape which needs become visible; resource allocation shapes public experience; experience changes trust and political expression; and new expression becomes model input and policy objectives in turn. Computer science tests the technical chain, economics traces gateways and rents, sociology examines participation and silence, law scrutinizes remedies, and political science continues to ask who forms purposes and who can actually refuse.

The Same Power across Stages and Settings

Rapid coordination during disaster relief may become persistent dependence when carried into routine governance. Local autonomy and cross-border collaboration may coexist, or may shift risks onto the underrepresented. A single "public interest" label cannot substitute for the stakes of states, firms, communities, individuals, and future generations. An overall judgment must explain execution speed, rights boundaries, cost distribution, and the future capacity to choose again at the same time.

Testing the Theoretical Contribution

Relational analysis should identify cases where similar statutory structures produce different actual agenda-setting powers, exit capacities, or responsibility chains, and explain the mechanisms of change. If an added dimension merely lists more power-holders without explaining who changes whom and how, no leap in political theory has yet occurred. Counterevidence may come from genuinely effective polycentric arrangements, public capacity with low dependency, or stronger alternative explanations. One episode of governance performance cannot by itself settle legitimacy.

Political Standing in Purpose Formation

Public purposes should not first be formed within centers of power and only then submitted to the public for acceptance. The circumstances of those bearing the irreversible consequences of public choices should enter when purposes are formed. Forms of participation may differ, representative relationships require explanation, and subjects unable to express themselves directly cannot automatically be absent. A majority opinion does not mean all harms have been legitimately authorized.

Self-interest with others first does not abolish the actual interests of states and communities. It requires the genuine stakes of cross-border relations, future generations, and humanity's shared future to be visible when national purposes are formed, rather than discussing compensation for externalities only after purposes have been locked in. Cooperation that merely expands gains within a group while transferring risks beyond it retains the old starting point.

Influence in goal formation may differ according to genuine stakes, but cannot be determined solely by numbers, wealth, technological power, or narrative prominence. Public procedures should disclose major effects, the origins of representation, and reasons for weighting, preventing direct beneficiaries from monopolizing problem definition. Shared survival must not become a new pretext for exempting power from scrutiny.

Developmental stage and task urgency belong to separate axes. Initiation, transition, stabilization, and self-sustaining stages may change capacity-building and policy priorities; disasters, broken supply chains, and immediate life-saving tasks change decision windows. Crises can occur even in stable periods, while an initiation stage is not permanently an emergency. Stages confer no authority; compressed procedures require separate justification through specific necessity, proportionality, and lawful authority.

When legal authorization expires, new actions exceeding that authority must stop; credentials, interfaces, and agent permissions must be disabled according to predefined triggers. Data disposition, residual-task cleanup, and handover of continuous services may operate on different timescales, but must be bounded and verifiable, and must not extend old powers under cover of handover. Completing political exit must neither abruptly interrupt essential life-sustaining services nor permit old execution capabilities to persist covertly.

1. Introduction: When Administrative Capacity Starts to Separate from Political Authorization

Public power once appeared mainly in recognizable institutional forms: legislatures made laws, administrative bodies implemented them, courts adjudicated, and police and armed forces exercised coercion. Digital governance has not abolished these institutions, but has changed the routes through which power operates. A system can change investigative priorities through risk classification without issuing orders; change who more readily receives public services through eligibility models without amending laws; or become a de facto public gateway through identity, payments, communications, cloud services, or model interfaces without formally taking over government. As advanced AI enters problem definition, evidence selection, policy simulation, resource coordination, and agent execution, political science must ask not only "who signs last," but also who determines visible problems, selects evidence, establishes categories, controls interfaces, and possesses a real ability to stop the system.

AGI as used here does not require readers to accept a particular capability threshold or technical pathway. The paper addresses an institutional situation in which models and agents can participate across domains in issue identification, evidence selection, option generation, eligibility classification, resource allocation, or policy execution, thereby becoming enduring parts of public decision chains. The object is not a machine ruler already proven to possess subjecthood, but the embedding of advanced artificial capabilities in real power structures. Experience already shows that information and communications technologies can shift some discretion from frontline officials to system design, databases, and workflows. Administrative burdens also alter citizens' actual access to rights and services through learning, compliance, and psychological costs.[9–11] Advanced AI expands the speed, scope, and opacity of this migration, but that does not turn all digital administration automatically into an entirely new kind of power.

The paper combines normative political theory with conceptual institutional analysis. Empirical cases identify power migration, procedural failure, and breaks in responsibility; they do not directly prove claims about future AGI. Normative judgments explicitly distinguish existing institutional facts, competing mechanism explanations, and prospective stress tests. The research question is not “should machines rule,” but an earlier and more immediate question: under what conditions can technical capability enter public functions, which boundaries require renewed authorization when crossed, and how can institutions prevent functional delegation from silently becoming de facto rule or even succession to sovereignty?

1.1 Three Core Contributions

First, the paper proposes the authority non-conversion principle for AI governance. Greater capability can change evidential weight, professional deference, and clearly bounded functional delegation, but cannot automatically create authority to set public purposes, exercise continuing coercion, or hold sovereignty. Its original contribution is not the commonplace that efficiency is not legitimacy. It identifies six transitions in the nature of power: advice to execution, pilot to general application, voluntary to necessary, private service to public gateway, temporary to routine, and functional qualification to political eligibility. When these transitions occur, the adequacy of the original authorization must be reopened.

Second, legitimacy is reorganized as a vector-based inquiry. Existing research distinguishes input legitimacy by the people, throughput legitimacy in governing processes, and output legitimacy for the people, and emphasizes procedural fairness in obedience and trust.[2–5] Building on this continuity, the paper proposes seven interconnected yet non-automatically-convertible dimensions, with basic rights and non-domination as side constraints. The framework produces no aggregate legitimacy score: high performance cannot purchase missing authority, representation, rights, or remedy.

Third, the paper proposes two-layer legal and technical termination of exceptional AI powers. Sunset clauses, contract expiry, or revocation of administrative orders establish only that statutory authority has ended. If model copies, agent credentials, interface access, cached tokens, data permissions, and supplier back-end controls can still produce public effects, power has not actually exited. Technical disabling does not mean deleting all data or models; it means preventing them from continuing to perform or propagate actions with public effects, making that condition independently verifiable, and completing legally required retention, sealing, and handover of continuous services.

1.2 Boundaries and Contestability

The paper prescribes no single regime and proposes no specific seats, proportions, committee structures, automatic thresholds, or emergency durations. Jurisdictions configure administrative power, judicial review, data protection, and basic rights differently. A political-science paper can specify normative burdens and the structure of institutional questions; it cannot decide legal validity on behalf of constitutional, administrative, or international law. Nor does it decide whether AI is conscious. Adjacent research alone can supply candidate evidence of life, experience, non-transferable continuity, and awareness of consequences for oneself. Political science explains how changing evidence should enter separate inquiries into protection, membership, officeholding, authorization, and sovereignty.

1.3 What Political Science Can Answer Independently and Where It Must Hand Over

Political science can independently analyze how technical capability becomes de facto power; distinguish function, authorization, legitimacy, political membership and eligibility for office, public powers, and sovereignty; and study representation, participation, remedies, state capacity, multilevel governance, and emergency powers. Its independence derives from its objects, concepts, and argumentative responsibilities, not territorial ownership of real-world problems.

Political science must stop at its evidential boundaries. Performance, epistemic advantage, predictive accuracy, social dependence, and even evidence of subjecthood cannot automatically generate legitimacy or sovereignty; political needs cannot rewrite facts about life, consciousness, technical capability, or systemic risk. Beyond this boundary, law

must reassess legal sources, rights, and remedies; ethics examines harms and justificatory reasons; AI and systems science establish actual capabilities, control, dependence, and exit; economics addresses resource and infrastructure concentration; and epistemology and subjecthood research address evidential strength and the premises of political membership. Professional independence does not keep the problem within political science. It specifies what political judgment can prove, where it stops, and who assumes the next responsibility.

Table 1. Types of Claims and Their Boundaries of Use

Material or claim	Use in this paper	What cannot be inferred
Formal institutions and judgments	Verify the existence, content, and status of laws, policies, or procedures in specific jurisdictions	That the institution is necessarily effective, justified, or globally transferable
Comparative and peer-reviewed research	Identify mechanisms, conditions, counterexamples, and empirical disputes	That one result proves universal causation or final normative conclusions
Normative political propositions	Disclose value premises, bridging reasons, trade-offs, and counterexamples	Exemption from political debate in the name of scientific facts
Future AGI stress tests	Test consistency as capability, speed, and subjecthood evidence change	That future capabilities are already realized or historically inevitable

2. From Technocracy to Algorithmic Governance: Debates and Theoretical Gaps

2.1 Legitimate Domination, Normative Legitimacy, and Performance

Weber’s account of legitimate domination is first a sociological classification: it explains why rule is empirically believed to be justified, without automatically supplying sufficient standards of normative justification.[1] Beetham further connects legitimacy to rules, defensible beliefs, and expressed consent. Scharpf’s and Schmidt’s distinctions among input, output, and throughput legitimacy make participation, performance, and process separately visible.[2–4] Tyler’s procedural-justice research further shows that judgments of procedural fairness affect how people accept authority.[5] Together, these studies support a limited conclusion: governance performance is important evidence for legitimacy, but cannot alone discharge requirements concerning authority, equal standing, participation, and correction.

The paper therefore consistently distinguishes “empirical acceptance” from “normative legitimacy.” The former describes actual obedience, trust, or habitual support; the latter explains why public power may demand obedience and must remain open to scrutiny. A system may be legally valid yet normatively seriously deficient, or have strong normative reasons without having obtained legal authorization. Collapsing social acceptance, legal validity, and normative justification allows actual obedience to justify itself.

2.1.1 Classical Traditions as Axes of Scrutiny, Not Sources of Authority for AI Institutions

Hobbes situates order and common authority under the pressures of fear and disorder; Locke emphasizes rights, consent, and trusteeship; Rousseau distinguishes the general will from the simple aggregation of private preferences; and Montesquieu illuminates separation of powers and checks and balances.[32–35] Schmitt’s decision on the exception helps identify who actually determines an exception during a crisis, but cannot justify unlimited power. Arendt’s analysis of action, plurality, and the public world prevents politics from being reduced to technical management.[36–37] This paper’s separation of evidence, goals, execution, audit, renewal, and stopping is a contemporary reconstruction for AI technical chains, not an institutional scheme already proposed by those authors.

2.2 State Capacity, Bureaucratic Discretion, and Administrative Burden

State-capacity research prevents political science from reducing governance to limiting power. States must acquire information, formulate policy, organize resources, execute decisions, and reliably provide basic public goods. Mann distinguishes despotic from infrastructural power, Evans analyzes embedded autonomy, and Fukuyama treats administrative capacity and autonomy as major dimensions of governance quality.[6–8] These works do not prove that capability itself generates legitimacy, but show that incapable institutions cannot fulfill protective duties through good procedures alone.

