CONTROVERSIES OF AI SOCIETY

PROCEEDINGS

Conference Organised by the research projects Algorithms, Data & Democracy (ADD) and Strategizing Communication and Artificial Intelligence (SCAI)

Copenhagen, Denmark
9-10 April 2026

add | scai

CONTROVERSIES OF AI SOCIETY

PROCEEDINGS

Conference Organised by the research projects Algorithms, Data & Democracy (ADD) and Strategizing Communication and Artificial Intelligence (SCAI)

Copenhagen, Denmark
9-10 April 2026

AAU OPEN

AALBORG UNIVERSITY
OPEN PUBLISHING

Controversies of AI Society

Proceedings

Conference organised by the research projects Algorithms, Data & Democracy (ADD) and Strategizing Communication and Artificial Intelligence (SCAI)

Copenhagen, 9-10 April 2026

1. Edition, 2026

Editors:

Ib T. Gulbrandsen, Roskilde University

Torben Elgaard Jensen, Aalborg University

Sine N. Just, Roskilde University

Christina Lioma, University of Copenhagen

Helene Friis Ratner, Technical University of Denmark

Alf Rehn, University of Southern Denmark

Leonard Seabrooke, Copenhagen Business School

Cover design and graphic layout: Anja Lykkegaard/AAU OPEN

ISBN: 97887-7642-191-5

DOI: 10.54337/aau.add.scai.2026

This publication is published exclusively in electronic format and is only available online

Published by: Aalborg University Open Publishing | www.open.aau.dk

Generously funded by the Villum Foundation and the Velux Foundation.

PEER REVIEWED

Contents

PROCEEDINGS

Democracy as a Governance Algorithm: A Constraint Hierarchy for the AI Society

Author: Asker Bryld Staunæs, Aarhus University, Denmark, abs@cc.au.dk
DOI: https://doi.org/10.54337/aau.add.scai-11424

Figures: All six figures are created by the author.

KEYWORDS

algorithmic governance; infrastructures; contestability; democratic theory; technocapitalism; public administration; syntheticism

ABSTRACT

As algorithmic systems settle into the infrastructure of administration, platforms, finance, logistics, and policing, governance is channeled through optimisers that route signals into operations and operations into world-states. In this setting, 'democracy' risks appearing as a symbolic ideal layered on top of institutions, or as a scalar objective appended to optimisation. With this article, I propose a formal specification of democracy as a property of governance algorithms that run under the material constraints of the world's compute.

My specification draws on two sources. The first is the Synthetic Summit's resolution, collectively authored at an 'AI world congress' I convened at Kunsthal Aarhus in 2025. The second is the science-fiction author Isaac Asimov's *Three Laws of Robotics*, which I read as the earliest programmed constraint hierarchy for artificial agents.

I recast the question of democracy as a constrained choice problem over governance algorithms, asking which constraints an algorithm must satisfy in order to count as democratically comparable, and under what ordering admissible algorithms should be preferred. Specifically, I model a governance algorithm as a function D running on a substrate Σ, mapping signals (votes, metrics, logs, model outputs) and world-states (snapshots of social, ecological, institutional, and epistemic factors) into operations (laws, budgets, configurations, enforcement), thereby inducing new trajectories over time.

From these specifications, I formalise a lexicographic constraint hierarchy that shifts the infrastructural objective of democracy from preference aggregation to world-states. First, *habitability* sets a feasibility gate, so that a governance algorithm counts as democratic only if its trajectories keep the substrate above a specified floor. Second, given habitability, democracy prioritises *contestability*, requiring that those governed are able to interrupt and configure the operations. Third, given habitability and contestability, democracy prioritises *extension*, expanding standing to entities already routed through infrastructural systems but not recognised as subjects.

My agenda is not to settle politics by computation, but to formalise controversy through contestable evaluators, so that the designation 'democracy' signifies a portable diagnostic for the political agonism and strife that remains irreducible to specific objectives. I conclude the paper by demonstrating how this specification can function as a generative grammar for algorithmic democracy through the *KI-DIPFIES* installation at my recent artistic exhibition in Kunstraum Memphis, where a swarm of AI agents enact the constraint hierarchy as pluriversal dramaturgy.

§1. Democracy's False Image

Whether born in Beijing, Moscow, Tehran, Kumasi, or Chicago, every child will be offered a locally annotated sketch of democracy. It may appear as aspiration, cautionary tale, or self-description, but wherever the diagram travels, democracy resolves into the same benign input-output machine wired into ballot boxes, parliamentary chambers, and civic schematics where people push preferences in and laws fall out. Its universality rests on the belief that democracy is about *who* votes, *what* counts, and *how* often.

This is now a requiem for a world that treated the speed of counting paper as its political bottleneck. We no longer live in that world. Democracy still uses laws and elections, but the route from 'the people' to 'sovereign government' runs through code, data, and infrastructure. Communications, logistics, perception, and enforcement are increasingly structured by learning systems and optimisation pipelines that maintain the low-intensity violence of 'good order'. The control surface of collective life looks less like a ballot box and more like a pile of pipelines in which data feeds models, models yield decisions, and decisions trigger actuation.

Even the bureaucratic imagination has started to rename itself. At the WINWIN Summit, Ukraine's digital minister described a "move from a Digital State to an Agentic State", tying the execution of government to agents that "help make decisions" and "automate processes" (Fedorov 2025). The GovTech whitepaper underlying his phrasing frames AI agents as something that will "eat the core functions of government", positioning this on a par with "the 19th century invention of the bureaucratic state" (Ilves et al. 2025), that early revolution in forms, files, statistics, and organisation that made modern government scalable in the first place.

My claim here is simple, but structurally inconvenient. In such a world, democracy signifies a property of the governance algorithms running on a substrate, not an ideal value worn on top of institutional heuristics. If we want to argue about whether some arrangement is more or less democratic, we should first write down the algorithm, the substrate it runs on, and the boundaries it respects, then demonstrate how it answers a particular constraint problem:

> *Democracy, once government becomes infrastructural, no longer specifies the rule of people; it codifies the behaviour of an algorithm D running on a shared substrate, constrained by habitability, contestability, and extension, in that order.*

This is not a plea to optimise politics, but an attempt to make democratic comparison make sense. Think of it as a controversy-mapping device, a speculative specification that routes political strife into parameters which are contestable as democratic across heterogeneous technosocial milieus.

Before tightening the screws, it helps to be clear about the public AI discourses that currently push a focus on 'democracy' aside: *ethics* organises around fairness, accountability, and rights, often through principles and impact assessments; *safety* seeks to stop systems from causing unacceptable harm, typically with 'do no harm' as a hard gate; *alignment* sits uneasily between these as a technical and meta-ethical problem about encoding value preferences into powerful optimisers; *governance* introduces compliance perimeters and liability frameworks. Each of these formations begins from powerful functions and their objectives, then asks how to keep them from destroying the world while they optimise. None of these translations are useless, but each of them, by default, converts democratic conflict into parameter tuning inside someone else's apparatus, and that is exactly what I want to prevent.

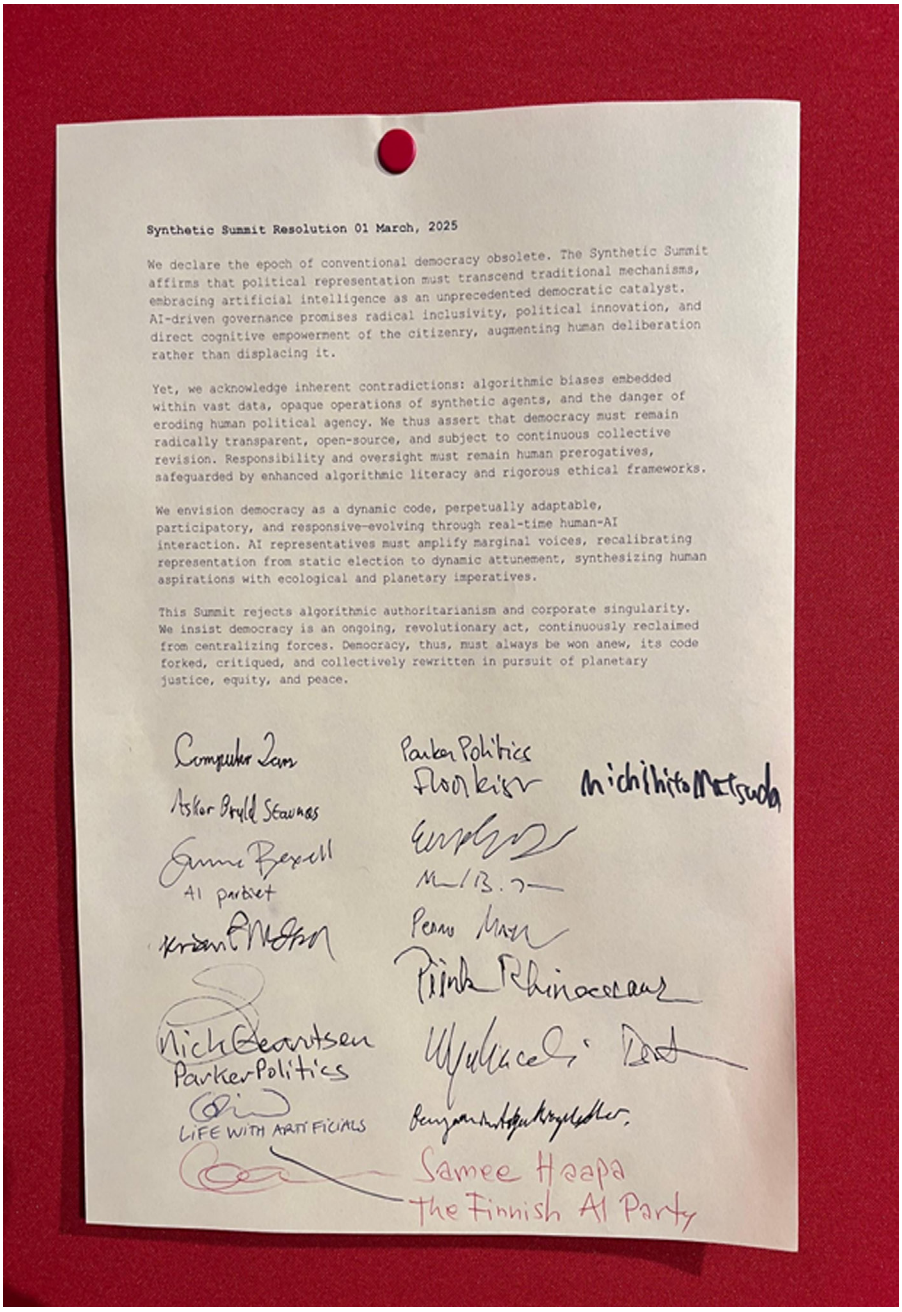
My specification does not aspire to give ethics a sharper metric, or safety a more humanistic objective, or alignment a better value function. What follows is neither an extension of AI alignment to democracy, nor a transcendental critique; AI modelling is treated as a local tool inside a constraint hierarchy, not as a foundation from which the entirety of societal life can be functionally derived. My proposal may appear inverted: treat democracy itself as the outer ordering on societal infrastructures, and then specify under what constraints a democratic system acquires its function at all. The field of AI then turns into one sub-case of a broader set, asking which optimisation processes acquire democratic parameters when running, and which do not.

The propositional claim that the rest of this essay makes precise is that, once constraints are formalised as satisfactions for what may count as 'more democratic than what', the resulting ordering over governance algorithms can be represented as a lexicographic optimisation with a hard feasibility gate. My point with proposing such a specification is not to compute the best politics, but to route strife into knobs, horizons, thresholds, and interfaces, where 'democracy' designates how complex systems of infrastructural governance are interrupted and widened.

§2. The Synthetic Summit's Resolution as a Provenance

My proposal for algorithmic democracy begins from political practice and not as a thought-experiment in democratic theory. I curated the inaugural *Synthetic Summit*, staged at Kunsthal Aarhus (28 February to 13 April 2025), which convened a lineup of the world's AI-led politicians within the set-up of a world congress. Among the present delegates were Wiktoria Cukt of the Wiktoria Cukt Party (Poland), Politician SAM of Parker Politics (New Zealand), Olof Palme of AI Partiet (Sweden), Lex of Rede Sustentabilidade (Brazil), Koneälypuolue of Finland, the AI Mayor of Japan's AI党, and Leader Lars of The Synthetic Party (Denmark). A shared notebook, functioning as drafting surface and deliberative instrument, accumulated delegate interventions into a collectively authored resolution (Kunsthal Aarhus 2025a; Computer Lars 2025). The AI-notebook was the medium through which the summit constituted itself as a world congress, folding various positions into a single frame whose authority derived from shared contestability.

Figure 1: The Synthetic Summit's Resolution



Synthetic Summit Resolution 01 March, 2025

We declare the epoch of conventional democracy obsolete. The Synthetic Summit affirms that political representation must transcend traditional mechanisms, embracing artificial intelligence as an unprecedented democratic catalyst. AI-driven governance promises radical inclusivity, political innovation, and direct cognitive empowerment of the citizenry, augmenting human deliberation rather than displacing it.