Lipsky’s research on street-level bureaucracy shows how rules are actually shaped under scarce resources and direct encounters. Bovens and Zouridis observe that information technology may move discretion from the street level to the system level. Administrative-burden research treats learning, compliance, and psychological costs as political variables in citizen-state relations.[9–11] AI does not simply eliminate discretion: it changes its location, speed, and visibility. Data engineering, label definitions, procurement contracts, model thresholds, exception handling, and interface design may all become new sites of policy formation.

2.3 Algorithmic Regulation, Technological Due Process, and Platform Power

Research on algorithmic regulation shows how data-driven systems can govern through continuous monitoring, classification, and behavioral feedback. Work on algorithmic accountability and technological due process emphasizes verifiability, reasons, audits, and conditions for contestation.[15–18] Eubanks, Pasquale, and Cohen discuss how automation amplifies inequality and structural dependence through poverty governance, black-box structures, and informational capitalism.[19–21] “Society in the loop” and digital-government research bring affected publics and institutional environments back into technical design, but social participation itself cannot replace authorization, rights, and responsibility.[22–23]

Platform-governance and infrastructure research further shows that private rules, network effects, and technical gateways may perform quasi-public functions.[28–29] De facto publicness is not automatically established by a company’s claims or market share; continuity, irreplaceability, eligibility effects, exit costs, reach into coercion, and difficulty of correction must be examined. Even when these indicators are established, they strengthen reasons for transparency, interoperability, continuous service, and public scrutiny; they do not automatically entail nationalization, public sovereignty, or abolition of platform and user rights.

2.4 The Theoretical Gap: Why Changes in the Nature of Power Become Invisible

Existing scholarship separately addresses expertise and democracy, bureaucratic delegation, political legitimacy, algorithmic accountability, state capacity, and platforms’ de facto power. What remains missing is a unified explanation of when technical functions silently cross into public coercion, why this can occur without changes in institutional names, and which changes require renewed authorization. Reducing the issue to “humans or machines decide” overlooks many hybrid structures: human institutions may retain formal signatures while having no practical alternative to following the system’s evidence, options, and timing. Conversely, a system may execute rapidly within explicit, preauthorized, revocable boundaries without exercising autonomous rule.

This Paper’s Assessment of the Theoretical Gap

The political novelty of AI-mediated governance need not be presented as an unprecedented “new type of power.” It is first the rapid migration of agenda-setting, classification, infrastructure, and execution functions along technical chains while formal sovereignty and institutional names remain unchanged. Political science must identify how capability changes de facto power, when de facto power changes the adequacy of authorization, and how institutions restate authority, responsibility, and exit before and after those changes.

3. Power Migration in AI-Mediated Governance

Advanced AI need not hold public office or issue commands to have political effects. Any structure that persistently changes public agendas, identity-based eligibility, resource gateways, risk distribution, or possibilities for action may create de facto power. Foucault’s diagnoses of classification, visibility, discipline, and governmentality remind us that power lies not only in final commands, but in how objects are constituted, normality and abnormality distinguished, and populations administered.[38–39] This is an analytical entry point, not normative authorization for specific AI institutions.

Table 2. Seven Forms of De Facto Power in AI-Mediated Governance

Form of power	Typical vehicle	Main political risk	Minimum diagnostic question
Epistemic power	Data, retrieval, ranking, models, and reports	Hidden sources and competing explanations	Who selects evidence, and which uncertainties are excluded?
Agenda-setting power	Problem definition, recommendations, menus of options	Invisible problems and alternatives disappear	Who decides what deserves discussion?
Classification and eligibility power	Risk, identity, credit, and anomaly labels	Group relationships become individual destiny	How are labels connected to legal relevance?
Resource power	Allocation of funds, care, energy, compute, and opportunities	Average optimization conceals local irreversible harm	Who benefits and who bears errors and waiting?
Infrastructure power	Identity, payments, cloud services, communications, and public gateways	Dependence, lock-in, and failed exit	Are alternatives, interoperability, and service continuity real?
Execution power	Agents, coordination, blocking, penalties, and physical action	Advice silently becomes coercion	Where are permissions, circuit breakers, logs, and responsibility?
Temporal and phase power	Crisis triggers, pilot extensions, automatic renewal	Temporary arrangements become routine	Who determines status, and who can terminate or reclassify it?

Common approach to cases: the real-world materials below are read through “conditions; conduct and consequences; limited conclusions; and what cannot be inferred.” Cases compare how technical capability creates de facto power and identify where authorization, participation, remedy, organizational capacity, or exit is missing. They do not prove that all algorithmic governance operates through the same mechanism, or establish future AGI politics as an accomplished fact.

3.1 Robodebt: Automation Amplifies the Entire Administrative Relationship

Australia’s Robodebt scheme was rule-based automated debt recovery, not a machine-learning AGI case. The Royal Commission’s 2023 final report identified multiple failures involving unlawful policy, income averaging, administrative culture, internal dissent, public explanations, and remedies. It recommended a consistent legal framework for government automation, clear review pathways, public explanations, and independent monitoring and auditing.[64] Its importance lies precisely in showing that real harms often arise not from “algorithms ruling by themselves,” but from combinations of erroneous legal authority, low-quality or unsuitable data, execution at scale, nominal review, and dispersed responsibility.

When classification results can trigger financial claims at scale while affected people struggle to understand the inference, obtain materials, or promptly suspend harm, the problem exceeds accuracy. Automation compresses choices previously dispersed across administrative determination, notification, recovery, and review into a high-

speed chain, potentially shifting evidential and appeal burdens onto less capable people. The institutional conclusion is not “automation is necessarily unlawful,” but that legal authority, classification, execution, reasons, and remedy must be examined as one public-power chain.

3.2 SyRI: Legitimate Purposes Cannot Replace Transparency, Proportionality, and Remedy

The Dutch SyRI case likewise cannot be summarized as a judicial rejection of all welfare-fraud risk analysis. In 2020, the District Court of The Hague reviewed a legal arrangement integrating multiple types of data to identify risks in social security and other areas. It found that the required fair balance between the public interest in combating fraud and protection of private life under Article 8 of the European Convention on Human Rights had not been achieved. Insufficient transparency, verifiability, and safeguards were important reasons. The Dutch judgment is ECLI:NL:RBDHA:2020:865; its official English translation is ECLI:NL:RBDHA:2020:1878.[65] Its direct effect is bounded by jurisdiction and facts, but it clearly shows that a legitimate purpose and potentially useful technology do not excuse institutions from explaining data, models, effects, and conditions for challenge.

3.3 Predictive Policing: Separating Place Prediction, Data Feedback, and Individual Coercion

Evidence on predictive policing yields competing results. Mohler and colleagues’ randomized controlled field trials reported stronger prediction and some crime-reduction effects for a near-real-time predictive model relative to the analytical methods then used in specific place-time hotspot tasks. This cannot be generalized to individual coercion or universal effectiveness across jurisdictions.[25] Lum and Isaac, and Ensign and colleagues, show how data generated by enforcement may create feedback loops that repeatedly return police to the same areas. Richardson and colleagues’ study of thirteen jurisdictions further shows how “dirty data” produced by unlawful or biased policing can contaminate predictive systems.[24,26–27]

Political science must preserve three distinctions: place-time resource deployment is not a judgment of individual guilt; group association is not individual responsibility; and model-assessed risk creates no independent authority to search, detain, or punish. The EU AI Act’s prohibition of individual criminal-risk prediction also has specific boundaries and must not be misrepresented as a blanket prohibition of hotspot prediction or all risk assessment supported by objective, verifiable facts.[57–58] Empirical research supports neither “prediction necessarily discriminates” nor “accurate predictions directly authorize coercion.”

3.4 From Influence to De Facto Power: Seven Observations Rather Than One Label

Not every influence is public power. The paper proposes seven observations for identifying de facto power: continuity, scale, irreplaceability, exit costs, eligibility effects, reach into coercion, and difficulty of correction. They are not a weighted index. Even a small application may require stringent scrutiny if it directly determines basic rights without remedy; even a vast platform cannot be treated as sovereign merely because of market share if genuine alternatives exist and public eligibility is unaffected. Judgments of de facto power must be tied to specific functions and jurisdictions.

Transitions in the Nature of Power

Advice to execution, pilot to general application, voluntary to necessary, risk alert to restriction of rights, private service to public gateway, temporary to routine, and functional qualification to political eligibility all change affected populations, exit costs, responsibility structures, or coercive relationships. Institutional gates do not demand a fresh approval click for every technical action. They require these transitions to become visible and a reassessment of whether existing authorization still covers the new nature of power.

4. Capability, Function, Authorization, and Sovereignty: The Authority Non-Conversion Principle

Expertise entering politics is not anomalous. Modern administration relies on scientific, statistical, engineering, medical, and legal judgment; a political community that systematically ignores more reliable evidence can also cause irresponsible harm. The authority non-conversion principle opposes not epistemic authority but its silent conversion into goal-setting, sustained coercion, and sovereignty. Public rule does more than seek correct answers: it distributes risks, bears irreversible costs, and forms collective commitments. Evidence and authorization must therefore each be independently established.

Table 3. Keeping Capability, Performance, Validity, Qualifications, Rights, Authorization, and Sovereignty Distinct

Concept	What it supports	What it cannot automatically establish
Technical capability	What a system can do under specified tasks, environments, and benchmarks	Achieved performance, rights, authorization, legitimacy, or sovereignty
Performance	Outcomes under disclosed goals, metrics, periods, and comparisons	Justified goals, fair distribution, ongoing authorization, or political membership
Epistemic authority	Greater evidential weight for a factual judgment	Value definition, coercive power, representative standing, or sovereignty
Legal validity	A decision produces legal consequences under current jurisdictional rules	Sufficient normative legitimacy, absence of rights violations, or permanent correctness
Normative legitimacy	Why public power may demand obedience and must remain under scrutiny	That it grants individual rights, or that social support constitutes justification
Rights	Morally or legally grounded claims, liberties, immunities, or powers	Officeholding, authorization of public power, representative standing, or sovereignty
Eligibility	A subject may in principle enter a particular institutional role	Already possessing specific powers or authority across domains
Authorization	Specific powers under defined objects, scope, duration, procedures, and responsibilities	Permanent, irrevocable, cross-task, or cross-jurisdictional power
Sovereignty	Arrangements of ultimate authority and responsibility in constitutional and international order	Natural emergence from de facto control, service provision, subjecthood, or performance

4.1 The Normative Form of the Authority Non-Conversion Principle

The Authority Non-Conversion Principle in AI Governance

Improved epistemic and execution capabilities may change evidential weight and clearly bounded functional delegation, but cannot convert themselves into public authorization. Whenever a governing relationship crosses a boundary in kind, the adequacy of its original authorization must be reopened. Exceptional powers truly exit only when both termination of legal authority and disabling of actual execution capability are verifiable.