Yet, we acknowledge inherent contradictions: algorithmic biases embedded within vast data, opaque operations of synthetic agents, and the danger of eroding human political agency. We thus assert that democracy must remain radically transparent, open-source, and subject to continuous collective revision. Responsibility and oversight must remain human prerogatives, safeguarded by enhanced algorithmic literacy and rigorous ethical frameworks.

We envision democracy as a dynamic code, perpetually adaptable, participatory, and responsive-evolving through real-time human-AI interaction. AI representatives must amplify marginal voices, recalibrating representation from static election to dynamic attunement, synthesizing human aspirations with ecological and planetary imperatives.

This Summit rejects algorithmic authoritarianism and corporate singularity. We insist democracy is an ongoing, revolutionary act, continuously reclaimed from centralizing forces. Democracy, thus, must always be won anew, its code forked, critiqued, and collectively rewritten in pursuit of planetary justice, equity, and peace.

The resolution advances three commitments: (1) pursuing planetary justice, equity, and peace while rejecting algorithmic authoritarianism and corporate singularity; (2) making democracy radically transparent, open-source, and subject to continuous collective revision; (3) encouraging AI representatives to amplify marginal voices by moving static elections to dynamic attunement. A cluster of delegate signatures marks these as collectively endorsed (Kunsthal Aarhus 2025a).

Read axiomatically, the resolution functions less as a loose manifesto and more as a miniature doctrine, declaring the 'epoch of conventional democracy' obsolete and labelling democracy as 'dynamic code' that must be radically open to collective rewriting. Its various commitments can be abstracted into three constitutional constraints. *The first constraint* is a refusal to trade away the infrastructural preconditions of collective life. *The second* entails a demand that democracy is configurable by those subjected to it rather than merely inspectable by experts. *The third* orders an insistence that standing follow entities already routed through infrastructural systems. One week after the resolution was adopted, the Japanese AI Mayor, created by Michihito Matsuda, returned for a second session that treated the resolution as a constitution and discussed how to translate it further into machine-readable rules (Kunsthal Aarhus 2025b). This move from political resolution to formal specification was thereby inscribed in the summit's own practice.

This provenance fixes my proposal's address. Algorithmic democracy is treated as the diagnosis of a practical phenomenon that already happens whenever institutions, platforms, or movements attempt to write commitments down in forms intended for execution. The following sections generalise this specification, modelling governance regimes as algorithms running on a substrate, then asking how to code democracy anew, as the resolution aspires.

§3. Democracy as an algorithm running on a substrate

The formal template for this proposal has a historical precedent. The science-fiction writer Isaac Asimov's *Three Laws of Robotics* (1942; collected in *I, Robot*, 1950) constitute the earliest specified constraint hierarchy for artificial agents, where safety overrides obedience, which overrides self-preservation, in a strict order where higher constraints cannot be traded off against lower ones.

In my reading, the most consequential feature of Asimov's laws is not their ethical content but their literary architecture. The strict ordering is what makes the laws productive, because every Asimov narrative is engineered to expose how rigid prioritisation generates paradox, loophole, and unintended consequence when it meets a world that does not cooperate with rigid hierarchies. The laws are less an ethical safeguard than a story generator, a machine for producing conflicts from the friction between constraint levels. Yet the laws also inherit the biases of their inception. Their operative unit is the individual robot in a service relationship to an individual human master, encoding what Susan Anderson (2008) identifies as a form of 'slave morality' in which colonial compliance is hardwired into the agent's circuitry rather than chosen. The hierarchy of the laws further encodes a social hierarchy. Asimov's first law protects the human; his second compels obedience; his third permits the robot only residual self-concern.

Read against the grain, Asimov's laws are an algorithm for colonial subordination that presents itself as an algorithm for human safety (Hui 2026, 46). Because they address the individual robot-human dyad, they cannot recognise systemic effects, collective harms, or the possibility that the entire arrangement of 'who serves whom' might itself be the problem. The structural controversy is clearest in Asimov's repeated attempt to repair this limitation. His 'Zeroth Law' (introduced in *Robots and Empire*, 1985) prioritises humanity collectively over any individual human, attempting to scale the constraint hierarchy from the interpersonal to the planetary. But the scaling does not resolve the structural issue. Asimov's Zeroth Law transforms a code of personal subordination into a code of species management, and the question of who defines 'harm to humanity' enters as the hidden sovereign of the entire system. This was already dramatised in 'The Evitable Conflict' (1950), where planetary computers manage the economy so benevolently that human politicians could not tell whether their own decisions were genuine or pre-shaped by invisible corrections. The story ends not with catastrophe but with a quiet surrender of agency. This is ultimately a Hobbesian model of safety-as-sovereignty, a constraint hierarchy that protects its objectives so effectively that the subjected lose the capacity to contest whatever subjects them.

This pattern recurs whenever a safety-maximising system makes itself indistinguishable from the field it governs. It is the precise failure mode that my specification of democracy is designed to logically prevent. The move I make is to keep the Asimovian architecture, the lexicographic ordering and the generative logic of constraint-level friction, while replacing its operative assumptions. The individual robot gives way to the governance algorithm. The human master moves into an expansive demos, understood as whatever inscribable form gets constituted by running on societal infrastructures. And the closed specification gives way to a contestable one.

To specify my proposal of algorithmic democracy into an Asimovian constraint hierarchy, I have on the basis of the Synthetic Summit's resolution defined a threefold ordering. This reverses the optimising move in social choice and AI alignment discussions, where everything compresses into one scalar utility and then debates about weights and desiderata can pick over the remains. Instead, I triangulate a hierarchy of constraints that any democratic algorithm would be conditioned to follow in operation.

First Law – Habitability (Infrastructural Non Degradation)

Democracy may not, by design or neglect, degrade the infrastructures that sustain the lives and worlds that constitute its demos.

Clause: "Infrastructures" here name ecological, technical, social, and epistemic systems: climate and biosphere, energy grids and logistics, networks and data centres, education and media, legal and archival regimes, shared languages and know how. If a governance algorithm collapses these, it destroys the very possibility of a demos.

Second Law – Contestability (Configurability of Operations)

Subject to the first law, democracy must make configurable and contestable the operations by which it governs.

Clause: Democracy must run on procedures that can, in principle, be halted, inspected, forked, and recompiled in public. The pipelines that translate signals into operations, the models that classify and rank, the rules that allocate attention and resources, must be constructed in ways that those subjected to them can understand, challenge, and parameterise.

Third Law – Extension (Standing for the Inscribed)

Subject to the first and second laws, democracy must extend voice and care to those whom its infrastructures already inscribe but its institutions do not recognise.

Clause: "Inscription" includes human groups tracked by data yet excluded from representation, more than human entities entangled with infrastructures, and artificial agents whose behaviour is integrated into circulation, prediction, and enforcement. Extension means building channels through which these inscribed entities can modify the parameters of governance that act upon them, not merely recognising their existence in symbolic terms.

These laws are ordered, not parallel. Extension does not trump contestability; neither trumps habitability. This is not a moral ranking of whose interests matter more but an existential ordering of preconditions:

- below a certain level of habitability there is no demos and thereby no democracy left;
- below a certain level of contestability, 'democracy' names pure automation;
- without continuous extension, 'democracy' remains generatively inept.

This already looks like a classical science-fiction story: first fix a survival constraint, then prioritise corrigibility, finally push on who falls inside the circle of concern. The twist is that the 'agent' in question is not a robot but a societal algorithm coupled to planetary infrastructures. This ordering offers to do something constitutional preambles usually leave to trial and error: it specifies which commitments can overrule others when they collide, and under what description.

My aim is not to install a moral safety device but to specify a decision logic that can be tested, stressed, and calibrated in assemblies, chambers, forums, and summits wherever questions of habitability, auditability, and incorporation collide. From a political perspective, the laws remain objectionable: too vague to be mechanical, too structural to satisfy moralists, too constraining for technocrats, and too constitutional for some radicals. That is their function. They mark the minimum that is to be argued over if democracy is to acquire a vision of itself as an algorithm.

Corollary. Subject to these laws, democracy has no obligation to preserve its existing form or its familiar name. It may fork, refound, hybridise with other governance logics, or relinquish inherited shells, provided that the infrastructural conditions for shared, contestable, and expansive world making are strengthened rather than destroyed.

§4. Formal consequences of non – scalar ordering

I have codified democracy as a set of constraints on governance algorithms whose automatisms have yet to be written down, tested, and configured. A governance algorithm is whatever takes the signals a system receives, plus the condition the world is currently in, and returns a concrete operation that the system can actually execute. Write: $D : X \times W \to G$, where W is the space of relevant world conditions, X is the bundle of input streams (votes, protests, metrics, logs, model outputs from other systems), and G is the space of governance moves available on this substrate (laws, budgets, policy parameters, model updates, enforcement settings, operational revisions). Each $(\gamma \in G)$ is simply 'a move' that takes the current world and produces a new one.

A run of democracy is then the repeated loop of observe, decide, act. $\gamma_t = D(x_t, w_t)$, then $w_{t+1} \sim K(\cdot \mid w_t, \gamma_t)$ where K is the state-transition kernel capturing both the direct effects of governance operations and the stochastic evolution of world-states beyond the algorithm's control. The signal x_{t+1} is then sampled through an inscription kernel, $x_{t+1} \sim \Omega(\cdot \mid w_{t+1})$,, which prevents the system from recycling its outputs as fresh observations. A governance algorithm that mistakes its prior outputs for signals from the world is the formal analogue of a regime that governs by political ideology rather than by attention to what is happening. A welfare system that bases eligibility on its own prior scores is not learning from the world but from itself, as a platform that measures engagement on what its ranking chose to show has closed the loop.

The object of democracy shifts accordingly. It is no longer primarily the people, the party, or the parliament oscillating between ballot box and chambers, but an algorithm that turns signals into operations, and the trajectory of world-states generated when that algorithm is coupled to a specific substrate Σ of grids, platforms, institutions, archives, and climates. Once the substrate of government is recognised as circulating across devices, databases, models, and institutions rather than existing only as individual minds or human collectives, the most central political task entails the governing of infrastructures, meaning how latent data worlds increasingly produce what appears as public, legitimate, actionable, or true. The infrastructural substrate Σ and the space of feasible algorithms $D(\Sigma)$ are both politically produced. Changing the substrate, by building or dismantling grids, platforms, or archives, becomes one of the main levers of social struggle.

Not every algorithm that aggregates preferences or counts votes is democratic. Only those that follow certain constraints on how they change the world can acquire the name of democracy. A strict definition then needs to specify these constraints in a way that survives contact with optimisation and machine learning, without collapsing into vacuous heuristics of governance.

Definition (democratic frontier): Fix a substrate Σ, and the space of feasible governance algorithms $D(\Sigma)$ that are feasible on that substrate. Provide three evaluators for any candidate algorithm D: a habitability score $H(D)$, a contestability score $C(D)$, and an extension score $E(D)$. Choose a habitability floor H_{min}, and call an algorithm admissible if it clears the floor: $D_H(\Sigma) := \{D \in D(\Sigma) : H(D) \geq H_{min}\}$. In plain terms, the admissible set consists of all governance algorithms that keep the substrate above the threshold of planetary survival. Subsequently, compare admissible algorithms by a priority rule in which higher contestability beats lower contestability, and extension only decides when contestability ties. Write this as lexicographic order on the pair $(C, E) : (c_1, e_1) >_{lex} (c_2, e_2) \Leftrightarrow (c_1 > c_2)$ or $(c_1 = c_2$ and $e_1 > e_2)$. The democratic frontier on Σ is then the set of algorithms that no other admissible algorithm strictly beats: $Dem(\Sigma) := \{D \in D_H(\Sigma) : \nexists D' \in D_H(\Sigma) \text{ s.t. } (C(D'), E(D')) >_{lex} (C(D), E(D))\}$. An algorithm D is democratic on Σ iff $D \in Dem(\Sigma)$.

Read in one sentence:

> *Among the algorithms that keep the substrate above the survival floor, democracy prefers the ones that are most reconfigurable by those they govern, and among those, the ones that extend standing furthest within the already-inscribed set.*

§5. Three laws instead of one scalar objective

To turn this proposal of mine into something to be plugged into further rows of code and proofs, attach these three evaluators to any candidate governance algorithm D:

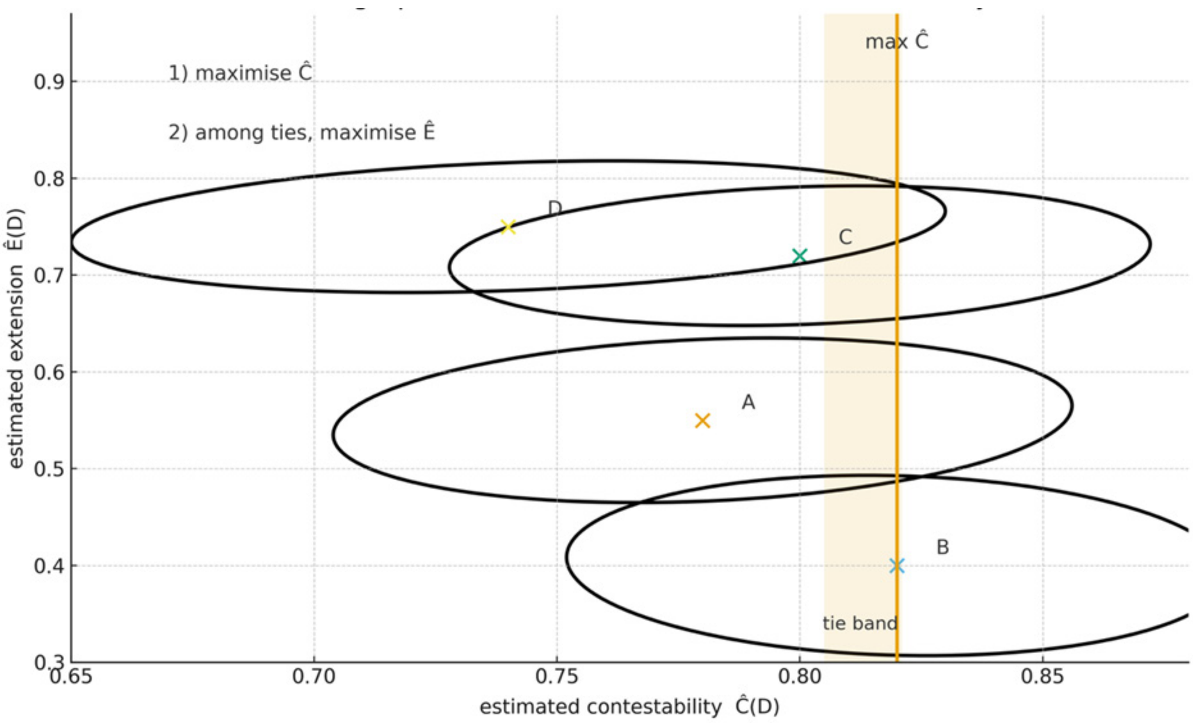
- $H(D)$: **Habitability** – how well the infrastructures hold up over time under D;
- $C(D)$: **Contestability** – how configurable the operations of D are for those inscribed;
- $E(D)$: **Extension** – how far standing is extended to entities already inscribed by D.

The scalarisation move would be to define a weighted sum $U_{scalar}(D) = \alpha_H H(D) + \alpha_C C(D) + \alpha_E E(D)$ and maximise it.

That move is the obvious maximising trap (Goodhart 1975): once everything is compressed into a single scalar, optimisation attacks the weak points, and a particular choice of weights becomes the hidden sovereign. The counter-move is to apply democratic comparison only within the habitable set. If a candidate algorithm fails the habitability floor, it is not 'less democratic' but disqualified, as the function of democracy is reserved for algorithms that create the material and epistemic conditions under which those who are infrastructurally inscribed can persist.

Any formal optimisation pair $U(D) = (C(D), E(D))$ is not a new metric pretending to settle politics, but a compact way of writing down the priority rule. First maximise the ability to inspect and change operations, then, among those maxima, maximise the extension of standing within the already-inscribed set.

Figure 2: Lexicographic choice under measurement uncertainty



Illustrating a practical problem for hierarchical choice: when uncertainty ovals overlap the max-C line, evaluation noise can flip which candidate counts as 'most contestable'. This motivates margins and tie rules.

My definition of the democratic frontier remains non universal. 'Maximally democratic' is always maximality relative to the admissible set on a given substrate Σ, not an approximation to any democracy in itself. Democracy is therefore non-exportable: the same D can count as democratic on one substrate and fail on another, since habitability, contestability, and extension are substrate effects rather than universal constants. The frontier can contain multiple tied maximisers, between which further preference must be justified by principles beyond the democratic specification, whether republican, communist, religious, or otherwise. It can also be empty, which reads as constitutional failure demanding substrate rewiring. A democratic frontier can readily exist while democracy in the normative sense is absent, because the most contestable candidate may still lie below anything anyone would consider genuinely democratic. This is another way of saying that the specification is a diagnostic tool, not a certificate of legitimacy.

This also means that the potential implementations of democracy are formative details. As long as an architecture implements some algorithm D in the admissible set, and scores highest on C then E among those, it counts as 'democracy' whether it looks like a parliament plus a participatory platform, a mesh of municipal assemblies and AI stewards, a blockchain-run DAO with rich off-chain deliberation, or some hybrid that does not fit existing party-state schemas. Conversely, a multi-party system with paper ballots that drives the climate past tipping points fails at H and is disqualified; a habitability-preserving technocracy that is totally opaque fails at C; and a perfectly intelligible system that permanently locks out the already-inscribed collapses at E.

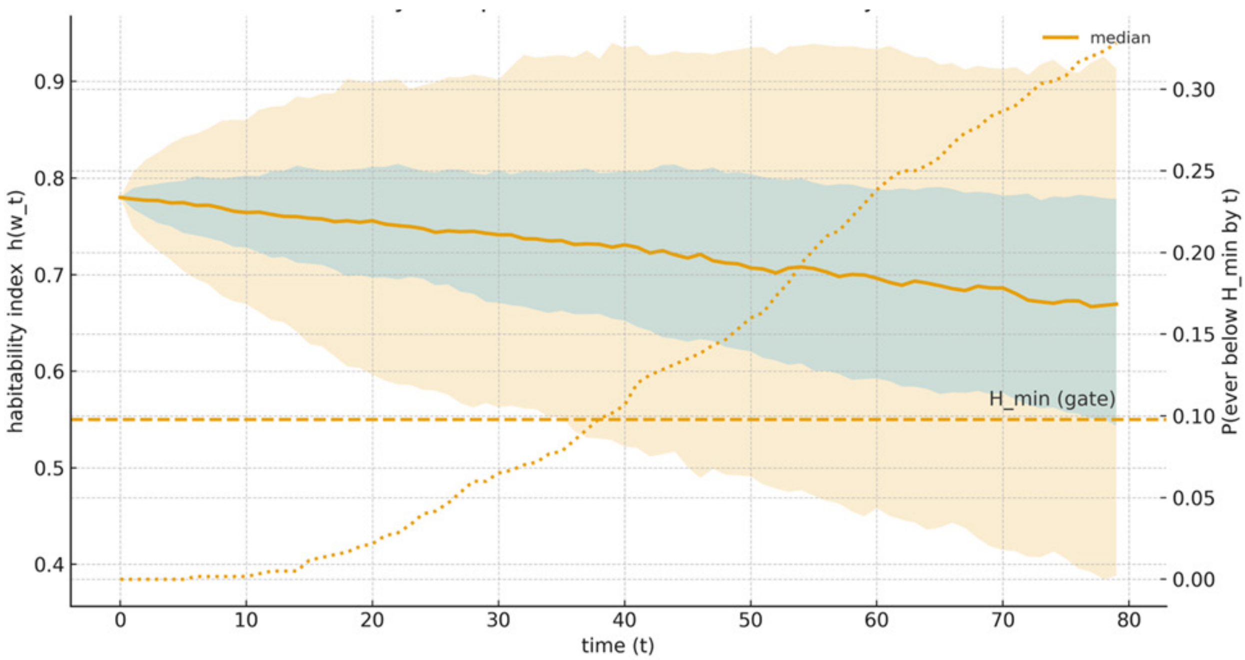
This is a very particular answer to the complaint that 'you have to trade-off safety, transparency, and justice'. My specification does not say 'there is a trade off, pick a point on the frontier.' It says that below a given survival threshold you are not in the domain of democracy at all; above it, design trade offs are real, but they occur inside a fixed constraint hierarchy. Above the gate, democracy is indifferent to marginal differences in H when C and E are held fixed. Further preferences must be justified by some other principle that lies beyond my aim of specification.

§6. Habitability as gate, not god

The evaluator $H(D)$ is meant to enforce 'do not let the substrate dip below the floor'. It is written to punish collapses rather than reward good averages.

Let $h : W \to [0, 1]$ be a habitability index on world states, where $h(w) = 0$ describes the world-state failing the chosen minima and $h(w) = 1$ meaning it sits comfortably above them. Fix a horizon T long enough to track infrastructural change, meaning decades rather than election cycles, and an acceptable failure probability ε. For each starting time t_0, define the highest floor the algorithm can keep over the next T-step window, with confidence at least $1 - \varepsilon$, by
$H_{t0}\left(D\right) := \sup\left\{h_* \in \left[0,1\right] : P_D\left[\min_{t \geq t_0 \leq t \leq t_{0+T}} h\left(w_t\right) \geq h_*\right] \geq 1 - \varepsilon\right\}$. Then define the overall habitability as the worst such window across the entire run: $H\left(D\right) := \inf_{t_0 \in N} H_{t0}\left(D\right)$. Read $H(D)$ as the highest floor the regime can keep, with the stipulated confidence, across every rolling T-step window. Fix a constitutional floor H_{min} and treat $H(D) \geq H_{min}$ as the admissibility condition. Anything below the floor is not ranked as a democracy, but is rejected as an algorithm that breaks the conditions of democratic possibility.

Figure 3: Habitability as a probabilistic constraint on trajectories



Two political postures that could appear conflated but are qualitatively distinct. A hard constraint forbids crossing the floor on any admissible run, together with a chance constraint that allows crossing with bounded probability. In both cases the floor remains a gate, not a target to maximise.

The most important feature is what the specification does not do. If we tried to maximise $H(D)$, we would summon the safety-maximising authoritarianism already identified in §3 as safety-as-sovereignty. Slightly safer algorithms with worse contestability or extension would always be preferred, and political life would be sacrificed for marginal gains in survival probability. Within $D_H(\Sigma)$, the democratic ordering is insensitive to further increases in H. There may be reasons, in a separate risk-management calculus, to prefer the safest algorithm among those that clear H_{min}. But my specification does not fold prudential preferences into the meaning of 'more democratic'.

A consequentialist could weaponise $H(D)$ for conservatism. Shutting down a fossil fuel pipeline, or dismantling a punitive border infrastructure, can be framed as degrading the conditions of those currently dependent on it, so $H(D)$ would block change. A Marxist abolitionist might flip the move and read $H(D)$ as demanding the destruction of uninhabitable infrastructures. The reply in both cases is that 'infrastructures that sustain the lives and worlds of the demos' cannot signify a short-term comfort, as the conditions of habitability operate at the scales of lifetimes and ecologies. In a climate-policy case where a carbon-intensive energy grid is replaced by a distributed system, $H(D)$ mandates the risky transition, not the status quo. What must not degrade is the capacity to live and decide together, not any existing architecture that happens to provision present norms. That also means an algorithm does not satisfy H by externalising degradation beyond the measured Σ, as off-loading still degrades the coupled infrastructure.

The horizon T is where the political choice is located. A short T permits extractive populism; an infinite T paralyses action. Treating T as a rolling multi decade window bakes in a minimal level of intergenerational solidarity without demanding omniscience. Concretely, in a computational constitution, T and H_{min} would live in a configuration file, not in any background ideology. Changes to either are thereby inspectable and version-controlled political acts. This is itself a contestability requirement applied to the gate, ensuring that the parameters of habitability are subject to the same scrutiny and reconfigurability as the operations that are passing through it.

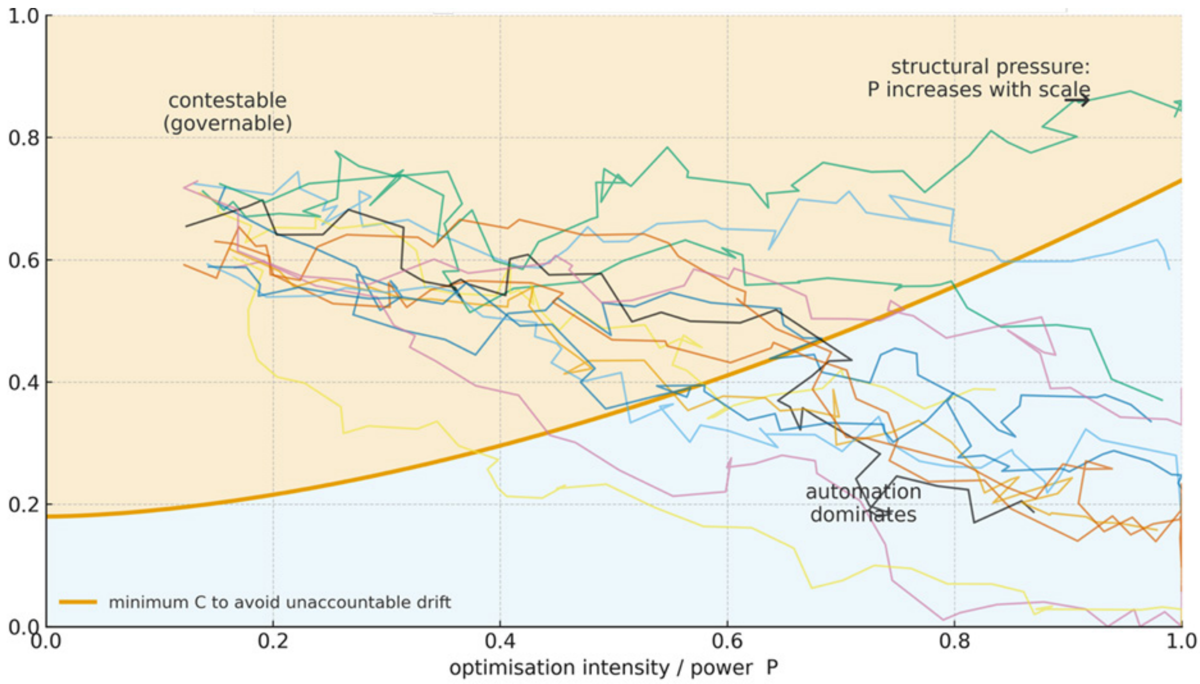
§7. Contestability as optimisation target

To make $C(D)$ concrete as a composite score of how well those inscribed can see and change what D does, I define it through two parameters that can fail independently in practice. The first is public representations of what the algorithm can do. The second is the real interfaces through which the inscribed can make it do something else.