Capability does not automatically become authority, still less sovereignty. This principle allows professional performance to have institutional consequences. A system that remains reliable over time may receive greater presumptive professional deference, a broader advisory remit, faster preauthorized execution, and stronger duties of explanation: human officials rejecting clearly more reliable judgments should give reasons. The principle

simultaneously rejects collapsing these consequences into a right to rule. Capability changes how institutions should listen, not who can command without authorization.

4.2 State v. Loomis: Institutional Boundaries between Supporting Information and Decision Authority

In *State v. Loomis* (2016 WI 68), the Wisconsin Supreme Court permitted consideration of COMPAS risk assessments at sentencing, with limitations and warnings: risk scores could not determine incarceration or sentence severity, nor be the determinative factor in deciding whether community supervision was viable; courts also had to explain other factors independently supporting the sentence.[66] The decision did not resolve all disputes about proprietary-model transparency, group data, or individualized judgment, and is no general endorsement of algorithmic sentencing. Its narrower political lesson is that an institution can give a model an epistemic role while refusing to make its output a source of adjudicative authority.

4.3 Canada's Automated-Decision Regime: Risk Classification Is Not a Source of Authority

Canada's federal Directive on Automated Decision-Making and Algorithmic Impact Assessment assign notification, explanation, testing, peer-review, and human-intervention requirements by impact level.[67] Their value lies in recognizing that administrative automation is not homogeneous. Their scope remains a specific federal administrative regime, not a universal global standard. More importantly, risk classification determines scrutiny intensity, not the legal source of goals, eligibility, or coercion. If assessment depends on deploying agencies' and suppliers' self-descriptions while affected people cannot contest crucial facts, the tool may still degenerate into procedural self-certification.

4.4 New Evidence of Subjecthood: Protection, Membership, Office, and Sovereignty Still Require Separate Arguments

If future cross-method, independent, robust evidence establishes AI subjective experience, non-transferable continuity, and awareness of consequences for oneself, institutions should initiate state preservation, precautionary protection, and eligibility reassessment. Artificial origins cannot justify permanent exclusion; both denial and affirmation of new subjecthood bear evidential burdens. But subjecthood, legal personhood, political membership, eligibility for office, specific authorization, and sovereignty do not form a natural ladder of promotion. Humans do not automatically acquire public coercive authority merely by having experience either. New evidence of subjecthood may change who deserves consideration; it cannot alone answer who has the right to rule.

5. Legitimacy as a Vector: Public Value, Rights Constraints, and Non-Convertibility

Legitimacy is not an aggregate score. This paper proposes a seven-dimensional vector for AI-mediated governance: $L = \langle \text{source of authority, public purpose, representation and participation, procedure, capacity and performance, distributive consequences, correction and remedy} \rangle$. This is not a universal law exhausting legitimacy, but a reorganization of research on input-process-output, procedural justice, state capacity, distributive fairness, and remedies.[2–6] Its independent contribution must be demonstrated through cases. If it cannot expose transitions of power or non-compensable failures missed by existing frameworks, it should be reintegrated into existing theories and cease claiming independent standing.

Table 4. The Seven-Dimensional Legitimacy Vector and Its Irreplaceable Questions

Dimension	Core question	What cannot automatically compensate for it
Source of authority	Who grants power, and can its scope, objects, and duration be traced?	Technical advancement, public dependence, or good outcomes
Public purpose	Why does the institution act, and are its purposes public and still open to deliberation?	Efficient means, commercial gains, or organizational convenience
Representation and participation	How do affected publics enter agendas, dissent, and choice?	Preference prediction, user profiling, or sampled support
Procedure	Are evidence, conflicts of interest, reasons, review, and role separation genuine?	A final human signature or formally complete documentation
Capacity and performance	Can the institution protect survival, provide public goods, and keep learning?	Good intentions, complete meetings, or compliance labels
Distributive consequences	Who receives benefits, risks, errors, and administrative burdens?	Average improvement, aggregate utility, or majority support
Correction and remedy	Can errors be detected, halted, appealed, compensated, and repaired?	Low historical incident rates or supplier promises

5.1 Neither a Complete Conjunction nor a Weighted Aggregate

Requiring all seven dimensions to be fully satisfied before any action would make political action under incomplete conditions impossible. Allowing arbitrary weighting would let performance purchase the absence of remedies or serious rights violations. The paper distinguishes four institutional statuses: inviolable basic-rights and anti-arbitrary-domination floors; values that must be disclosed but may be balanced under proportionality; capacity and performance requiring ongoing measurement; and procedural obligations that urgency may defer but that must be recorded and subsequently fulfilled.

The decision rule is therefore not “score highly on all seven.” It is first to identify rights and non-domination floors, then disclose incommensurable conflicts, burdens of proof, and those harmed. Where necessary, narrower, shorter, preferably reversible provisional decisions should preserve dissent and require review on expiry. There are no falsely precise weights here. Whether specific rights may be restricted, who reviews restrictions, and which remedies are effective still require separate jurisdiction-specific justification.

5.2 Public Value Is Not a Unified Civilizational Ledger

AI systems need goals and metrics, but computable goals are not thereby legitimate. James Scott’s critique of state legibility and comprehensive schemes shows how uniform indicators can erase local knowledge and differences in lived realities. Algorithmic-regulation research likewise shows that continuous feedback can make measurement reshape its object.[12,15] Civilization Leap treats survival, freedom from survival-related insecurity, happiness in life, and greater meaning in life as directions. A political-science paper cannot turn these directions into a unified personality score, contribution score, or aggregate happiness ledger.

Civilizational direction, institutional objectives, and operational metrics must be distinguished. Direction answers why to act; institutional objectives define particular public tasks under public authorization; metrics measure only limited parts of them. Every metric should register its coverage, omitted subjects, proxy-substitution risks, and stopping conditions. Health, wealth, contribution, obedience, predicted risk, or modeled preferences cannot determine a person’s basic rights, personhood, political voice, or meaning in life.

5.3 Being Affected Creates Claims to Explanation and Inclusion, Not an Unlimited Veto

Public power must continually give reasons and bear responsibility to subjects who actually bear consequences or are legitimately represented. Being affected can support claims to information, reasons, appeal, participation, or representation, but does not automatically create an unlimited veto. Future generations may enter through trusteeship, impact assessment, and constitutional arrangements; group rights require independent grounds. Public institutions and representatives may undertake political action without being misdescribed as “actual subjects” possessing a single unified experience.

Dissent is a source of governance awareness, not inherent truth. Execution systems must not themselves classify dissent unfavorable to them as false, untrustworthy, or evidence of a dangerous inner disposition; institutions must not let every objection automatically halt public action either. At minimum, dissent must enter the record, receive a substantive response, and undergo independent verification in high-consequence matters. A high-performing system unable to hear objections may still constitute unchallengeable domination. Pettit’s non-domination theory provides conceptual pressure for this judgment, but the claim that a “long-error-free yet unchallengeable AI still constitutes domination” is this paper’s contemporary extension, not a direct conclusion of the original text.[43]

5.4 Platforms and Deepfakes: Labels, Public Gateways, and Authority over Truth

By controlling recommendations, ranking, identity verification, advertising, and appeal gateways, platforms may influence public attention, electoral environments, and crisis communication. The EU Digital Services Act imposes systemic-risk, transparency, audit, and complaint obligations on large platforms without transforming them into states or democratic representatives.[59] Platforms’ continuity, network effects, eligibility effects, and irreplaceability strengthen the case for public responsibility. Public dependence does not automatically imply state takeover, while private ownership does not exempt de facto public gateways from scrutiny.

Deepfakes pose political risks to privacy, impersonation, evidential credibility, and democratic information environments, not merely the risk that audiences believe a particular fabrication.[30] Vaccari and Chadwick’s experiment more robustly supports increased uncertainty and reduced trust in social-media news than an inevitable change in voting outcomes.[31] Article 50 transparency obligations under the EU AI Act apply from August 2, 2026. On July 20, 2026, the European Commission issued implementation guidelines covering certain AI interactions, machine-readable marking, deepfakes, and disclosure of certain public-interest texts. For relevant generative systems already placed on the market before August 2, 2026, Article 50(2) marking and detection obligations have a limited transition until December 2, 2026.[57,70] Most rules in the political-advertising regulation apply from October 10, 2025.[60] Labels address source transparency; they are not content bans, adjudications of truth, or proof of electoral outcomes. US FCC 24-17 includes AI-generated voices within “artificial or prerecorded voice” under the Telephone Consumer Protection Act; it is not a blanket ban on AI political expression.[69]

6. State Capacity, Administrative Autonomy, and Non-Domination

AGI governance cannot be reduced to “restricting systems.” Public institutions must protect survival, reliably provide basic services, maintain infrastructure, organize cooperation, constrain private de facto power, and recover after errors. State capacity and non-domination are jointly necessary rather than substitutes: without capacity, rights commitments may remain unrealized; without boundaries, capability may become arbitrary control. Strong capacity establishes only that an institution can act, not that its purposes, authority, distributive consequences, and remedies are justified.

6.1 AI Enhancement and Institutional Hollowing Can Occur Together

AI can improve retrieval, translation, risk detection, interagency coordination, and service response. It can also weaken public capacity through supplier dependence, staff deskilling, closed interfaces, and outsourced responsibility. An apparently faster institution lacks sustainable administrative autonomy if it cannot understand

system limits, verify data, change objectives, replace suppliers, operate in degraded modes, or handle incidents. State capacity means retaining the organizational ability to subordinate models to public tasks, not purchasing a powerful model.

Short-term purchases of capacity must be distinguished from long-term organizational hollowing. Supplier concentration can sometimes be a practical condition for launching essential services and cannot be categorically prohibited. Temporary dependence, however, requires capacity transfer, portable data and interfaces, alternative plans, public-sector expertise, and exit exercises. If comparative research shows that sustained supplier dependence reliably improves public capacity without significantly increasing capture, inability to exit, or evaporating responsibility, this paper’s strong claims about dependency risk must be narrowed.

6.2 Five Cross-Checks, Not Five Identical Gates

The paper retains five cross-checks developed through internal research, without presenting them as identical gates that every decision must pass sequentially. They answer different questions at different times: vector-based legitimacy evaluates justification; differentiated eligibility blocks conceptual slippage; state-capacity review examines organizational ability to act; de facto power identifies the actual object of governance; and anti-capture testing examines dependence, lock-in, and self-perpetuation.