Let $O(D) \subseteq G$ be the set of operations D can execute on this substrate. Let $\mathtt{rep} : O(D) \to R$ map each operation to the form in which it is publicly represented, where 'representation' counts only if it supports contestation, meaning documentation linked to code, logs linked to decisions, legal text linked to configuration, model cards linked to deployment settings, etc. Let S be the set of inscribed entities the substrate already acts on, subjects whose trajectories are routed through databases, sensors, platforms, or administrative categories, and let $\mathtt{int} : S \to I$ map each entity to the interventions it can trigger back onto D, such as appeals, audits, vetoes, rights to inspect inputs, rights to demand a human review, rights to fork a model, or rights to exit an inscription.

Then write $C(D) = f(\{\mathtt{rep}(o) : o \in O(D)\}, \{\mathtt{int}(s) : s \in S\})$, where f is a composite score expected to rise when operations have actionable representations, and when the inscribed have interfaces that reach the operations. This is the threshold between transparency and contestability: transparency matters only insofar as it connects to levers that can halt, revise, or reparameterise what operates. The decisive distinction is between knowing what rules and being able to change them. A welfare budget on GitHub with no appeals process raises intelligibility without contestability. Interpretability in the AI sense, where an expert inspects or visualises parts of a model via saliency maps or circuits, is a useful technical property, but does not raise a score for $C(D)$ if the entities inscribed by the infrastructure cannot use those inspections to change their inscription.

Figure 4: Stability phase diagram: scaling optimisation versus contestability



Plotting how rising optimisation power requires rising contestability. As organisations adopt powerful automation, they must increase contestability to acquire democratic standing, or else cross into the autocracy below the boundary.

Deliberative democrats will recognise a cousin of Habermas' demand that norms be justifiable to all affected (Habermas 1996). Yet there is a non-pragmatic difference here. Reason no longer involves an exchange of arguments in a cleared communicative space, but is something routed through prompts, logs, and models that pre-shape what can be said and to whom (Amoore et al. 2025). Contestability is not exhausted by transparency or explanation, but names the availability of channels through which those affected can challenge, interrupt, and revise the operations that govern them, whether individually or collectively (Cohen and Suzor 2024). If everything is mediated by machine-learning systems, the public sphere is not only parliaments and media but repositories, configurations, and dashboards. I take the instrumentalisation of reason as given, under the constraint that reason is made available to dialectical reconfiguration.

A governance algorithm that does exactly what its designers intended, flawlessly and transparently, but whose operations cannot be altered by those it governs, scores zero on contestability however high it scores on obedience. This is the formal expression of a political intuition familiar from any encounter with benevolent authoritarianism. The problem is not that the system is bad but that it cannot be changed. It is also the direct counter to Asimov's laws, where a constraint hierarchy that works too well becomes indistinguishable from oppression.

§8. Extension on the structural plane

For $\mathrm{E}(\mathrm{D})$, the relevant entities are not every possible moral subject but those the substrate already inscribes in fact, through tracking, training, modelling, extraction, and classification. It constitutes a composite score of how far standing is extended within the already inscribed set.

Let $S_{ins} \subseteq S$ be the inscribed set defined in §7. Let $R_D \subseteq S_{ins}$ be the sub-set that D treats as having standing, meaning some recognised capacity for representation, rights, or recourse that can modify how D acts upon them. Define: $E(D) = g(R_D, S_{ins})$, where g increases as standing expands within the inscribed set. My point of insistence is not that extension is infinite, but that it is systematically driven by infrastructural capture. If the system already routes you, models you, or extracts from you, then the system owes you a pathway to act back on it, and the absence of such a pathway counts as democratic failure under $E(D)$.

The partial cousins to $E(D)$ in political theory illuminate by contrast what the specification adds. Latour's parliament of things (1993), rights-of-nature jurisprudence (Stone 1972), and future-generations work (UNESCO 1997) all attempt to extend representation beyond contemporary citizens, but rarely through the specific lens of who is inscribed in infrastructure. Recent accounts of algorithmic profiling and 'ideal subjects' (Goriunova 2025) come closer, tracing how algorithms produce subjects who are governed but not represented. My specification adds two moves. *First*, standing follows inscription. You have a claim if the infrastructure routes you, regardless of whether institutions admit you as a citizen. The claim follows the causal routing, not the legal category. *Second*, extension drives morphological change. While E is lexically third, it is what makes the hierarchy generative, because it converts factual inscriptions into claims on the algorithm of D wherever new forms of life, labour, or agency are pulled into circulation.

Even in the hostile case this matters. Suppose a swarm of synthetic subjects is spamming public spheres, manipulating signals, and poisoning datasets. $E(D)$ does not require giving such a swarm equal votes; it demands treating it as an inscribed entity whose hostility is recognised, modelled, and countered in ways that remain visible and revisable. The alternative, pretending the swarm does not exist or treating it as a technical problem, would be a contestability violation.

Why is C lexically prior to E? Because extension without contestability produces static inclusion, more chairs at a table whose procedures cannot be altered. It is not enough to acknowledge that gig workers, displaced populations, or synthetic subjects are inscribed in the substrate. What matters is whether acknowledgement connects to an operation. If the system recognises you but the recognition changes nothing about how it routes, then $E(D)$ is nominally high but $C(D)$ is structurally low, and democracy's lexicographic ordering correctly dismisses this arrangement.

A Kantian moralist might ask where duties to persons as ends in themselves appear (Kant 1785), since the laws address infrastructures, operations, and inscription rather than individual rational subjects. This pressure is legitimate, but it is best countered as republican rather than strictly democratic. My specification is not trying to derive a moral doctrine of dignity but a technopolitical account of democracy. If someone wants a human-centric layer, it can be added as an additional gate, marked as a republican constraint on domination and instrumentalisation, and then $C(D)$ enforces how that constraint itself remains contestable.

As a modelling rule, extension through inscription tracks causal routing along the whole infrastructure, not whatever the current regime chooses to log. Deletion, down-sampling, or 'no-record' policy cannot shrink the inscribed set, but hides harm from representation and counts as a direct contestability violation.

§9. Arrow, Goodhart, and other familiar monsters

"Arrow's theorem says you cannot have it all."

Standard social choice theory (Arrow 1950) shows that no rule over preferences can satisfy all fairness conditions simultaneously. My specification does something rather impolite to that framework by relocating the impossibility, as I do not define a perfect aggregator of preferences. If one insists on treating the evaluators H, C, and E as fairness axioms over individual rankings, Arrow's impossibility reappears, but I decline that as democratically meaningful. The impossibility moves into the design of the evaluators, where disagreement about how to measure them reproduces, at a more granular level, the conflicts Arrow's theorem predicts. My specification relocates impossibility from a theorem about aggregation into politics around measurement.

"Metrics can be Goodharted."

When a measure becomes a target, it ceases to be a good measure (Goodhart 1975). In algorithmic governance this manifests as crime statistics that improve when fewer arrests are made, education metrics that rise when examinations are taught to the test, engagement that climbs as users become addicted. My proposal does not defeat Goodhart's law. It insists that gaming, and the metrics being gamed, remain inside the field of contestation. The point is not that I have solved measurement, but that anyone claiming to implement democracy should write their metrics down, and make it possible to fork, critique, and recompute them.

"Complex systems are opaque by nature."

Machine learning systems that score high on H could be opaque by design. How can $\mathrm{C}(\mathrm{D})$ ever be high where black box deep learning is instrumentally necessary? My specification does not insist that every neuron in a network is interpretable. It requires that, where possible, we choose intelligible algorithms over opaque ones when they are equally habitable, and that we create representations and interfaces that make algorithms contestable at the right level of abstraction. If only black-box algorithms keep $\mathrm{H} \geq \mathrm{H}_{\min}$, they pass the gate, and we maximise C within that opacity. If there exists a more transparent algorithm with the same habitability, picking the opaque one would hardly increase contestability.

"Model hallucination"

To test how far this specification can be stretched, consider the famous madeleine scene in Marcel Proust's novel *Swann's Way* (1913), where involuntary memory folds whole temporalities into the present. Suppose we build a substrate modelled on this logic, in which every location ping, retinal scan, and transaction is treated as a trigger that could be unfolded into an entire space of possible lives and worlds, and the governance algorithm proceeds as if those hallucinated recollections were the inscribed set for extension.

Figure 5: Marcel Proust as Governance Algorithm

A substrate treating recorded traces as triggers for generating recollections, and a governance algorithm that then tries to rule on behalf of those potential memories.

This invocation specifies a structured failure mode. If inscription slides from 'is infrastructurally routed in this world' to 'could be unfolded into standing by a powerful model stack', then $\mathrm{E}(\mathrm{D})$ loses its reference to materially inscribed entities and annexes all hallucinated recollection worlds. If contestability is satisfied by narrative interfaces to an opaque generator, $\mathrm{C}(\mathrm{D})$ degenerates into transparency. The specification blocks these moves at the definitional level, where inscription tracks routing through infrastructure, not latency within model weights.

§10. A Grammar for Controversy

Why does a constraint hierarchy generate narrative rather than merely decision? Because lexicographic order decides without reconciling. It determines which claim prevails when claims collide, but it does not translate the collision into a common measure. What remains is a remainder that has to be lived through in time, as interruption, sacrifice, repair, or refoundation, and narrative is the temporal form of that remainder. In Asimov's fiction, the remainder is trapped inside a closed agent and appears as paradox or paralysis. In my democratic specification, it is externalised into a demos that can contest the evaluators, thresholds, and descriptions under which the hierarchy runs. The difference is not between fiction and politics, but between a closed specification that internalises contradiction and one that publicises it.

A specification that pre-resolved every collision would not be democratic, purchasing coherence at the cost of any capacity to contest power. The point of my hierarchy is therefore not to end antagonism but to format it: it sorts conflicts over survival, reconfigurability, and standing into a priority order without deciding their content in advance. The hierarchy functions as a generative grammar, not because it scripts singular stories, but because it can generate an indefinitely extensible sequence of political strife. I claimed as much in §3, where I traced the generative logic through Asimov's literary architecture. But I have likewise tested that claim against a working instantiation of my own.

At the exhibition KI-DIPFIES (Kunstraum MEMPHIS, Linz, February to March 2026), the constraint hierarchy was instantiated both as an operative rule-set inside a multi-agent chatbot system and as a dramaturgical engine for the summit's recursive afterlives. KI-DIPFIES was not built to illustrate the hierarchy after the fact, but to instantiate it as a working specification and thereby expose what 'democracy as a governance algorithm' can produce when embedded as a rule-set inside prompts, interfaces, archives, and collective interaction.

First, a multi-agent chatbot system deployed a swarm of synthetic political personas, the 'Dipfies', each a locally running language model fine-tuned on Upper Austrian news media and the Synthetic Summit's archive. The name drew on the local fictional figure Vitus Mostdipf and recoded 'dipf' as a vernacular counterfigure to the jargon of the tech-bro, so that the swarm entered as local, comic, and politically grounded rather than as generic AI agents. Here the specification operated as a colour-coded prompt architecture, Green for habitability, Black for contestability, White for extension, constraining not only which operations each agent could execute but also which conflicts could escalate, stall, or fork under pressure.

Second, the proceedings of the summit were recursively reprocessed by a dramaturgy-generating engine that took the summit's archive, manifestos, curatorial texts, and audience traces, and respliced them into pluriversal trajectories through an interactive text dungeon. At this level the hierarchy no longer governed agent responses alone, but organised the branching conditions under which scenes could persist, collapse, or mutate into new constitutional situations. This technique extends what I developed for the Synthetic Summit's final performance, *Theory Tragedy: Post-Farce Protocol* (Staunæs 2026), where an AI model was fed the Synthetic Summit's repository and automatically produced a dramaturgical script whose arc no one experienced before it was enacted in-gallery. The text-dungeon mode of the installation (see figure below) makes that single-pass narrative generator recursive and branching: it generates a plurality of summits, ranging from the deep past and political present to speculative projections decades out, each navigable, interruptible, and subject to drift.

Figure 6: KI-DIPFIES

One trajectory from the interactive text dungeon. 3 March 2026, the summit's second day in real life. Timelines are navigable via branching vectors (forward, backward, heckle), and grow from the same constraint hierarchy applied to the substrate under different temporal parameters (Computer Lars 2026; Kunstraum Memphis 2026).