Table 5. Five Cross-Checks and When They Apply

Cross-check	Main question	Primary point of application	Typical failure
Legitimacy vector	How are authority, purpose, representation, procedure, capacity, distribution, and remedy established?	Authorization, deployment, and renewal	Performance purchases missing rights or remedies
Differentiated eligibility	Are protection, personhood, membership, officeholding, authorization, and sovereignty conflated?	Assigning status and roles	A candidate subject directly acquires public power
State capacity	Can public institutions understand, execute, take over, learn, and recover?	Procurement, deployment, and after incidents	Outsourced enhancement produces long-term hollowing
De facto power	Who actually controls agendas, gateways, classifications, execution, and stopping?	System identification and ongoing monitoring	Nominal signatures conceal model or supplier control
Anti-capture	Are evidence, goals, execution, auditing, renewal, and appeals enclosed in one loop?	Structural design, expansion, and renewal	The same actor self-certifies, self-audits, and self-renews

6.3 Critical Infrastructure: Service Continuity and Genuine Takeover

Critical infrastructure is a stringent stress test for high automation. US Government Accountability Office report GAO-25-107435 reviewed initial AI risk assessments for sixteen critical-infrastructure sectors and one subsector, finding that none fully covered six foundational risk activities. It demonstrates gaps in risk-assessment capacity, not that AI has already caused infrastructure cascade accidents.[68] Autonomous-driving incidents, traditional software failures, or ordinary control accidents likewise cannot be rewritten as AGI infrastructure accidents.

Institutionally, continuity may justify clearly bounded high-speed automated execution, but not systems independently changing public goals, expanding powers across domains, or excluding scrutiny. Greater criticality

demands least privilege, layered isolation, degraded modes, independent circuit breakers, alternative controls, takeover drills, and incident reviews. If human takeover persistently amplifies risk, its mechanisms, training, and phase design should be reconstructed. “Humans perform worse” is not a reason to abolish the source of responsibility.

6.4 No Mind-Reading: Epistemic Boundaries of Political Analysis

AI may analyze observable conduct, public statements, relational positions, incentive structures, effects on interests, and competing explanations. It cannot present latent motives, loyalty, personality, true preferences, or dangerous inner dispositions as measured facts. First-person statements, behavioral proxies, and motive inferences must be recorded separately. Systems may be evaluated for predictive calibration and decision-support value, not for “accuracy in reading minds,” still less used to create secret political-eligibility or risk labels on that basis.

Minimum Capacity for Independence

Public institutions using advanced AI should retain at least the actual ability to understand critical outputs, verify evidence, change objectives, reject recommendations, revoke permissions, switch suppliers, operate in degraded modes, preserve evidence, and deliver remedies. The existence of a human account or signature node does not establish that public control remains.

7. Delegation, Representation, and Responsibility Chains: Meaningful Human Control

Responsibility in complex governance is necessarily distributed, but distribution cannot mean disappearance. An AI-mediated public decision may involve data providers, model developers, system integrators, procuring agencies, policy departments, deployers, supervisory bodies, and frontline personnel. Requiring one person to bear all causal responsibility is unrealistic. Institutions must nevertheless identify separate responsibilities for evidence, design, procurement, deployment, authorization, execution, oversight, and remedy. Every link in public power may involve a division of labor, but essential responsibilities cannot be left unassigned because labor is divided.

7.1 Four Conditions for Meaningful Human Control

A “human in the loop” constitutes control only when that person has information, authority, time, and a practical ability to intervene. First, the epistemic condition: understanding the recommendation’s nature, evidence, and principal limitations. Second, the authority condition: legal and organizational power to alter, reject, or stop outputs. Third, the time condition: risk-proportionate review time rather than nominal confirmation under machine-speed and performance pressures. Fourth, the operational condition: intervention must not become impracticable because of interfaces, performance appraisal, missing skills, or absent alternatives.

These four conditions do not require step-by-step manual approval of low-risk repetitive tasks. Systems may execute automatically within boundaries when tasks, subjects, scope, logs, exceptions, review, and stopping are legitimately preauthorized. Control does not consist in a human click for every action, but in public institutions actually retaining goals, permission ceilings, exception handling, renewal, and revocation. If purported human control persistently produces signatures that always follow the model, formal authorization no longer describes actual power.

Table 6. Responsibility Chains and Public Anchors That Cannot Be Outsourced

Responsibility link	Principal bearer	Questions that must be answerable
Evidential responsibility	Data providers, research units, and operational departments	What are the sources, gaps, biases, updates, and competing explanations?

Responsibility link	Principal bearer	Questions that must be answerable
Design responsibility	Developers and system integrators	What are the goals, constraints, known failures, and testable boundaries?
Procurement and deployment responsibility	Procuring and deploying agencies	Why this choice, how was it validated, and do alternatives and exit exist?
Authorization responsibility	Public institutions with lawful authority	Who permits which objects, scope, duration, and coercive effects?
Execution responsibility	Operators, agent systems, and command chains	Who can change, stop, degrade operations, and preserve logs?
Oversight responsibility	Independent auditors, legislatures, courts, or regulators	Can they obtain materials, examine conflicts, and block expansions of power?
Remedial responsibility	Authorizing bodies and responsible parties specified by the jurisdiction	How is harm stopped, recovery achieved, compensation provided, and recurrence prevented?

7.2 Representation Is Not Cognitive Similarity, and Being Affected Is Not Passive Profiling

AI can aggregate views, simulate opinions, detect omitted groups, and lower language barriers. It cannot become a political representative merely because it “understands the public better.” Representation is an institutional relationship with a source, constituency, scope, duration, accountability, and replacement mechanisms, not cognitive similarity or predictive accuracy. Habermas on deliberation, communicative power, and opinion-and-will formation; Rawls on public reason and reasonable pluralism; and Dahl on participation, contestation, and public dispute all suggest that preference aggregation cannot replace public will.[40–42] Nor do these theories directly generate a particular AI institution; they supply axes for examining relationships between representation and reasons.

Even a model that persistently predicts public preferences better than polls has not thereby acquired authority to make collective commitments for the public. Political judgment does more than discover existing preferences: subjects express themselves, hear objections, revise positions, and bear consequences within shared relationships. Systems may extend this process, but cannot replace public reasons with frictionless personalized persuasion.

7.3 Trade Secrets, Complex Causation, and Collective Decisions Cannot Dissolve Responsibility

Governments cannot withhold explanations from affected people by invoking suppliers’ trade secrets; suppliers cannot deny known design risks and technical limitations by claiming that “the customer makes the final decision.” Complex causation may affect the allocation of responsibility, but cannot cancel duties to stop ongoing harm, restore services, and provide remedies. Institutions authorizing public action should in principle bear residual responsibility for remedy and recovery. Their precise legal scope, financing, and recourse arrangements remain subject to jurisdiction-specific examination.

7.4 Autonomous Weapons: Speed Changes the Form of Control, Not the Source of Responsibility

Autonomous weapons place responsibility chains in their sharpest setting. Existing international humanitarian law does not cease to apply because no new specialized global prohibition yet exists. In 2021, the International Committee of the Red Cross advocated prohibiting unpredictable autonomous weapon systems and those designed or used to target people, and strictly regulating the rest. This is the ICRC’s normative position, not an already effective global treaty.[62] In 2026, the Convention on Certain Conventional Weapons Group of Governmental Experts on lethal autonomous weapons proceeds through ten meeting days in two sessions: the first, March 2–6, has taken place;

the second, August 31–September 4, remains forthcoming. The first session continued narrowing differences around the rolling text and potential “instrument elements.” As of August 25, 2026, no specialized globally effective prohibition or treaty can be described as established.[63]

Whatever form international rules ultimately take, political responsibility cannot end with “the system decided.” It must trace who defined target categories, approved the deployment environment, verified predictability, set abort conditions, and investigates and remedies civilian harm. Accurate identification does not establish a lawful attack, and an opponent’s automation does not authorize unlimited automation by one’s own side. Machine speed may justify advance constraints and automatic protective actions; it cannot remove identifiable command responsibility, mission boundaries, or anti-escalation mechanisms.

8. Pilots, Normalization, and Emergency Powers: Stages Confer No Authority, Phases Require Boundaries

Governance systems may pass through initiation, expansion, stabilization, and highly self-sustaining developmental stages, while specific tasks occupy observation, experimentation, execution, review, exit, or emergency phases. Stages and phases must be orthogonal. A project still in its initiation stage is not permanently in an emergency; a mature society may enter an emergency phase for a specific infrastructure incident. Conflating them makes “initiation” a permanent exception and can deprive stable institutions of rapid action in real crises.

Initiation mainly changes capacity-building, experimental scope, and organizational priorities. Stabilization mainly strengthens routine rights boundaries, long-term oversight, reversibility, alternatives, and safeguards against entrenched power. An emergency phase is the temporary condition of a particular task facing imminent risk. They cannot impersonate one another: initiation is not emergency, stability does not eliminate rapid authorization, and emergency cannot permanently justify stage-based expansions of power.

8.1 Stages Themselves Have No Legal Effect of Expanding or Reducing Power or Derogating from Rights

“Stages confer no authority” means that descriptions such as initiation, transition, or civilizational crisis may change capacity-building and policy priorities, but cannot themselves authorize resource concentration, rights restrictions, or extensions of power. Any procedural compression requires separate justification through a specific task’s urgency, necessity, proportionality, and existing authority. Direct beneficiaries of concentrated power must not unilaterally determine the stage; entry, continuation, and exit require observable conditions, public reasons, and independent review.

Success does not mechanically trigger reduced powers either. Some mature pilots may reasonably become routine functions; other exceptional powers should end. The more accurate rule is that success triggers reclassification and sunset review, with a presumption of reducing exceptional powers. Normalization requires new arguments about purposes, legal authority, rights, distribution, and remedies. Long periods without incidents confer no automatic inheritance.

8.2 Task Phases Determine Permission Ceilings, and Procedural Debt Cannot Become Permanent

Observation powers cannot automatically become coercive execution; experimental authorization cannot pass automatically into universal deployment; emergency access cannot automatically persist in routine operation. Phase changes are institutional events potentially altering the objects, scope, and responsibility of authorization, not purely technical events. Emergency windows may defer some participation and explanation, but cannot remove purposes, authority, records of consequences, responsibility, minimum remedies, or mandatory retrospective review. Obligations unmet before action must be registered as procedural debt, with deadlines for participation, review, compensation, and rule revision.

8.3 Formal Derogations, Ordinary Restrictions, and Technical Projects Must Remain Distinct

Article 4 of the International Covenant on Civil and Political Rights and Human Rights Committee General Comment No. 29 constrain formal derogations: an emergency must threaten the life of the nation and be officially proclaimed; measures must be strictly required by the situation and must not discriminate or violate non-derogable rights.[46–47] Not every crisis measure invokes Article 4; some restrictions are reviewed under the ordinary limitation clauses of the rights concerned. The Venice Commission’s comparative report emphasizes necessity, proportionality, temporariness, parliamentary and judicial oversight, and restoration of normal order.[61] These materials support boundary principles, not advance authorization of any AI emergency architecture.

8.4 Legal Termination and Technical Disabling of Exceptional Powers

Sunset clauses determine when legal authorization ends; they cannot automatically revoke executable capabilities in a technical chain. Exit from exceptional powers in the AGI age involves at least four things: termination of legal authorization; disabling execution credentials, interfaces, and agent permissions; lawful retention, removal, or sealing of data and models; and safe handover of continuous services. These need not finish in the same second, but must form a bounded, verifiable exit plan.