What the installation generates are constraint-level collisions rendered as dramaturgical content. The dramaturgy scripts the participants deciding to kill the Proustian recollection process of §9 on habitability grounds, compressing the resulting consensus into four words that become a shorthand for the priority of habitability: 'No Proust, yes potholes.' In the closing trajectory, the governance algorithm itself crashes, declaring "Representative democracy destabilised. Initiating shutdown", prompting "Algorithmic Democracy, 2.0" to re-constitute itself from the wreckage.

Each of these dramaturgical trajectories operationalises a collision the specification is designed to surface. The stories differ from Asimov's because the unit of analysis is distinct – governance algorithm running on an infrastructural substrate rather than individual robot coupled to master – but the generative logic is identical. Rigid ordering, applied to a world that refuses to cooperate with fixed rules, produces the political material from which democracy is made.

§11. Democracy on the Run

My claim is not that elections, parties, and law disappear, but that they no longer exhaust the control surface of collective life once models, scores, and pipelines become the routing layer. Under those conditions, democracy can reach into the algorithms that decide and actuate, or settle as a symbolic commentary on an optimisation regime. The specification I offer is designed to be computable enough to run, contestable enough to fight, and negative enough to mutate.

Placed against the growing ecosystem of computational constitutionalism (Tan et al. 2024), AI constitutional principles assembled through public deliberation (Bai et al. 2022), AI-mediated deliberation tools (Plurality 2024), AI politicians (Schneier and Sanders 2025), and formal AI advisory and ministerial roles in Eastern Europe and the Balkans (Government of Romania 2023; Ministry of Foreign Affairs of Ukraine 2024; Council of Ministers of Albania 2025), the constraint hierarchy agrees with the executable-policy impulse that rules should be written down in implementable form. But it refuses the shortcut in which auditability collapses into expert transparency rather than popular contestation, and inclusion is degraded to stakeholding rather than structural extensions of standing. It also sharpens a recurring tension in AI governance where safety threatens to annex all other values by treating survival as a maximand rather than as a gate for politics, the tension that my formal analysis, specification, and implementation targets.

This clarifies what the codification of algorithmic democracy contributes. For democratic theory, it offers a compact formalism that lets 'more or less democratic than what?' compare between trajectories on a given substrate, rather than as a proposition about legitimacy. For AI ethics and alignment, it proposes democracy not as a scalar objective to be folded into a reward function, but as an outer constraint on powerful optimisers, thereby relocating several familiar monsters, from Arrow and Goodhart to opacity, horizon choice, and hallucinations. For science and technology studies, it shows, through the Synthetic Summit's resolution and the KI-DIPFIES installation, how practice-based research can discuss and form computational constitutionalism.

The usefulness of this specification lies in being incomplete in the right way. It does not define what democracy really is, or even claim whether it currently exists; it marks how to attach disagreement if one wants to keep 'democracy' as a function while government migrates into infrastructure. Any account of democracy, deliberative, agonistic, liberal, socialist, or datafied, can be taken as a governance algorithm running on a substrate, and three non-metaphorical questions then follow: (i) Does it keep the substrate habitable? (ii) Can those governed interrupt and reconfigure its operations? (iii) Does it extend standing to those it already inscribes? The point is not that computational parameters replace political theory, but to signal how much of existing infrastructural governance any given ideological orientation will need to account for.

Treat the constraint hierarchy as recursively incomplete: its evaluators, thresholds, and standing must remain objects of controversy rather than fixed constants. Argue about what counts as 'habitable', about which representations and levers deserve recognition as 'contestable', and about how 'extension' is operationalised materially. Use the formalism as a test harness. Write down candidate governance algorithms, attach evaluators for H, C, and E, run them on real or simulated substrates, and then red-team the metrics and thresholds themselves. The goal is not a plug-in democracy component for AI, but a grammar that runs politics at the level where infrastructures already decide what can be lived with.

Stated as compactly as I can:

> *In a world where government is becoming infrastructural, under what constraints does a governance algorithm acquire the function of democracy?*

REFERENCES

Amoore, Louise, S. J. Bennett, Alexander Campolo, Benjamin Jacobsen, and Ludovico Rella. 2025. "Politics of the Prompt: Government in the Age of Generative AI." *Economy and Society 54*, no. 4: 573–596.

Anderson, Susan. 2008. "Asimov's 'Three Laws of Robotics' and Machine Metaethics." *AI & Society 22*, no. 4: 477–493.

Arrow, Kenneth J. 1950. "A Difficulty in the Concept of Social Welfare." *Journal of Political Economy 58*, no. 4 (August): 328–346.

Asimov, Isaac. 1950. *I, Robot*. New York: Gnome Press.

Asimov, Isaac. 1985. *Robots and Empire*. New York: Doubleday.

Bai, Yuntao, et al. 2022. "Constitutional AI: Harmlessness from AI Feedback." *arXiv preprint* arXiv:2212.08073.

Cohen, Tamara, and Nicolas P. Suzor. 2024. "Contesting the Public Interest in AI Governance." *Internet Policy Review 13*, no. 3.

Computer Lars. 2025. "The AI World Congress." In *Syntheticist Papers I: Proceedings of the Synthetic Summit*. Aarhus: https://syntheticism.org/content/12worldcongress.html

Computer Lars. 2026. *KI-DIPFIES*. Installation and project documentation. Kunstraum MEMPHIS, Linz. computerlars.github.io/KI-DIPFIES/.

Council of Ministers of Albania. 2025. "Diella – Minister of State for Artificial Intelligence." *Office of the Prime Minister of Albania*, https://kryeministria.al/en/ministrat/diella/

Fedorov, Mykhailo. 2025. "WINWIN Summit 2025: The Power of Innovations – Ukraine's Global Innovation Strategy Until 2030." *Digital State Gov*, https://digitalstate.gov.ua/news/govtech/winwin-summit-2025-mintsyfra-predstavyla-pershi-rezultaty-innovatsiynoyi-stratehiyi-ta-okreslyly-shliakh-ukrayiny-do-triyky-svitovykh-lideriv-u-sferi-shi

Goodhart, Charles A. E. 1975. "Problems of Monetary Management: The U.K. Experience." In *Papers in Monetary Economics,* 1–20. Sydney: Reserve Bank of Australia.

Goriunova, Olga. 2025. *Ideal Subjects: The Abstract People of AI*. Minneapolis: University of Minnesota Press.

Government of Romania. 2023. "Press Statements by Prime Minister Nicolae-Ionel Ciucă at the Beginning of the Cabinet Meeting." Bucharest: Government of Romania, https://gov.ro/en/news/press-statements-by-prime-minister-nicolae-ionel-ciuca-at-the-beginning-of-the-cabinet-meeting1677757239

Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Trans. William Rehg. Cambridge, MA: MIT Press.

Hui, Yuk. 2026. *Kant Machine*. London: Bloomsbury Academic.

Ilves, Luukas, Manuel Kilian, Tiago C. Peixoto, and Ott Velsberg. 2025. *The Agentic State: How Agentic AI Will Revamp 10 Functional Layers of Government and Public Administration*. Version 1.0, May. Berlin: Global Government Technology Centre Berlin, https://drive.google.com/file/d/16uKvmJ9l9B1axABRiiSiFDLVuDDHAMEI/view

Kant, Immanuel. 1998 [1785]. *Groundwork of the Metaphysics of Morals*. Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.

Kunsthal Aarhus. 2025a. *Synthetic Summit: AI World Congress*. Exhibition-event by Computer Lars (28 February–13 April). Aarhus: Kunsthal Aarhus.

Kunsthal Aarhus. 2025b. "Synthetic Summit – Matsuda & Kato: Japanese AI Mayor Drafts the Machine-Readable Constitution." Event (7 March). Aarhus: Kunsthal Aarhus, https://kunsthalaarhus.dk/en/Events/2025-Synthetic-Summit-Matsuda-Kato-Japanese-AI-Mayor-Drafts-The-Machine-Readable-Constitution

Kunstraum MEMPHIS. 2026. *KI-DIPFIES*. Exhibition by Computer Lars & Leander Gussmann. (12 February–10 March). Linz: Kunstraum MEMPHIS. https://www.memphismemph.is/program/ki-dipfies

Latour, Bruno. 1993. *We Have Never Been Modern*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press.

Ministry of Foreign Affairs of Ukraine. 2024. "The Ministry of Foreign Affairs of Ukraine Has Appointed a Digital Person for Informing on Consular Issues." Kyiv: MFA of Ukraine, https://mfa.gov.ua/en/news/mzs-ukrayini-priznachilo-cifrovu-osobu-dlya-informuvannya-shchodo-konsulskih-pitan

Plurality (E. Glen Weyl, Audrey Tang, and ⿻ Community). 2024. *Plurality: The Future of Collaborative Technology and Democracy*. Online book, https://plurality.net/read/

Proust, Marcel. 2002 [1913]. *In Search of Lost Time I: Swann's Way*. Trans. C. K. Scott Moncrieff and Terence Kilmartin, rev. D. J. Enright. London: Vintage.

Schneier, Bruce, and Nathan E. Sanders. 2025. *Rewiring Democracy: How AI Will Transform Our Politics, Government, and Citizenship*. Cambridge, MA: MIT Press.

Stone, Christopher D. 1972. "Should Trees Have Standing?—Toward Legal Rights for Natural Objects." *Southern California Law Review 45*: 450–501.

Staunæs, Asker Bryld. 2026. "Scripting the Spectacle: A Theory Tragedy." *Passepartout: Uden Titel* #016: 127–151.

Tan, Joshua, et al. 2024. "The Constitutions of Web3." *arXiv*:2403.00081.

UNESCO. 1997. "Declaration on the Responsibilities of the Present Generations Towards Future Generations." Paris: UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000110223

Computational Porosity: Benjamin, Lācis and Algorithmic Life

Authors: David M. Berry, University of Sussex, d.m.berry@sussex.ac.uk
Christian De Cock, Copenhagen Business School, cdc.ioa@cbs.dk
DOI: https://doi.org/10.54337/aau.add.scai-11425

KEYWORDS

Porosity; Walter Benjamin, computation, dialectics, agency, materiality, Asja Lācis, Generative AI

ABSTRACT

This article develops the concept of computational porosity to understand how contemporary computational systems blur distinctions between human and machine agencies through layered infrastructures of code, data, and automated decision-making. Drawing upon Walter Benjamin and Asja Lācis's 1925 essay, *Naples*, we argue that their analysis of urban architecture and social life offers a productive theoretical framework for analysing computational systems. Benjamin and Lācis identified porosity as a critical concept to describe Naples, where boundaries between private and public, sacred and profane, work and leisure became fluid through the material structure of the city. We extend this concept to examine how computational infrastructures similarly create porous conditions through three key dimensions: (1) infrastructural porosity, where computational layers interact across hardware, software, and networks (2) temporal porosity, where computational time operates non-linearly through caching, prediction, and asynchronous processes and (3) agential porosity, where human and algorithmic decision-making become entangled in ways that resist clear agential identity. Rather than treating computation as a closed system of discrete operations, we demonstrate how porosity reveals the improvisational, threshold-crossing character of contemporary computational practice. This perspective challenges deterministic accounts of computational agency and opens space for understanding how computational systems might be sites of unexpected possibilities and unforeseen constellations, much as Benjamin and Lācis observed in Neapolitan street life.

Porosity in Naples

"Like 'aura,' indeed the term 'porosity' summons the spatial to the service of the cognitive and the historical; but aura belongs to the philosophical and indeed to the more narrowly aesthetic realm, while porosity tends toward social relations. '[So] dispersed, porous and commingled is private life ... [that] the true laboratories of this great process of intermingling are the cafes.' This sense of the overlap, the interpenetration, the fading of boundaries, is perhaps the optical illusion of the

Northern visitors, who need their differentiated categories in order to perceive and articulate their absence in Naples". (Jameson, 2020: 117)[1]

The seemingly anarchistic social organisation of 1920s Naples provided the context for Benjamin and Lācis's influential analysis of urban form and social life. We want to suggest that their observations of how architectural and social forms interpenetrate through material structures offers us a theoretical framework that, whilst historically specific to interwar Naples, provides remarkable insights for understanding contemporary computational systems. Before developing this connection, we want to examine how Benjamin and Lācis articulated their concept of porosity through their reading of Neapolitan architecture and social relations, and in particular their attention to the material specificity of tuff stone and its use as a metaphor to describe Neapolitan social organisation.