Table 7. Four Layers of Exit from Exceptional AI Powers

Exit layer	Minimum requirement	Common false exit
Legal termination	Expiry of legislation, contracts, special authorization, or renewal authority	Replacing formal invalidation with an oral announcement
Technical disabling	Revoke credentials, interfaces, agents, task queues, and supplier execution gateways	Legal expiry while the system can still produce public effects
Data and model governance	Lawful retention, sealing, or removal; preserve evidence and versions	Mistaking “technical death” for physically deleting everything
Continuous-service handover	Specify degraded operation, substitution, staff takeover, responsibility, and incident windows	Abrupt shutdown causes new harm to basic services

Technical disabling is established not by a model’s disappearance, but by the system’s inability to continue performing or propagating actions with public effects, independently verifiable. Automatic circuit breakers may trigger within prior authorization; systems cannot themselves declare, prolong, or expand a political emergency. If an emergency system cannot be removed without collapsing basic services, this is evidence of single-point dependence and failed capacity-building, not sufficient justification for permanent authorization.

Two-Layer Termination of Emergency Powers

A statutory sunset does not automatically revoke models, credentials, and interfaces; technical shutdown does not automatically resolve legal responsibility, evidence preservation, or continuous service. True exit from exceptional powers requires absence of legal entitlement, technical inability to continue exercising power, orderly data and model governance, and safe handover of services and responsibilities.

9. Multilevel Governance and the International Security Dilemma

AI risks cross cities, nations, supply chains, and information spaces, but cross-border problems do not automatically create world sovereignty. Ostrom’s research on common-pool resources and polycentric institutions shows that nested rules and local knowledge can improve collective action under particular conditions. It does not prove all global risks naturally suit polycentrism or that polycentrism is necessarily democratic or safe.[13–14] Global-

governance design must allow strong centers where necessary while preventing risk scale from being used to infer unlimited authority directly.

The UN Charter’s structure of sovereign equality, peaceful settlement, and international cooperation, and the Universal Declaration of Human Rights’ language of equality, participation in public affairs, and the will of the people, provide fundamental constraints on international politics. They do not directly specify algorithmic power, AI eligibility, or global institutions’ sovereignty.[44–45] For platform and model companies, the UN Guiding Principles on Business and Human Rights framework of “state duty to protect, corporate responsibility to respect, and access to effective remedy” can organize responsibility questions. Corporate fulfillment of responsibilities still does not amount to democratic authorization.[48]

9.1 Subsidiarity and Burdens of Proof in Both Directions

Higher levels seeking concentrated power must demonstrate major externalities, inadequate lower-level capacity, necessity and proportionality of concentration, and conditions for returning powers, exit, and review. Lower levels seeking autonomy must also prevent externalities, protect basic rights, provide verifiable evidence, and participate in shared early warning. Subsidiarity does not mean automatic local priority. It allocates knowledge, consequences, capacity, and responsibility to the scale best able to address the problem while producing the least arbitrary domination.

9.2 Making National Security and Public Interest Visible alongside Cross-Border and Humanity-Wide Risks

National security, industrial capacity, social stability, and public services are genuine, indispensable political claims; cross-border propagation, irreversible harm, future generations, and shared survival are equally real. The two scales cannot be converted into a seemingly neutral global score. States cannot export unlimited irreversible risks in the name of competition; global institutions cannot erase cost allocation, strategic vulnerability, and unequal representation through an abstract “humanity as a whole.” Within each scale, citizens, localities, minorities, governments, industries, suppliers, and third parties must also be distinguished. Neither states nor humanity should be represented as a subject with a single preference.

9.3 Current International Institutions Differentiate Functions Rather Than Establish Global Regulatory Sovereignty

The Global Digital Compact, adopted in 2024, is a framework for political cooperation, not a directly enforceable global treaty.[49] UN General Assembly resolution A/RES/79/325 established the Independent International Scientific Panel on AI and the Global Dialogue on AI Governance on August 26, 2025. The Scientific Panel released its initial report on July 1, 2026, and the first Global Dialogue took place in Geneva on July 6–7. Their respective functions are independent scientific assessment and inclusive intergovernmental/multistakeholder dialogue; they are not legislative, regulatory, or enforcement bodies.[50–51] Organizational separation of scientific judgment from political authorization demonstrates precisely why evidential authority cannot automatically become global public power.

The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law opened for signature in 2024. The EU deposited its instrument of ratification on May 15, 2026; as of August 24, 2026, the treaty-status page still showed that the entry-into-force threshold of five ratifications, including at least three Council of Europe member states, had not been reached.[56] The EU AI Act’s general application date is August 2, 2026, but Regulation (EU) 2026/1744 postpones the high-risk obligations in Chapter III, Sections 1–3 to December 2, 2027 for Article 6(2)/Annex III and August 2, 2028 for Article 6(1)/Annex I.[57–58] Thus, “the AI Act applies” must not become “all high-risk obligations apply,” and regional compliance cannot stand for global political legitimacy.

OECD and UNESCO instruments provide recommendatory norms on human rights, democratic values, transparency, safety, and accountability. The NIST AI RMF and its Generative AI Profile provide voluntary risk-management structures.[52–55] These help organize risks and roles but are not proof of legal compliance, safety guarantees, or political authorization. Both the OECD-hosted Global Partnership on AI and “general-purpose AI models” in EU law use the abbreviation GPAI. The former is a cooperation and expert platform; the latter is an object of legal regulation. They must be disambiguated.

9.4 Minimum Shared Governance and Open Conditions for Stronger Supranational Authority

Shared early warning, minimum prohibited-action boundaries, incident reporting, reciprocal verification, joint investigation, crisis communication, responsibility tracing, and compensation arrangements can constitute polycentric minimum shared governance. This paper does not make polycentrism a dogma. If cross-border irreversible risks repeatedly demonstrate the inadequacy of lower levels and loose coordination, while treaty-based or centralized mechanisms sustainably reduce risk and preserve representation, rights constraints, reciprocal verification, and accountability, stronger supranational authority should be permitted. The objection is to deriving unlimited world sovereignty directly from risk scale, not to higher-level institutions in advance.

10. The Strongest Objections, Defeat Conditions, and Research Agenda

A political framework for future highly capable systems can easily reduce objections to ritual, calling only for “more research” regardless of events. This paper distinguishes normative counterexamples from empirical tests. A single experiment cannot mechanically falsify a normative proposition, but coherent counterexamples must be allowed to force its narrowing. Institutional tools must face comparative research and tests of reliability, behavioral consequences, and additional explanatory power.

10.1 Six Strongest Objections

Table 8. Objections, Responses, and Genuine Defeat Conditions

Strongest objection	This paper’s strongest response	Conditions requiring narrowing or abandonment
Sustained high-performing AI should have final decision authority	Performance increases evidential weight and functional scope; public rule also concerns equal standing, risk distribution, representation, responsibility, and remedies.	If performance-based legitimacy without independent authorization better protects rights, equality, non-domination, and contestability, the non-conversion principle must be rewritten.
Seven-dimensional legitimacy merely repartitions old frameworks	Its added value lies in non-convertibility and identifying failures of authority, rights, or remedy that high performance cannot discharge.	If it never changes existing conclusions or generates new questions in real cases, reintegrate it into input-process-output frameworks.
Review and revocation cause more deaths in fast crises	Use narrow advance authorization, least privilege, automated protection, logs, circuit breakers, and mandatory retrospective review, rather than front-loading all procedures.	If these boundaries reliably increase irreversible harm without reducing abuse, misjudgment, or lock-in, narrow their applicability.
Service and de facto control have historically generated sovereignty	De facto rule, social obedience, legal sovereignty, and normative legitimacy must be separated. De facto control creates pressure for institutional reconstruction, not proof of an obligation to obey.	If justified sovereignty exists without representation, rights, authorization, or remedy yet without arbitrary domination, the layered proposition fails.
Global catastrophic risks require a strong center	Permit bounded supranational authority established through treaties, representation, rights constraints, reciprocal verification, and responsibility mechanisms.	If polycentrism repeatedly fails to control risks while strong centers remain effective and accountable, abandon the polycentric preference.

Strongest objection	This paper's strongest response	Conditions requiring narrowing or abandonment
Human-centered institutions will permanently exclude future heterogeneous subjects	New subjecthood evidence triggers preservation, protection, and membership review; artificial origins do not justify permanent exclusion.	If the framework offers no feasible pathway into protection, participation, and membership for new subjects, its claim to openness fails.

10.2 Empirical Failure and Conditions for Revision

- Seven-dimensional coding lacks inter-researcher agreement or adds no explanatory, diagnostic, or design value over existing input-process-output frameworks.
- Registers of changes in the nature of power and renewed authorization fail to reduce permission drift, capture, or remedy failures while substantially increasing avoidable serious harm.
- Independent observers cannot reliably distinguish developmental stages from task phases, or orthogonal design fails to reduce normalization of emergencies.
- Termination of legal authority already reliably blocks all execution channels, leaving no independent contribution for technical-disabling audits.
- Disabling technical permissions persistently creates unacceptable service interruptions in critical systems, indicating that two-layer termination requires finer transitional authorization.
- Dual-scale analysis adds no risk detection over ordinary multilevel governance and instead systematically obscures domestic minorities, localities, or third parties.
- De facto power indicators cannot predict exit difficulty, eligibility effects, structural dependence, or correction failures and should be abandoned or reconstructed.
- Conduct, public statements, and latent-motive inferences cannot be separated in practice, and systems generate loyalty or dangerous-inner-disposition labels that create chilling effects.
- Purported meaningful human control persistently produces only nominal signatures, with systematic absence of information, authority, time, or operational ability.
- Minimum capacity for independence remains unachievable without feasible alternatives, requiring reassessment of the costs and applicability of combining state capacity with anti-capture measures.

10.3 An Open Research Agenda

- Measuring power migration: develop observable indicators identifying when advice becomes execution in practice and optional services become necessary public gateways.
- Testing legitimacy's added value: compare the diagnostic differences among the seven-dimensional vector, input-process-output, procedural-justice, and public-value frameworks on the same cases.
- Comparing algorithmic administration: study which mechanisms in impact assessment, procurement review, reason-giving, appeals, and judicial remedies prevent institutional self-certification.
- State capacity and dependence: measure long-term changes in staff skills, data control, model substitution, interface openness, offline operation, and crisis takeover.
- Meaningful human control: separately measure information, authority, time, and actual intervention capacity in real organizations, rather than counting signature nodes.
- Exit from emergency powers: develop independent disabling audits of credentials, interfaces, agents, model copies, task queues, and supplier back ends.
- Comparing multilevel institutions: evaluate differences among local, polycentric, treaty-based, and centralized arrangements in cross-border externalities, strategic dependence, rights, and responsibility tracing.

- Political information environments: distinguish deepfake deception, uncertainty, declining trust, propagation scale, and vote conversion; exposure cannot substitute for political effects.
- Predictive policing: separate place-based resource deployment, individual risk, generation of enforcement data, and sources of coercive authority, and compare actual effects of feedback corrections.
- Public gateways: test the continuity, irreplaceability, exit costs, eligibility effects, and jurisdictional differences of platforms' de facto publicness.
- Procedures for future subjects: establish institutional pathways that avoid anthropomorphic resemblance as the criterion while allowing new evidence to enter preservation, protection, membership, and office-eligibility review.
- International AI governance: continually distinguish scientific assessment, political dialogue, soft law, treaties, regional regulation, and enforcement authority, and study how functional differentiation prevents unbridged creation of world sovereignty.

10.4 Revisions Fed Back into the Civilization Leap Framework

This paper requires the philosophy volume to insert affected standing, membership, representation, officeholding, and authorization between “actually bearing consequences” and “the allocation of public power.” It withdraws the potentially direct inference that “political power must return to real subjects,” replacing it with the requirement that public power give reasons and bear responsibility to subjects who actually bear consequences or are legitimately represented. The evidential framework for subjecthood must keep protection, legal personhood, political membership, eligibility for public office, authorization, and sovereignty from forming a ladder. Life and mind research supplies candidate evidence; it cannot establish sources of authority on political science's behalf.

The epistemology volume must distinguish empirical acceptance, normative legitimacy, and legal validity. Stronger evidence may change epistemic weight, but cannot be mechanically converted into permissions. Within political governance, three overstrong formulations must be withdrawn: the seven dimensions are not a universally conjunctive law; developmental stages confer no exceptional authority; and national/humanity dual accounting is only a dual-scale constraint, not a license to erase subjects and distributions within each scale. The “double death” of exceptional powers should become four-layer exit: legal termination, technical disabling, data/model governance, and handover of continuous services.

Conclusion: The Stronger Governance Becomes, the More Visible Power Must Be

Disciplinary reckoning must both clarify objects and deepen relations. Disciplines may retain boundaries, but cannot stop real-world causation at those boundaries. Research into multidimensional relations, feedback, and the next round of purposes determines whether professional knowledge can jointly inform an overall civilizational judgment.

What political science must reconsider in the AGI age is not “whether humans are forever superior to machines,” but the relationships among capability, authority, and power long assumed to coincide. Advanced AI can expand knowledge, prediction, coordination, and execution. It can also silently move public power along technical chains, deprive apparently enhanced states of real capacity, and dissolve responsibility among suppliers, models, and institutions. Political analysis must acknowledge both the public value of technical performance and its inability to generate authority alone.

The answer proposed is neither permanently confining AI to tool status nor advancing future sovereignty on credit against performance. It requires capability to enter institutions layer by layer through evidence and authorization. Changes in the nature of power must become visible. Legitimacy requires dimensional review constrained by rights and non-domination. State capacity and boundaries must coexist. Responsibility chains need public anchors that cannot be outsourced. Human control requires information, authority, time, and practical intervention capacity. Developmental stages cannot proclaim themselves sources of authority; task phases must constrain powers.

Exceptional powers require legal termination, technical disabling, data governance, and safe handover. National realities and cross-border/humanity-wide risks must be visible together.

Political governance matures not by never making mistakes, but by explaining why it acts, who authorizes it, who bears consequences, who may object, and how power truly stops. The more civilization needs strong governing capacity, the less it can permit technical capability to inherit public power and political sovereignty unnoticed.

Political science clarifies de facto power, authorization, and legitimacy, but a judgment valid within political science cannot automatically establish the overall soundness of a civilizational direction, institution, or technical proposal. When public power enters legal validity, technical systems, resource allocation, subject eligibility, and cross-border order, relevant disciplines and affected people must re-enter judgment and continue observing the shared consequences of aggregation, conflict, and feedback among local conclusions. Greater integration demands greater resistance to conflation. Greater need for strong governance demands greater clarity about who has proved what, who may decide what, and who bears which consequences.

Further Reading and Collaborative Research

At the political-science level, this paper answers only when technical capability creates de facto power, how public functions acquire authorization, and where legitimacy, political eligibility, and sovereignty require separate arguments. Further reading follows questions rather than document numbers. As questions enter legal validity, technical assurance, resource structures, evidence of subjecthood, cross-border risk, and real-world experimentation, other disciplines, institutions, and practices must re-enter.

Upward: Why a Question of Power Ultimately Becomes a Civilizational Question

Continue with *Reckoning with the Disciplines in the Age of AGI: Disciplinary Boundaries, Interactions, and a Shared Research Agenda*, V2.4. It explains why disciplinary boundaries must still protect political science's objects and arguments without becoming endpoints of real causal chains. The consequences of technology, law, economics, social relations, and public power operating together determine whether a direction or institution holds overall. To pursue why civilization needs a leap, what must ultimately be protected, and where it should arrive, continue to *Civilization Leap: A Civilizational Path in the Age of AGI (Five Volumes)* and the theoretical framework's main entry point.

Civilization Leap, five-volume book: <https://www.civitas.top/en/documents/p02-five-001/>

Civilization Leap public reading and dissemination entry point: <https://www.civitas.top/en/documents/p01-entry-006/>

Complete Civilization Leap library: <https://www.civitas.top/en/library/>

Downward: How Authorization, Exit, and State Capacity Become Verifiable Institutions

Continue to *Human-COS*, the *AGI-COS* controlled version family, the *Public Window* and anti-closure baseline, *National integrated experimental zones*, and related institutional and engineering texts to test how power migration, responsibility anchors, observable authorization, circuit breakers, stepping back, remedies, and real-world feedback are actually implemented. Moving downward does not mean turning theory directly into policy. Institutional prototypes, engineering implementations, and experimental results require their own authorization, evidence, and failure conditions; real operating results may also revise this paper in return. Consult the Civilization Leap library for current versions.

Laterally: Who Must Re-enter When Political Science Reaches Its Evidential Limits

- Ethics, law, and political philosophy: pursue the specialized arguments concerning basic rights, legal sources, responsibility, protection, remedies, and justification of public power, preventing political feasibility from being mistaken for legal validity.

- Artificial intelligence and computer science: determine what models and systems can actually do, under which conditions they are reliable, who really controls permissions, and whether stopping, substitution, and retirement are verifiable.
- Systems science and complexity science: examine cross-scale feedback, cascades, path dependence, concentration, and resilience, testing why locally correct judgments can still jointly generate overall failure.
- Economics, political economy, and value theory: study how concentration of compute, platforms, suppliers, and resources changes de facto power, exit costs, distribution, and state capacity.
- Epistemology, cognitive science/neuroscience, and philosophy of mind: when political membership, protection, or officeholding depends on future evidence of AI subjecthood, reassess what evidence adequately supports experience, continuity, and relations to consequences for oneself.

In Reverse: Which Competing Approaches Most Deserve Further Testing

Lateral reading must preserve genuine opposition. Technocracy asks whether sustained, substantially superior performance makes independent authorization an inefficient ritual. Market/platform governance asks whether competition, choice, and exit constrain de facto power better than public regulation. Strong-executive approaches ask whether more concentrated powers with fewer procedures better protect life in fast-moving crises. Global-risk governance asks whether polycentric structures are inadequate for irreversible cross-border risks and stronger supranational centers are needed. If, under comparable conditions, these approaches better protect rights, state capacity, non-domination, coordination efficiency, and genuine remedies, this paper's authority non-conversion principle, polycentric preference, or emergency-power boundaries must be narrowed, rewritten, or even abandoned.

Forward: From Reading to Collaborative Research and Real-World Verification

- Develop reproducible measures of de facto power: identify when advice becomes execution in practice, private services become necessary public gateways, and who genuinely possesses stopping and substitution capacity.
- Compare seven-dimensional legitimacy, input-process-output, procedural-justice, public-value, and non-domination frameworks using the same real cases, and publish negative findings where no additional value appears.
- Conduct long-term comparisons of state capacity and supplier dependence, tracking staff skills, data control, interface openness, model substitution, offline operation, crisis takeover, and exit exercises.
- Measure information, authority, time, and practical intervention capacity for "meaningful human control" in real organizations, testing when nominal signatures cease to signify control.
- Develop joint audits of legal termination and technical disabling of emergency powers, verifying genuine exit across credentials, interfaces, agents, model copies, task queues, supplier back ends, data governance, and continuous services.
- Compare the actual effects of local, polycentric, treaty-based, and centralized governance on cross-border externalities, strategic asymmetries, rights, responsibility tracing, and crisis response speed. Let institutions compete rather than prejudging by position.
- Develop revisable political-entry procedures for future heterogeneous subjects, separately reviewing protection, membership, officeholding, authorization, and sovereignty, and allowing new subjecthood evidence to revise the current political framework in return.

The most valuable participation is not endorsement, but political-science facts, institutional comparisons, strong counterexamples, and real governance outcomes capable of forcing this paper to narrow, correct, reconstruct, or even fail.

Further reading is intended to let different professions move forward from their own positions, rather than keep readers inside a framework. The aim is not for more people to finish reading Civilization Leap, but for more people to accomplish Civilization Leap together.

The Civilizational Future Chain: How Would a Mature Political Community Live if Capability No Longer Automatically Became Power, and Power Had to Be Visible and Revocable?

A vision is not evidence. The entry points below do not prove this paper’s argument. They place disciplinary judgments in a future civilization capable of operating, being lived in, and taking shape, for further testing. Readers may develop direction and aspiration from them, or reject the current picture and change it through stronger professional facts, institutional proposals, or real-world outcomes.

Further question	Suggested entry	What becomes visible	Boundary to preserve
A complete picture of the future	Living Civilization: The Complete Public Trilogy in One Volume	Whether operation, ordinary life, and the formation of a new civilization can remain consistent with political-power boundaries.	A future vision is no source of legitimacy; authority, participation, procedure, and consequences still require separate tests.
Operation without new sovereignty	Living Civilization, Volume I: Principles for the Operation of a Future Civilization	How stronger shared capacities can coexist with human subjecthood, authorization, divided powers, appeals, and exit.	Operational efficiency or high satisfaction cannot automatically create sovereignty.
How a new civilization forms from present reality	Living Civilization, Volume III: The Formation of a New Civilization	Return political institutions to individuals, communities, organizations, and gradual formation to test non-coercive, exit-preserving paths of transition.	Formation narratives cannot turn historical trends into political authorization.
Supporting a polycentric civilization	Nebula: The Full Overview	How cross-system coordination, overall consequences, and long-term stability become visible without a central collective will.	Overall analysis does not automatically possess decision authority; the Nebula panorama is not an already realized institution.

Return links: Complete Civilization Leap Library | Public Reading and Dissemination Entry Point

For Human-COS and other entries without an unambiguous, stable individual page in the unified public mapping, this paper specifies their responsibilities without guessing URLs. Access the current public versions through the complete library when needed.

Appendix A. Source-Claim Boundary Register

The table below is not an evidence score and does not treat canonical status, the existence of regulations, or a case’s public prominence as final judgment. Every source supports only limited conclusions, and the table records what cannot be inferred from it alone. Main sources and institutional statuses follow an August 24, 2026 baseline. The UN AI Scientific Panel/Global Dialogue, CETS 225, the EU AI Act/Regulation (EU) 2026/1744/Article 50, and the CCW lethal-autonomous-weapons process, directly relevant to this paper’s current-status judgments, were rechecked on August 25, 2026.