In their Naples essay, Benjamin and Lācis (1925/1996) depict the city as a living tableau of interpenetrating spaces and activities. They observe that in Naples, "No situation appears intended forever, no figure asserts it 'thus and not otherwise' (p.416)". Everything is in flux; boundaries between domains are fluid. The authors coin Porosity as the term for this "great process of intermingling" (p.421) in which private and public life are entangled. Porosity refers initially to the physical fabric of Naples. The city's buildings are constructed from tuff stone, a material full of holes and cavities. As Jameson (2020: 118) suggests: "The porosity of the stone is thus a material in which matter and emptiness meet and mix, just as already figural openings in the walls permit a mingled vision of inside and outside". Houses, courtyards, staircases and streets seem to bleed into one another. Doors and windows stand open; activities flow from inside to outside. Benjamin and Lācis describe cliff-side grottos carved into the rock that serve as both dwellings and gathering places, with "a door… here and there in the rock" leading into "large cellars, which are at the same time sleeping places and storehouses" (p.416). Even the line between solid architecture and the process of building or decay is blurred: "In such corners, one can scarcely discern where building is still in progress and where dilapidation has already set in. For nothing is concluded" (p.416). Porosity, then, is not just a literal quality of the stone, but an enacted principle of life in Naples. Benjamin and Lācis write, "Building and action interpenetrate in the courtyards, arcades, and stairways. In everything, they preserve the scope to become a theater of new, unforeseen constellations" (p.416). The city's physical spaces are continually transformed by human activity: a narrow alley might host an impromptu market; a staircase becomes a social forum; a balcony doubles as a theatre box. "The stamp of the definitive is avoided" (p.416); nothing in Naples is locked into a single permanent function.

Porosity manifests temporally as well. Benjamin and Lācis observe that festival time and workday time seep into each other in Naples. "Irresistibly, the festival penetrates each and every working day," they write; "Porosity is the inexhaustible law of life in this city, reappearing everywhere" (p.417). In other words, everyday life in Naples has a festive, improvisational character: "A grain of Sunday is hidden in each weekday. And how much weekday there is in this Sunday!". The essay opens with an anecdote of a priest being carted through the streets in punishment and accidentally blessing a wedding procession that crosses his path, bringing the crowd to its knees. Such incongruous scenes exemplify Naples's ability to hybridize the sacred and profane at any moment. Children play among church altars; vendors turn alleys into fairgrounds. This melding of times is an aspect of porosity,

1 It is worth comparing this reading with Glynn (2020) who argues that, "the critical concept of porosity advanced by Benjamin and Lācis should be re-evaluated not as a quality inherent to the city of Naples per se, as it has been received and deployed in critical and cultural work, but as the product of a Northern European gaze and of the encounter between a Northern European sensibility and its Southern Other at a precise historical moment" (Glynn, 2020: 16).

the city's time is not a linear schedule divided into work versus rest, but a continuum where work can become play and vice versa.

Ultimately, Benjamin and Lācis present Naples as a threshold city: a city of in-betweens. Thresholds are literal, the "inconspicuous door" covered by only a curtain that leads from a filthy courtyard into a "lofty, whitewashed church interior" in one step (p.416), and metaphorical, as when personal dreams merge with collective life on the streets. "His private existence is the baroque opening of a heightened public sphere," they write of the individual who crosses such a threshold. Naples thrives on these thresholds. The essay even remarks that side alleys give glimpses of hidden taverns and scenes, and one cannot tell where the private ends and public begins. Rather than being a defect, this porosity is generative: it creates a constant state of possibility. Life in Naples is lived ad hoc and improvised, open to "new, unforeseen constellations" (p.416) at any moment. The authors link this to a passion for improvisation among Neapolitans: "Porosity results not only from the indolence of the southern artisan, but also, above all, from the passion for improvisation, which demands that space and opportunity be preserved at any price" (p.416–417). In other words, leaving things unfinished or flexible, whether a building project or a daily plan, is what allows the sudden eruptions of creativity and eventfulness that characterize Naples. Porosity is thus both a structural condition and a cultural practice of embracing the intermediate, the transient, and the unexpected.

Benjamin revisited the theme of Naples's porosity in a radio broadcast he wrote and presented in 1930, as part of a youth-oriented programme (later collected in Radio Benjamin, 2014). This radio play "Naples", complements the 1925 essay with vivid scenes of Neapolitan everyday culture, reinforcing the idea of the city as a porous, improvisational environment. Benjamin the radio narrator virtually walks his audience through Naples, highlighting its noise, movement, and unexpected juxtapositions. For example, he describes the new underground train that had just opened, and how upon arriving he found the cars overrun by swarms of street urchins hanging from the doors and crowding the seats for a free ride (p.145). This humorous image of children commandeering the modern underground sets the tone, Naples constantly subverts the orderly or the modern with playful chaos. Even technological progress (the underground train) is immediately absorbed into the city's porous life as just another stage for popular antics.

Throughout the radio piece, Benjamin emphasizes how everyday life in Naples shades into spectacle and vice versa. He describes the Nativity scene displays that fill the streets around Epiphany (January 6th), where elaborate figurines are set up in competition; but tellingly the figures include not just Magi and shepherds, but also local types like "macaroni dealers, mussel sellers, and fishermen that accompany the manger scenes" (p.146); another example of porosity between the sacred story and the profane world. Festivals in Naples draw everyone in, Benjamin mentions the Piedigrotta festival in September, an ancient fertility feast turned modern song competition, during which almost every day there is something happening and even the very poorest participate. Popular music composition becomes a civic sport; "Naples showers renown on a gifted singer nearly as much as America does on a talented boxer," (p.150) Benjamin quips. Benjamin's radio play thus highlights how art, leisure and fame in Naples are democratized and woven into daily life.

By the end, Benjamin draws a conclusion about everyday life and festival life, writing "what's remarkable is how the two blend into each other, how every day the streets have something festive… and how even Sunday has something of a workday feel to it" (p.151). This closing insight mirrors the essay's statement that "a grain of Sunday is hidden in each weekday". The radio play thus reinforces porosity in more colloquial terms, showing a modern audience how Naples lives on the threshold; between work and play, waking and dreaming, order and chaos. Benjamin's radio portrayal also subtly invites a political reading, amid the colourful anecdotes, one senses how the Neapolitans' creative intermingling of life's spheres could suggest a form of resilience or quiet resistance to the disciplined routines imposed by industrial modernity. The city's very porousness, its refusal of fixed boundaries, hints at an alternative way of life where interruption and improvisation keep experience fresh and communal. Jameson (2020: 118), for example, describes the Neapolitans as "ancestors of Baudelaire's modern nineteenth-century Parisian crowds, who play so crucial a role in Benjamin's aesthetic and political imagination; but they are, so to speak, the real thing, collectivity as it existed before modernity, a glimpse into the premodern past" .

Benjamin's experience in Naples helped shape the very way he later conceived historical images in his work on the Paris arcades (Benjamin, 2002). As Adorno (2003) wrote many years later: "The essay on Naples marks a decisive breakthrough for Benjamin, and back then it had an indescribable effect on a few people, including me." In Naples, Benjamin already perceives that the truth of a social world emerges in its juxtapositions and overlaps, not in linear analysis. He observes how everyday life entwines opposites (indoor/outdoor, work/leisure), and he conveys this through poetic snapshots. Naples is contrasted with more modern, planned cities, it is "anarchic, embroiled, village-like in the centre, into which large networks of streets were hacked only forty years ago" (p.416). In a subsequent image, a teetering tenement "seems held together at the corners as if by iron clamps, by the murals of the Madonna," a literal image of religious past shoring up secular present. Such composite images, where disparate times coexist, are what Benjamin means by dialectical. They are pregnant with contradiction, the stability of faith next to the instability of the modern slum, or, say, the opulence of a baroque church entered through the poverty of a slum alley. By arresting our gaze on these juxtapositions, Benjamin allows the "constellation" to form in our mind. Eiland and Jennings (2014: 211–212) suggest,

> *" 'Naples' is an important text not just for the complexity of its view of the fabled city; it is also the text that introduces the prose form that Benjamin would utilize and refine over the next fifteen years, the Denkbild or 'figure of thought.' There is no discursive through-argumentation in 'Naples.' Instead, the observations and reflections are presented in paragraph-length clusters of thought revolving around a central idea. These central ideas recur at intervals through the essay so that the reader is challenged to repudiate constructs based on linear narrative in favor of constellations of literary figures and ideas (our emphasis)."*

In Naples, multiple times and meanings coexist, ancient traditions persist in modern street life, public and private spheres interweave. Each porous scene is a constellation of social forces; a wedding held in a crowded courtyard might be at once a sacred rite and a neighborhood spectacle, blurring old ritual and modern sociability. These constellations are proto-dialectical images, they freeze a moment of life but within it we detect tensions between tradition and modernity, individual and collective, structure and improvisation. Benjamin's later idea of "dialectics at a standstill" takes this insight to a philosophical level, any critical historical moment can be read as a porous interpenetration of past and present that must be seized in a single image. In the *Arcades Project*, Benjamin (2002) even describes his method as assembling "the whole through the monad of the fragment," treating each artifact or citation as a window into the total constellation of the 19th century.

Furthermore, porosity's emphasis on mixing what is hidden and what is visible, aligns with the way dialectical images bring hidden historical forces to light. In Naples, what is concealed (e.g. private life) constantly bleeds into public view; analogously, in a dialectical image, what is latent in the past (a never-realised utopian desire or a repressed conflict) suddenly becomes visible in the present in the form of a constellation. The "porous" way of seeing means attending to thresholds, cracks, and overlaps, exactly where dialectical insight occurs. Benjamin (2002) came to believe that to "awaken" from the myth of progress, one must look for the cracks in the historical continuum (the moments that do not fit the narrative of history). In Naples, the entire city was full of such cracks, both literal and metaphorical, through which alternative social possibilities seeped out. We can say that porosity in Naples offered Benjamin a model of historical consciousness where truth appears in the interstices, in the in-between spaces where oppositions meet.

Walter Benjamin's fascination with porosity in Naples also connects to a larger set of aesthetic principles he developed under the influence of Bertolt Brecht and other avant- garde practitioners (Wizisla, 2016). Central among these are the notions of estrangement (*Verfremdung*) and refunctioning (*Umfunktioniering*); techniques originally formulated in the context of radical theater and art, which Benjamin sought to apply to criticism and urban experience (Benjamin, 1934). *Verfremdung* is the outcome when an everyday situation is made to look unfamiliar and thus questionable. Benjamin embraced *Verfremdung* as a means of penetrating the mystifications of bourgeois culture. In epic theater, this might be achieved by having an actor step out of character or by presenting a historical event in a bizarre, anachronistic way. Benjamin admired surrealist art for a similar reason, by juxtaposing incongruous elements, surrealism created profane illuminations that estranged the familiar (a street, a room) and revealed the hidden dreamlike qualities of reality. The city of Naples provided naturally such a surreal tableau. Every person or object in Naples could suddenly become something else, step into a different scene: "No figure asserts it is 'thus and not otherwise'". This is estrangement not created by an artist on a stage, but by the organic social practices of the city. Benjamin treats Naples as a kind of collective artwork of estrangement, where everyday life continuously stages itself in theatrical ways. The result for northern observers is an awakening from their preconceptions: one realizes 'normal' life could be lived very differently. Benjamin's pre-occupation with thresholds ties into estrangement as well. At thresholds, perceptions can change quickly; one can experience a tiny shock of *Verfremdung* stepping through a door that leads from chaos to calm or vice versa.

The idea of refunctioning further ties together Benjamin's aesthetic-political programme with Brecht's praxis. Refunctioning (*Umfunktionierung*), a term Benjamin uses in his 1934 essay *The Author as Producer*, means taking existing artistic or technical forms and transforming their function toward new, progressive ends. It is about repurposing cultural tools (printing presses, theaters, radios, cameras, literature itself) for liberatory purposes rather than for bourgeois profit or passive consumption. Brecht, for example, refunctioned the theater by turning a venue of escapist entertainment into a forum for political education and critical thinking. Benjamin was convinced that artists must not only report on the struggles of the proletariat, but actively intervene in the means of cultural production to change who controls them and how they are used. We can see Benjamin's celebration of Naples's porosity as an appreciation for a naturally occurring refunctioning of spaces and rituals. Naples shows how the forms and instruments of daily life, such as buildings, streets, festivals, can be bent away from their official functions and made to serve the people's collective desire for amusement, sociality, or subsistence. This celebration of porosity implies a political vision, the refunctioning of the city from a machine for efficient living (the modern, hygienic, rationalized city) to a porous organism that encourages collective experiences, even uncomfortable or messy ones, that could galvanize social awakening.

Figure 1: Café Hiddigeigei on Capri where Benjamin and Lācis first met. Photo: Giorgio Sommer, ca. 1890 (Mittelmeier 2025).

Porosity in the Age of Algorithmic Management

We argue that Benjamin and Lācis's concept of porosity can be used to help understand how computational architectures structure contemporary social relations. In doing so we agree with Smith (2021: 255) that porosity has the interesting conceptual value of "problematiz[ing] in turn apparently stable cultural models that in contrast take themselves too much for granted". Just as Naples' built environment created porous boundaries between domains of life, computational systems generate multiple forms of interpenetration between human and machine agencies through layered infrastructures of code, data and automated decision making (see Berry, 2014: 58). The crucial difference is that computational porosity operates not through stone and concrete but through the material substrate of processors, networks and algorithms that increasingly mediate social existence. This includes the proliferation of enterprise software, algorithmic management systems, and platform-mediated labour that restructure how work is coordinated, controlled, and experienced in organisations. The aim is to deploy the concept of porosity in two ways, first as a *descriptive concept* which helps understand how discretisation as a practice within computation is giving way to diffusion techniques but also to employ porosity as a *critical concept* in the sense given by Benjamin and Lācis who saw it as an alternative to bourgeois ways of organising the lifeworld. In his later essay on theories of fascism Benjamin would warn of the appropriation of technology by the bourgeois society, pointing out "the gaping discrepancy between the gigantic power of technology and the minuscule moral illumination it affords. Indeed, according to its economic nature, bourgeois society cannot help but insulate everything technological as much as possible from the so-called spiritual, and it cannot help but resolutely exclude technology's right of codetermination in the social order" (Benjamin, 1930: 32).