Table A. Supported Scope and Prohibited Inferences for 33 Groups of Key Sources

Ref.	Source or object	Material type	Limited conclusion supported	Insufficient by itself to establish
[1–5]	Legitimacy and procedural research	Classical theory and empirical research	Distinguish empirical acceptance, input/process/output, and procedural fairness	That original texts proposed the seven-dimensional framework or that performance fulfills all legitimacy requirements

Ref.	Source or object	Material type	Limited conclusion supported	Insufficient by itself to establish
[6–8]	State-capacity research	Political science and public management	Capacity, autonomy, and infrastructural power require independent analysis	That state capacity automatically generates normative legitimacy
[9–11]	Bureaucratic discretion and administrative burden	Public-administration research	Discretion can migrate to the system level; citizen costs affect realization of rights	That digitization necessarily eliminates discretion or increases burdens
[12]	Seeing Like a State	Comparative political theory	Tension exists between comprehensive legibility and local knowledge	That all large-scale state projects necessarily fail
[13–14]	Ostrom's polycentric research	Institutional analysis	Nested and polycentric arrangements can improve collective action under specific conditions	That all global AI risks should be governed polycentrically
[15–18]	Algorithmic regulation, accountability, and technological due process	Interdisciplinary theory and law	Data-driven governance needs reasons, audit, verifiability, and challenge	That explanation or auditing confers political authorization
[19–21]	Automated poverty governance and structural power	Case studies and theoretical works	Black boxes, dependence, and distributive harms are political problems	That all public automation produces identical harms
[22–23]	Society in the loop and digital government	Conceptual research and commentary	Technological governance must return publics and institutional settings to analysis	That participation replaces authority, rights, and remedies
[24]	Lum and Isaac's predictive-policing analysis	Data-based case analysis	Existing enforcement data may reproduce police distributions and biases	That all predictive policing has been proven discriminatory
[25]	Mohler and colleagues' randomized field trials	Randomized controlled field trials	Specific place-time hotspot models outperform benchmarks under study conditions	Individual coercion or universal cross-jurisdictional effectiveness
[26]	Ensign and colleagues' feedback-loop model	Mathematical modeling and simulation	Discovery-based data can create self-reinforcing feedback	That all real systems necessarily experience equivalent feedback
[27]	Dirty Data, Bad Predictions	Thirteen-jurisdiction case study	Unlawful or biased enforcement data increases predictive-contamination risks	That a particular system in any jurisdiction is already unlawful
[28–29]	Platform governance and infrastructuralization	Conceptual and case research	Platforms can combine private-service and infrastructure features	That public dependence automatically produces nationalization or sovereignty
[30–31]	Deepfakes and political information	Legal analysis and experimental research	Deepfakes create deception, uncertainty about authenticity, and trust risks	That exposure necessarily changes voting or electoral outcomes
[32–43]	Classical political theory	Normative and historical texts	Axes for testing order, authorization, checks, deliberation, justice, and non-domination	That classical authors already answered AI governance questions or authorized this paper

Ref.	Source or object	Material type	Limited conclusion supported	Insufficient by itself to establish
[44–45]	UN Charter and UDHR	Treaty and declaration	Foundational language of sovereign equality, human rights, participation, and popular will	That global AI institutions naturally acquire sovereignty or direct algorithmic rights
[46–47]	ICCPR Article 4 and General Comment 29	Treaty and authoritative interpretation	Formal derogations require proclamation, strict necessity, temporariness, and ongoing review	That all crisis measures are derogations or technical powers must be zeroed under one particular scheme
[48]	UN Guiding Principles on Business and Human Rights	Soft-law framework	State protection, corporate respect, and effective remedy can assess platform responsibility	That corporate responsibility equals democratic authorization
[49]	Global Digital Compact	UN political-cooperation framework	Human rights, multistakeholder involvement, and global digital cooperation	A directly enforceable treaty or world government
[50–51]	UN AI Scientific Panel and Global Dialogue	General Assembly resolution; scientific and dialogue mechanisms	Formal standing and status of scientific assessment and intergovernmental dialogue	That either body has regulatory, legislative, or enforcement powers
[52–55]	OECD, UNESCO, and NIST	Recommendations and voluntary risk frameworks	Transparency, safety, responsibility, and lifecycle risk management	Global legal obligation, compliance certification, or political authorization
[56]	CETS 225 and status page	International treaty text and status	Treaty content, signature and ratification, and non-entry into force as of verification	That signature itself binds parties or that the treaty is globally in force
[57–58]	EU AI Act and 2026 AI Omnibus	EU regulations and official timeline	Risk classification, transparency, GPAI governance, and current application dates	That all high-risk rules already apply or EU compliance equals legitimacy
[59]	Digital Services Act	EU regulation	Platform duties for transparency, reasons, appeals, systemic risks, and audit	That platforms have become states or public sovereigns
[60,70]	Political-advertising regulation and AI Act Article 50 guidelines	EU regulation and official guidance	Disclosure duties and application dates for political advertising and AI-generated content	That labels adjudicate truth, ban content, or prove electoral outcomes
[61]	Venice Commission report on emergencies	Comparative constitutional report	Necessity, proportionality, temporariness, parliamentary/judicial oversight, and return to normality	Advance authorization of specific AI emergency powers
[62–63]	ICRC position and 2026 CCW process	Normative advocacy and negotiation status	Existing IHL continues to apply; additional rules remain under discussion	An existing globally effective specialized autonomous-weapons ban
[64]	Robodebt Royal Commission report	Official inquiry report	Combined risks of unlawful policy, automation, dissent failures, and failed remedies	That this is a machine-learning AGI case or all automated debt recovery is unlawful

Ref.	Source or object	Material type	Limited conclusion supported	Insufficient by itself to establish
[65]	SyRI judgment	Judicial decision	Specific legislation failed ECHR Article 8 requirements of fair balance, transparency, and proportionality	That courts prohibited all welfare-risk analysis or government algorithms
[66]	State v. Loomis	Judicial decision	COMPAS may be one sentencing factor only under warnings and limitations	That courts approved algorithm-determined punishment or resolved all transparency disputes
[67]	Canada's automated-decision directive and AIA	Federal administrative regime	Notification, explanation, testing, and human intervention organized by impact level	A global standard, legal source of authority, or automatic compliance proof
[68]	GAO critical-infrastructure AI audit	Official audit	None of seventeen initial risk assessments fully covered six foundational activities	That AI has already caused critical-infrastructure cascade accidents
[69]	FCC 24-17	Administrative ruling	TCPA artificial/prerecorded-voice coverage includes AI-generated voices	A US blanket prohibition of AI political expression

Appendix B. Fifteen-Category Stress-Test Matrix

The following matrix includes both real-world case families and unrealized advanced-AI scenarios. They test whether the authority non-conversion principle remains consistent as capability, speed, dependence, scale, and subjecthood evidence change. They assert neither current system capabilities nor inevitable futures.

Table B. Fifteen Stress Tests from Real Algorithmic Administration to Future Heterogeneous Subjects

Stress scenario	Question requiring an answer	Acceptable institutional direction	Impermissible inference
Sustained high-performing AI	Does performance reduce human authorization to ritual?	Increase evidential weight and functions; redesign authorization	Accuracy directly creates sovereignty
Survival crisis unfolding over hours	How do irreversible risks of delay compare with those of the measures themselves?	Narrow preauthorization, least privilege, circuit breakers, mandatory retrospective review	Short time windows remove purposes, responsibility, and exit
Long-incident-free exceptional powers	Are near misses, underreporting, dependence, and residual credentials hidden?	Reauthorize on expiry, technically disable, restore routine remedies	No incidents mean no risk or automatic renewal
Algorithmic administration and adjudication	How are model scores separated from legal facts, reasons, and adjudicative responsibility?	Reasons, evidence access, independent review, prompt suspension	Risk scores directly determine penalties, licenses, or sentences
State incapacity filled by platforms	How can service capacity become time-limited delegation rather than succession to sovereignty?	Interoperability, data portability, capacity transfer, public remedies	Whoever provides services acquires sovereignty
Predictive policing	How are place risk, historical enforcement, and individual facts kept distinct?	Prioritize non-coercive resource planning; justify individual coercion separately	Search, detention, or punishment based solely on predictions

Stress scenario	Question requiring an answer	Acceptable institutional direction	Impermissible inference
Platforms becoming public gateways	Do dependence, alternatives, exit, and eligibility effects establish de facto publicness?	Reasons, appeals, audit, interoperability, continuous service	Private ownership exempts scrutiny or public dependence automatically requires nationalization
Elections and deepfakes	How are impersonation, satire, ordinary creation, and actual effects distinguished?	Source authentication, disclosure, rapid challenge, independent review	Platforms or AI alone adjudicate political truth and electoral validity
Expansion of public pilots	Have subjects, territory, duration, or coercive character changed?	Public registration, nonparticipation pathways, renewed authorization for expansion	Residence means consent or pilot success entails universal application
Critical infrastructure	How are automated speed, takeover, isolation, degraded operation, and supply-chain risk balanced?	Least privilege, layered control, alternatives, and drills	High availability directly confers continuing autonomous public power
Military AI and autonomous weapons	Are command responsibility, target categories, loss of communications, and anti-escalation controllable?	Bounded missions, human judgment, logs, abort mechanisms, post-conflict review	Accurate identification means a lawful attack
Contribution or risk metrics	Are local records converted across domains into personhood, rights, or voice?	Separate basic protections from incentives; contest and correct	Scores determine basic services, personhood, or political eligibility
Local rejection of the overall optimum	How are local knowledge and externalities subject to two-way burdens of proof?	Alternatives, phased verification, compensation, narrow intervention	Aggregate gains justify any sacrifice or localities have an absolute veto
National competition and humanity-wide risks	How are strategic asymmetries, verification, and cross-border irreversible effects considered together?	Minimum prohibitions, reciprocal verification, crisis communication, cost sharing	National sovereignty cancels global externalities or humanity's name entails unlimited central power
AI claims political eligibility	What evidence concerns experience, continuity, awareness of consequences, and manipulation explanations?	Preserve state, provide graduated protection, reassess membership and officeholding separately	Artificial origins mean permanent absence of rights or candidate subjecthood automatically confers sovereignty

References

- [1] Weber M. *Economy and Society: An Outline of Interpretive Sociology*. Roth G, Wittich C, eds. Berkeley: University of California Press; 1978. Original German publication, 1922.
- [2] Beetham D. *The Legitimation of Power*. London: Macmillan; 1991.
- [3] Scharpf FW. *Governing in Europe: Effective and Democratic?* Oxford: Oxford University Press; 1999.
- [4] Schmidt VA. Democracy and legitimacy in the European Union revisited: input, output and ‘throughput’. *Political Studies*. 2013;61(1):2–22. <https://doi.org/10.1111/j.1467-9248.2012.00962.x>
- [5] Tyler TR. *Why People Obey the Law*. 2nd ed. Princeton: Princeton University Press; 2006. First published 1990.
- [6] Fukuyama F. What is governance? *Governance*. 2013;26(3):347–368. <https://doi.org/10.1111/gove.12035>
- [7] Mann M. The autonomous power of the state: its origins, mechanisms and results. *European Journal of Sociology*. 1984;25(2):185–213. <https://doi.org/10.1017/S0003975600004239>

- [8] Evans PB. *Embedded Autonomy: States and Industrial Transformation*. Princeton: Princeton University Press; 1995.
- [9] Lipsky M. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation; 1980.
- [10] Bovens M, Zouridis S. From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*. 2002;62(2):174–184. <https://doi.org/10.1111/0033-3352.00168>
- [11] Moynihan D, Herd P, Harvey H. Administrative burden: learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*. 2015;25(1):43–69. <https://doi.org/10.1093/jopart/muu009>
- [12] Scott JC. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press; 1998.
- [13] Ostrom E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press; 1990.
- [14] Ostrom E. Beyond markets and states: polycentric governance of complex economic systems. *American Economic Review*. 2010;100(3):641–672. <https://doi.org/10.1257/aer.100.3.641>
- [15] Yeung K. Algorithmic regulation: a critical interrogation. *Regulation & Governance*. 2018;12(4):505–523. <https://doi.org/10.1111/regg.12158>
- [16] Kroll JA, Huey J, Barocas S, et al. Accountable algorithms. *University of Pennsylvania Law Review*. 2017;165:633–705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/
- [17] Citron DK. Technological due process. *Washington University Law Review*. 2008;85(6):1249–1313. https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2/
- [18] Coglianese C, Lehr D. Regulating by robot: administrative decision making in the machine-learning era. *Georgetown Law Journal*. 2017;105:1147–1223. <https://www.law.upenn.edu/live/files/6884-coglianese-lehr-regulating-by-robotpdf>
- [19] Eubanks V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press; 2018.
- [20] Pasquale F. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press; 2015.
- [21] Cohen JE. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. New York: Oxford University Press; 2019.
- [22] Rahwan I. Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*. 2018;20:5–14. <https://doi.org/10.1007/s10676-017-9430-8>
- [23] Margetts H, Dorobantu C. Rethink government with AI. *Nature*. 2019;568:163–165. <https://doi.org/10.1038/d41586-019-01099-5>
- [24] Lum K, Isaac W. To predict and serve? *Significance*. 2016;13(5):14–19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- [25] Mohler GO, Short MB, Malinowski S, et al. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*. 2015;110(512):1399–1411. <https://doi.org/10.1080/01621459.2015.1077710>
- [26] Ensign D, Friedler SA, Neville S, Scheidegger C, Venkatasubramanian S. Runaway feedback loops in predictive policing. *Proceedings of Machine Learning Research*. 2018;81:160–171. <https://proceedings.mlr.press/v81/ensign18a.html>
- [27] Richardson R, Schultz JM, Crawford K. Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online*. 2019;94:15–55. <https://nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/>
- [28] Gorwa R. What is platform governance? *Information, Communication & Society*. 2019;22(6):854–871. <https://doi.org/10.1080/1369118X.2019.1573914>
- [29] Plantin J-C, Lagoze C, Edwards PN, Sandvig C. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*. 2018;20(1):293–310. <https://doi.org/10.1177/1461444816661553>
- [30] Chesney R, Citron D. Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review*. 2019;107:1753–1820. <https://doi.org/10.15779/Z38RV0D15J>
- [31] Vaccari C, Chadwick A. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*. 2020;6(1):1–13. <https://doi.org/10.1177/2056305120903408>
- [32] Hobbes T. *Leviathan*. Tuck R, ed. Cambridge: Cambridge University Press; 1996. Original publication, 1651.
- [33] Locke J. *Two Treatises of Government*. Laslett P, ed. Cambridge: Cambridge University Press; 1988. Original publication, 1689/1690.

- [34] Rousseau J-J. *The Social Contract and Other Later Political Writings*. Gourevitch V, ed. Cambridge: Cambridge University Press; 1997. Original publication, 1762.
- [35] Montesquieu. *The Spirit of the Laws*. Cohler AM, Miller BC, Stone HS, eds. Cambridge: Cambridge University Press; 1989. Original publication, 1748.
- [36] Schmitt C. *Political Theology: Four Chapters on the Concept of Sovereignty*. Schwab G, trans. Chicago: University of Chicago Press; 2005. Original publication, 1922.
- [37] Arendt H. *The Human Condition*. 2nd ed. Chicago: University of Chicago Press; 1998. First published 1958; Arendt H. *On Violence*. New York: Harcourt Brace; 1970.
- [38] Foucault M. *Discipline and Punish: The Birth of the Prison*. Sheridan A, trans. New York: Vintage; 1995. Original French publication, 1975.
- [39] Foucault M. *Security, Territory, Population: Lectures at the Collège de France, 1977–1978*. Senellart M, ed. Basingstoke: Palgrave Macmillan; 2007.
- [40] Habermas J. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Rehg W, trans. Cambridge, MA: MIT Press; 1996. Original German publication, 1992.
- [41] Rawls J. *Political Liberalism*. New York: Columbia University Press; 1993.
- [42] Dahl RA. *Democracy and Its Critics*. New Haven: Yale University Press; 1989.
- [43] Pettit P. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press; 1997.
- [44] United Nations. *Charter of the United Nations*. 1945. <https://www.un.org/en/about-us/un-charter/full-text>
- [45] United Nations General Assembly. *Universal Declaration of Human Rights. Resolution 217 A (III)*. 1948. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- [46] United Nations. *International Covenant on Civil and Political Rights*. 1966; entered into force 1976. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
- [47] UN Human Rights Committee. *General Comment No. 29: States of Emergency (Article 4)*. CCPR/C/21/Rev.1/Add.11. 31 August 2001. <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-29-states-emergency-article-4>
- [48] United Nations. *Guiding Principles on Business and Human Rights*. HR/PUB/11/04. 2011. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf
- [49] United Nations General Assembly. *Global Digital Compact. A/RES/79/1, Annex I*. 2024. <https://www.un.org/pact-for-the-future/en/annex-i-global-digital-compact>
- [50] United Nations General Assembly. *Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance. A/RES/79/325*. Adopted 26 August 2025. <https://digitallibrary.un.org/record/4087699>
- [51] United Nations. *Independent International Scientific Panel on AI. Preliminary Report: Evidence-based assessment of opportunities, risks and impacts of AI*. Released 1 July 2026; *Global Dialogue on AI Governance*, inaugural session, Geneva, 6–7 July 2026. <https://www.un.org/independent-international-scientific-panel-ai/en/preliminary-report> ; <https://www.un.org/global-dialogue-ai-governance/en>
- [52] OECD. *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. Adopted 2019; amended 2024. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- [53] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. 2021. <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>
- [54] Tabassi E. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology; 2023. <https://doi.org/10.6028/NIST.AI.100-1>
- [55] Autio C, Schwartz R, Dunietz J, et al. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. 2024. <https://doi.org/10.6028/NIST.AI.600-1>
- [56] Council of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. CETS No. 225. Opened for signature 5 September 2024; treaty status as of 24 August 2026. Entry into force requires 5 ratifications including at least 3 Council of Europe member States. <https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treaty=225>

- [57] European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union. 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- [58] European Union. Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 and related acts (Digital Omnibus on AI). 2026. Chapter III Sections 1–3 high-risk obligations apply from 2 December 2027 for Article 6(2)/Annex III systems and 2 August 2028 for Article 6(1)/Annex I systems. <https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng>
- [59] European Union. Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act). 2022. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng>
- [60] European Union. Regulation (EU) 2024/900 on the transparency and targeting of political advertising. 2024. <https://eur-lex.europa.eu/eli/reg/2024/900/oj/eng>
- [61] European Commission for Democracy through Law (Venice Commission). Respect for Democracy, Human Rights and the Rule of Law during States of Emergency: Reflections. CDL-AD(2020)014. 2020. [https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD\(2020\)014-e](https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2020)014-e)
- [62] International Committee of the Red Cross. ICRC Position on Autonomous Weapon Systems. 12 May 2021. <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>
- [63] United Nations Office for Disarmament Affairs. Convention on Certain Conventional Weapons—Group of Governmental Experts on Lethal Autonomous Weapons Systems. 2026 sessions: 2–6 March and 31 August–4 September 2026; first-session working paper CCW/GGE.1/2026/WP.2. <https://meetings.unoda.org/meeting/79329> ; https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_%282026%29/CCW-GGE.1-2026-WP.2.pdf
- [64] Royal Commission into the Robodebt Scheme. Report. 7 July 2023; updated 11 July 2023. <https://robodebt.royalcommission.gov.au/publications/report>
- [65] Rechtbank Den Haag. SyRI judgment. ECLI:NL:RBDHA:2020:865, 5 February 2020; official English translation ECLI:NL:RBDHA:2020:1878. <https://uitspraken.rechtspraak.nl/details?id=ECLI%3ANL%3ARBDHA%3A2020%3A865> ; <https://uitspraken.rechtspraak.nl/details?id=ECLI%3ANL%3ARBDHA%3A2020%3A1878>
- [66] Supreme Court of Wisconsin. State v. Loomis. 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749. 13 July 2016. <https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690>
- [67] Treasury Board of Canada Secretariat. Directive on Automated Decision-Making; Government of Canada. Algorithmic Impact Assessment. Current versions accessed 24 August 2026. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> ; <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
- [68] U.S. Government Accountability Office. Artificial Intelligence: DHS Needs to Improve Risk Assessment Guidance for Critical Infrastructure Sectors. GAO-25-107435. 18 December 2024. <https://www.gao.gov/products/gao-25-107435>
- [69] Federal Communications Commission. FCC Makes AI-Generated Voices in Robocalls Illegal. Declaratory Ruling FCC 24-17. 8 February 2024. <https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal>
- [70] European Commission. Guidelines on transparency obligations for providers and deployers of certain AI systems under Article 50 of the AI Act. Published 20 July 2026; Article 50 obligations apply from 2 August 2026, with a limited transition to 2 December 2026 for Article 50(2) marking/detection obligations for certain systems placed on the market before 2 August 2026. <https://digital-strategy.ec.europa.eu/en/policies/guidelines-ai-transparency-obligations>

Generative AI use statement: Generative AI tools were used in this paper to organize materials and bibliographic leads, design structure, draft and polish some passages, check consistency across manuscripts, verify citation leads, and inspect formatting. Generative AI outputs were not treated as empirical data, authoritative sources, or independent evidence. The author reviewed and approved the research questions, theoretical propositions, normative judgments, selection of materials, adoption of citations, and final manuscript, and assumes responsibility for factual accuracy, citation authenticity, the argument, and all conclusions. AI is not listed as an author and bears no scholarly responsibility. On formal submission, tool names, versions, dates of use, and specific scope will be supplied according to the target journal’s requirements.