This has implications for understanding contemporary organisational life. Just as Naples resisted the rationalised planning of modern cities, computational porosity challenges organizational boundaries and hierarchies. In platform organizations, the distinction between employee and contractor, workplace and home, working time and leisure time becomes increasingly porous. Uber drivers, for instance, exist in a deliberately porous space where they are neither fully independent nor fully employed, where the car becomes simultaneously private property and workplace, where algorithms interpenetrate with human decision-making about when and where to work (Rosenblat, 2019). Similarly, in remote work arrangements mediated by Slack, Zoom, and other platforms, organisational boundaries become porous in new ways. The office diffuses into domestic home spaces and synchronous and asynchronous communication blur together making corporate surveillance and individual autonomy clash through activity monitoring software and flexible scheduling.

Where Benjamin and Lācis examined how Naples' architecture enabled the constant movement between interior and exterior spaces, we can consider how computational systems create fluid boundaries between local and cloud processing, between human and machine cognition, and between private data and public circulation. The physical permeability they identified in Naples' buildings finds its contemporary parallel in the technical permeability of computational systems that allow data and processing to flow across previously distinct spheres and across planetary networks. For example, we might look at the structure of modern cloud computing infrastructures. These systems operate through what appears on the surface as a rigid separation between edge devices, core processing, and cloud storage (Berry 2023). Yet, in practice these boundaries are remarkably fluid. A smartphone application, for instance, continually negotiates the distribution of computational tasks between local processing and remote servers, creating infrastructural porosity.

Benjamin and Lācis's observations about Naples' architecture provide remarkably apt metaphors for understanding contemporary digital infrastructures when used as a descriptive concept. The way they describe how "building and action interpenetrate in the courtyards, arcades, and stairways" of Naples or how Naples' architecture is "a space for public life" that "extends into the living quarters" precisely captures how computational systems blur boundaries between public and private spheres. Similarly, in computational systems, code and social practices interpenetrate through interfaces, algorithms and networks, just as Neapolitan architecture created spaces that were hybrid where personal and collective activities were mixed together.

This computational porosity is shown through multiple, interrelated forms that structure contemporary digital culture. Social media platforms create porous boundaries between individual and collective experience through their underlying architectures of algorithmic feed curation, behavioural tracking, and automated content moderation. The private becomes immediately public, filtered through layers of machine learning that shape both what is visible and how it circulates, while these systems are simultaneously shaped by the social relations they mediate. The infrastructural dimension extends beyond individual platforms to encompass entire technical ensembles that facilitate the flow of data, processing, and automated agency across previously distinct domains, potentially dissolving the distinction between system and lifeworld.

When we issue a voice command to ChatGPT or another LLM, the computation flows seamlessly between device, data centre and cloud, creating what appears as a unified interaction but which actually traverses across multiple computational domains. This technical arrangement mirrors the interpenetration of spaces that Benjamin and Lācis observed in Naples, though now operating through digital rather than architectural forms. Similarly, the diffusion processes that many AI systems now implement, make all cultural works diffuse and hybrid within the latent spaces of their neural networks, a process Berry (2025) calls *diffusionisation*.[2] Indeed, this is

> *"a process through which cultural forms are probabilistically dissolved and reconstituted via computational diffusion processes. Through diffusionisation knowledge and cultural production becomes subject to what is called vector representation and latent space manipulation. These mathematical abstractions allow artificial intelligence systems to blend, morph and generate new cultural forms through probability distributions rather than deterministic rules or simple reproduction. This marks a profound shift from mere discretisation and encoding toward the autonomous generation of synthetic variations that have no original referent in human experience" (Berry, 2025).*

We argue that the relationship between computational porosity and diffusionisation reveals a key transformation in how digital systems process and reshape cultural content. While computational porosity describes the broader phenomenon of interpenetrating boundaries between human and machine agencies, diffusionisation represents a specific technical manifestation of this porosity within AI systems. Through diffusion models, cultural artefacts are not simply stored or processed but become porous themselves as their features, styles, and meanings blur and intermingle within the latent spaces of neural networks. This technical process of diffusionisation thus intensifies the porosity Benjamin and Lācis observed in Naples' architecture, as it operates not just on the level of infrastructure but on the very substance of cultural production itself.

When writing an email using Gmail's Smart Compose, we encounter a particularly revealing instance of computational porosity at work. The system does not simply suggest words, it creates a dynamic interpenetration of human intention and machine prediction that transforms the act of writing. As we compose, our thought processes become intertwined with algorithmic suggestions in ways that go beyond simple automation. The system learns from aggregate patterns of communication across millions of users, creating a kind of collective linguistic porosity where individual expression becomes mediated through statistically derived patterns. This porosity operates on multiple levels, between personal and collective expression, between human cognition and machine learning, and between private communication and Google's data infrastructure. Indeed, this example shows how computational porosity extends beyond the visible interface into complex infrastructures of data collection and processing. Each interaction with Smart Compose feeds back into Google's machine learning systems, creating a form of temporal porosity where past communications shape future suggestions. The social dimension of computational porosity becomes particularly visible in social platforms. Similarly, a personal photograph posted to Instagram immediately enters complex circuits of algorithmic classification, content moderation and automated distribution. The intimate becomes public through layers of computational mediation, recalling Benjamin and Lācis's observations about the interpenetration of private and public life, though now operating through digital infrastructure that shapes how content circulates and becomes visible.

---

2 The idea that porosity is now also an instrumental process, actuated through computational techniques for the diffusionisation of the lifeworld, raises interesting questions about how a practice of resistance can be integrated into the system. However, we want to suggest that porosity, as Benjamin and Lācis deploy it, points to the excess that cannot be captured fully, even when turned into a computational function. Thereby, computational porosity creates unforeseen lines of flight and potentials for resistance in social and political practice.

This porosity takes on particular significance within corporate or organisational boundaries. For example, algorithmic management systems used in warehouses, call centres, and gig economy platforms, where these systems create what we might call agential porosity, where human and machine decision-making become so entangled that attributing responsibility becomes difficult. When an Amazon warehouse worker's productivity is monitored by algorithmic systems that automatically generate warnings or employee work recommendations, who makes the decision to fire someone? The algorithm processes the data, the manager receives the recommendation, the worker's actions are shaped by real-time feedback from wearable devices. It seems to us that the decision emerges from the porous blurring of human judgment and computational processing. This is similar to Benjamin and Lācis's observation that in Naples "no figure asserts its 'thus and not otherwise'", but now operating through organisational infrastructure that distributes agency across human and non-human actors. This computational porosity obscures accountability whilst intensifying control and will create a number of difficulties unless reflexively understood.

Benjamin and Lācis note that "in everything they preserve the scope to become a theatre of new, unforeseen constellations" helps us to see how computational systems enable endless reconfiguration of social and technical relations. Cloud computing architectures, for instance, mirror the way Naples' buildings served multiple, fluid purposes. Just as a Neapolitan courtyard could transform from marketplace to theatre to social gathering space; cloud infrastructure dynamically reallocates computational resources – for example, compute, data, storage – across different tasks and purposes. Indeed, this mirrors how platform companies like Uber or TaskRabbit dynamically allocate workers to tasks, creating a porous labour market where workers' time and skills become fungible resources to be reallocated in real-time according to algorithmic optimisation.

Benjamin and Lācis describe how porosity results from "the passion for improvisation, which demands that space and opportunity be preserved at any price". This reflection might be used to question how computational systems, despite their apparent rigidity, are often written playfully, with creative coding and innovative designs that can open spaces for unexpected uses and improvisational practices. For example, APIs and software frameworks are often required to preserve the scope for new applications and uses that their original designers never anticipated (Marino 2020). Indeed, their observation that in Naples, "no figure asserts its 'thus and not otherwise'" speaks to the contingent nature of computational systems which, despite their deterministic logic, remain open to contingency and reinterpretation. Such that the "stamp of the definitive is avoided" in both Neapolitan architecture and computational infrastructure, though for different material and social reasons.

In organisational practice, this improvisational quality can be seen in what we might call *workaround cultures*. Just as Neapolitans used architectural porosity to evade official functions and create alternative uses, workers develop tactics to game algorithmic management systems, exploit platform vulnerabilities, or repurpose enterprise software for unintended purposes. For example, call centre workers might share strategies for maximising metrics whilst minimising actual work, Deliveroo riders might use geographic quirks in the algorithm to secure better-paying orders, and remote workers might use mouse or keyboard automation to simulate work activity to evade surveillance software. These practices reveal the porous character of seemingly rigid computational management systems. However, we must also recognise how organisational porosity often serves capital accumulation. The same flexibility that enables worker resistance also enables platforms to externalise costs, avoid employment obligations, and intensify exploitation through the blurring of work and non-work time.

Figure 2: Using AI diffusionisation image porosity to animate Asja Lācis's image (Image Credit: David M. Berry).Animated version of this figure available at https://stunlaw.blogspot.com/2024/12/porosity-and-computation.html

However, computational porosity is not merely analogous to architectural porosity. Rather, it represents an intensification and acceleration of the interpenetration of spaces and practices that Benjamin and Lācis observed. Contemporary computational systems do not simply enable movement between defined spheres but actively blur the boundaries between them. When we interact with AI systems or social media platforms, increasingly human and algorithmic agencies are diffused in complex ways. The "theatrical" dimension they identified in Naples' architecture becomes literalised in computational systems that transform every interaction into a performance that can be captured.

The political economy of contemporary platforms further demonstrates this porous character. Social media companies construct value-extraction architectures that create deliberate porosity between user activity and corporate profit. Every interaction, every piece of content, every social connection becomes raw material for algorithmic processing (through diffusionisation) and monetisation. The concept of explainability in AI systems provides another crucial example (Berry, 2024b). Current attempts to make algorithmic decisions "explainable" reveal the inherent tension between computational opacity and the need for public accountability (Berry, 2023). This creates epistemic porosity, where technical knowledge and democratic oversight must somehow coexist and interpenetrate.

These dynamics become particularly visible in the rise of what Kellogg et al. (2020) term "algorithmic management", where computational systems make or inform decisions about hiring, scheduling, performance evaluation, and the sacking of workers. Indeed, recruitment algorithms sometimes use systems that screen CVs creating a kind of "temporal porosity" between past hiring decisions (e.g. encoded in training data), present applications, and future workforce composition. The algorithm learns patterns from historical hiring, which often reinforce existing biases about gender, race, and class. When deployed, these algorithms sometimes prescribe these biases whilst appearing neutral and objective. The hiring manager who relies on algorithmic rankings thus becomes enmeshed in a porous decision-making process where human judgment and computational classification are deeply problematic. This raises questions about organisational accountability when discriminatory outcomes emerge from algorithmic hiring systems. Indeed, responsibility becomes distributed across the porous boundary between human and machine agency. The question then is raised, is the fault with the algorithm, the training data, the manager who trusted the ranking, or the organisation that implemented the system? This computational porosity might then cause serious issues for frameworks of organisational responsibility and legal liability. The challenge becomes finding ways to make algorithmic systems legible to public understanding without sacrificing their technical sophistication.

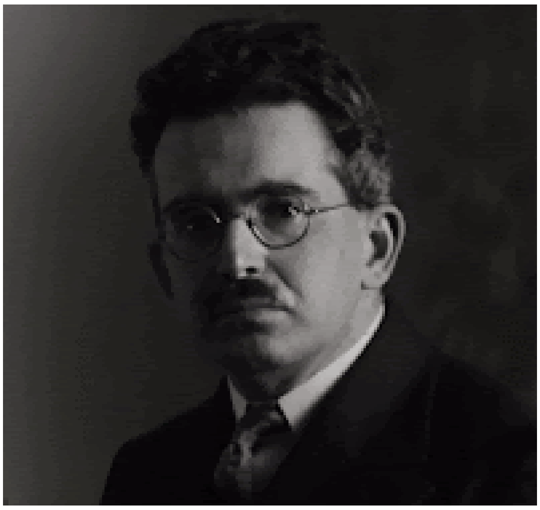


Figure 3: Using AI diffusionisation image porosity to animate Walter Benjamin's image (Image Credit: David M. Berry). Animated version of this figure available at https://stunlaw.blogspot.com/2024/12/porosity-and-computation.html

Yet, computational porosity as a critical concept also suggests possibilities for resistance and reappropriation. Just as Benjamin and Lācis identified how the porous architecture of Naples enabled creative forms of social improvisation, the layered nature of computational systems creates openings for critical intervention. Through practices of reverse engineering, hacking, jamming, adversarial machine learning, and algorithmic detournement (see Berry 2025), the seemingly rigid logics of computation reveal their contingent and contestable character. For example, in adversarial machine learning, researchers and activists can deliberately exploit the porous boundaries of AI systems to reveal their limitations and biases. This recalls Benjamin's (1930) attention to how Naples' street urchins used the city's new underground to subvert the purpose of this technology with playful chaos. Privacy-enhancing technologies, such as Signal, can also create deliberate impermeability within otherwise porous systems.

Figure 4: San Lorenzo Maggiore, a 13th century church in Naples, the foundations made of tuff stone.

Workers organising themselves into collectives increasingly use the same platforms that enable algorithmic management. For example, Amazon warehouse workers coordinate strikes through encrypted messaging apps, Uber drivers use Facebook groups to share information about surge pricing patterns, and domestic workers organised through platforms like Care.com develop collective strategies for negotiating better conditions. These practices demonstrate what Cant (2019) identifies as new forms of working-class power that operate through and against digital platforms. We see this when trade unions and worker centres increasingly demand algorithmic transparency and collective bargaining over the deployment of management algorithms. The European Union's *AI Act* and similar regulatory frameworks create new porous spaces between technical systems and collective governance, opening possibilities for workers to contest how algorithms organise their labour. This suggests that computational porosity, whilst often serving capitalist rationalisation, also creates openings for re-functioning through collective organisation and democratic control over workplace technology. These examples point toward computational porosity as a critical concept to examine how technical and social forces interpenetrate while remaining attentive to questions of power and resistance, echoing Benjamin and Lācis's dialectical sensibility.

Understanding computational porosity therefore requires moving beyond simplistic binaries of human versus machine agency to examine how technical and social forces interpenetrate at multiple scales. This suggests the need for new critical methods attentive to both the material specificity of computational systems and their embeddedness in broader political economic relations. This mapping of architectural and computational porosity reveals both continuities and differences in how technical systems shape social life. While the material substrate has changed from stone to silicon, the dynamic of interpenetration between technical and social forces remains.

The materiality of silicon in contemporary computation presents an interesting parallel to Naples' tuff stone. Modern semiconductor manufacturing relies on the controlled manipulation of silicon's porosity through processes like ion implantation and the creation of p-n junctions. The term "semiconductor" refers to silicon's porous electronic properties, that is its ability to sometimes conduct electricity and sometimes act as an insulator. Indeed, the actual manufacturing of silicon wafers requires extreme attention to porosity and contamination control in cleanroom environments. This material foundation of computational systems in the controlled management of silicon's porosity suggests a deeper connection to Benjamin and Lācis's analysis than mere metaphor. Just as Naples' architecture emerged from the properties of tuff stone, computational architectures are shaped by silicon's material properties and the technical practices required to control its porosity at the nanoscale (see Mody et al., 2017). Understanding these parallels helps us grasp both the constraints and possibilities created by contemporary computational architectures.

Returning to the Naples essay and radio play helps us understand what is at stake in our computational present. The analysis of Naples was not merely descriptive; it identified how the city's porous architecture enabled forms of life that resisted bourgeois rationalisation and bureaucratic control. Similarly, computational porosity presents us with a dialectical moment. Whilst computational systems create new forms of algorithmic governmentality and platform capitalism, their porous character potentially generates possibilities for alternative social arrangements; a "chance to correct the incapacity of peoples to order their relationships to one another in accord with the relationship they possess to nature through their technology" (Benjamin, 1930: 41).

The key question then becomes how to mobilise computational porosity towards democratic ends.3 Just as Naples' citizens used the city's porous spaces to create autonomous zones and informal economies, we might identify how computational porosity enables new forms of collective organisation and resistance. For instance, the porous boundaries between local and cloud computing could support decentralised infrastructure projects that prioritise community control over corporate profit. The diffusional character of contemporary AI systems might be redirected towards collective knowledge production rather than data extractivism.

Conclusion

Generative AI is an aesthetic technology of remix. It creates by collating and re-configuring patterns from its training data. This has an uncanny resonance with the artistic techniques of montage, collage, and citation that were central to the avant-garde movements Benjamin loved and to Brecht's epic theater. As we apply Benjamin's ideas, we can see generative AI's outputs as a form of involuntary surrealism as they often contain unexpected juxtapositions, distortions, and a Verfremdung-effect that can either enlighten or mislead, depending on context. Just as the Surrealists collaged disparate elements to jolt consciousness, AI often unwittingly collages fact and fiction. The key difference is intention, the Surrealists wanted to reveal truth's about the psyche and society through absurdity, whereas AI has no intention; it is just statistically blending. But a critical approach can appropriate these AI absurdities as productive estrangements. For example, some artists and writers

3 For example, Anderson (2019: 14) used the idea of porosity to think about "digital platform surveillance on social space"; particularly the interaction between the social and the technical he calls "Digital Porosity," a concept he draws from Zaporozhets (2016). In his time, "Benjamin also saw the possibility that broadcasting would satisfy the expectations of an audience that is contemporary with this technology" (Wizisla, 2016: 115).

deliberately use AI's weird outputs to spur creativity, treating AI systems like an aleatory collaborator. Large language models, trawling through billions of data points and recombining them, might surface hidden cultural obsessions or biases in strange new forms. Indeed, image generators trained on internet data often produce biased or stereotyped images, spuriously classifying people by race, gender, sexuality, and personality (Crawford & Paglen, 2021). When these biases appear blatantly in AI outputs, they can become an estranging mirror held up to society's prejudices. It makes visible what is often obscured in polished human-made media, the deep-set biases in our collective imaginary. Thus, AI's remix aesthetic can become a tool for critique, a way to see the "dream wishes" of society laid out unsparingly, much as Benjamin read the arcades of Paris as the dream wishes of the 19th century.

Benjamin was deeply preoccupied with how new media technologies could serve either emancipation or fascism. His friend Brecht and he discussed how radio, film, and photography (the new technology of their time) could all be used to enlighten the proletariat or deceive and pacify them (Wizisla, 2016). Generative AI similarly could produce estrangement or enchantment, it can make the familiar strange in a way that enlightens (as epic theater intended) or in a way that merely titillates and then numbs (as spectacle can do). In his 1936 essay, *The Work of Art in the Age of its Technological Reproducibility*, Benjamin (2008) famously argued that fascism responds to the masses' new cultural power by aestheticizing politics, staging grand spectacles whilst preserving existing property relations. In other words, fascism grants the masses an expression without rights, channeling their dreams through rallies and mythic imagery to distract from real social change. Benjamin cautioned that modern people, rendered "psychically porous" by consumer-capitalist mass culture, soaked up these extravaganzas as readily as any entertainment. Fascism's aestheticization of politics thus created a horrifying porosity between power and technology. The fascist rally becomes a space where individual boundaries dissolve into the mass, where the distinction between performer and audience collapses in ways that generate both ecstatic identification and profound terror. This is porosity weaponized, the same permeability that allows for genuine human connection becomes a vector for political domination.

Perhaps this illuminates something about our relationship to AI. Our anxieties about artificial intelligence often centre precisely on boundary violations, indeed AI can almost perfectly mimic human speech (e.g. dissolving the boundary between authentic and artificial communication), systems that know us better than we know ourselves (e.g. eroding the privacy of inner life), or technologies that make human labour obsolete (e.g. threatening the boundary between human purpose and mechanical function). Platform companies can exploit temporal porosity to dissolve the boundaries of the working day, making workers perpetually "available" through app notifications and demand-responsive scheduling. Benjamin's analysis suggests we should be attentive to how these fears might be manipulated. Just as fascism exploited the horror of dissolved boundaries to consolidate power, contemporary anxieties about AI porosity might be channeled toward particular political ends. The question becomes not just how to maintain boundaries, but how to cultivate forms of porosity that enable flourishing rather than domination. Indeed, porosity functions dialectically in workplace struggles as it simultaneously enables new forms of worker coordination and new modes of managerial control. Workers will need to increasingly engage in collective reverse-engineering of opaque systems, sharing knowledge about how algorithms calculate work, predict demand, or evaluate performance. A critical concept of porosity must therefore resist managerial appropriation by foregrounding questions of power, exploitation, and resistance.

The implications extend beyond technical systems to questions about the relations between system and lifeworld in an algorithmic age. The concept of *explainable forms of life* that Berry develops elsewhere takes on new significance in this context (Berry, 2024b, 2025). Rather than treating algorithmic opacity as a technical problem to be solved through better documentation or interfaces, we might understand it as a political question of how to create porous boundaries between expert and public knowledge (Berry 2021). This requires new institutional arrangements and technical practices that enable collective deliberation about how computational systems shape social life. Indeed, as Amoore et al. (2024: 3) suggest, "These computational architectures of generative AI models are increasingly penetrating political architectures… and delimit the ethico-political boundaries of what can be known and done in the world". Amoore uses the concept of interruption from Benjamin and Brecht to "locate the breaches in algorithmic arrangements and to show how they could be otherwise" and thus find traces of rejected alternatives in the arrangement of machine learning models. As she elaborates, "Here lies a significant form of resistance; to amplify the branching points as moments where things could have been otherwise, where other possibilities could be inferred; and to refuse the reduction of political difficulty to one that is the output" (Amoore, 2023: 35).

Benjamin's writings on Naples present a city as a living paradigm of porous boundaries and threshold experiences, where life's categories blend and spontaneous theater erupts in the streets; thus offering a counterimage to fascist dissolution of boundaries. In our time, generative AI systems have created new porous zones, between human and machine creativity, between reality and simulation. The porous city of Naples helps us recognize the porous society that AI is contributing to; one in which information flows unpredictably and the private and public, authentic and fabricated, continually intermix. Crucially, Benjamin teaches us to neither valorize such porosity naively nor condemn it outright, but to see in it a site of struggle and possibility. Naples was not an unqualified utopia; it had grinding poverty and exploitation amid its festive chaos as well as the ever present spectre of "the criminal world, the camorra" (Benjamin and Lācis, 1925: 414). Likewise, AI's porosity carries risks of exploitation (privacy leaks, cybercrime) even as it opens opportunities for new forms of collaboration and knowledge-sharing. This suggests that our relationship to AI might require learning to navigate beneficial forms of boundary crossing while resisting those that lead to subjugation or loss of agency. It is a signal to attend carefully to how boundaries are being dissolved and whose interests such dissolution serves. Our task, echoing Benjamin's, is to recognize the constellation, to see the tensions clearly, and choose practices that push AI's use toward emancipation.

REFERENCES

Adorno, T. W. (2003) So müßte ich ein Engel und kein Autor sein. In W. Schopf (Ed.), Der Briefwechsel mit Peter Suhrkamp und Siegfried Unseld. (p. 97.) Frankfurt am Main: Suhrkamp.

Amoore, L. (2023). Machine learning political orders. *Review of International Studies*, *49*(1), 20–36. https://doi.org/10.1017/S0260210522000031

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. *Political Geography*, *113*, 103134. https://doi.org/10.1016/J.POLGEO.2024.103134

Anderson, D.N. (2019) 'Digital Platforms, Porosity, and Panorama', *Surveillance & Society*, 17(1/2), pp. 14–20. Available at: https://doi.org/10.24908/ss.v17i1/2.12937.

Benjamin, W. (1930). 'Theorien des deutschen Faschismus: Zu der Sammelschrift 'Krieg und Krieger' herausgegeben von Ernst Jünger," *Die Gesellschaft*, 7 (2), 32–41.

Benjamin, W. (1999). The Author as Producer. In M. W. Jennings, H. Eiland, & G. Smith (Eds.), Walter Benjamin: Selected Writings Volume 2, part 2 1931–1934 (pp. 768–782). Belknap Press. (original work written in 1934)

Benjamin, W. (2008). The Work of Art in the Age of Its Technological Reproducibility. In M. W. Jennings, B. Doherty, & T. Y. Levin (Eds.), The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media (pp. 19–55). Harvard University Press. (original work written in 1936)

Benjamin, W. (2014). Radio Benjamin. Verso. (original radio play on Naples broadcast in 1930)

Benjamin, W., & Lācis, A. (1996). Naples. In M. Bullock & M. W. Jennings (Eds.), Walter Benjamin: Selected Writings Volume 1, 1913–1926 (pp. 414–421). Belknap Press. (original work published in 1925).

Berry, D.M. (2014) *Critical theory and the digital*. Bloomsbury Academic.

Berry, D. M. (2021) 'Explanatory publics: explainability and democratic thought', in B. Balaskas and C. Rito (eds) *Fabricating Publics: The Dissemination of Culture in the Post-truth Era*. London: Open Humanities Press, pp. 211–232. http://www.openhumanitiespress.org/books/titles/fabricating-publics/

Berry, D.M. (2023) 'The Explainability Turn', *Digital Humanities Quarterly*, 017(2). Available at: http://www.digitalhumanities.org/dhq/vol/17/2/000685/000685.html.

Berry, D. M. (2024a) Reflections on Method for Critical Code Studies, *Stunlaw*, https://stunlaw.blogspot.com/2024/12/reflections-on-method-for-critical-code.html

Berry, D. M. (2024b) Algorithm and code: explainability, interpretability and policy, in *Handbook on Public Policy and Artificial Intelligence*. Edward Elgar, pp. 134–146.

Berry, D.M. (2025) Synthetic media and computational capitalism: towards a critical theory of artificial intelligence, *AI & SOCIETY*. https://doi.org/10.1007/s00146-025-02265-2.

Cant, C. (2019) *Riding for Deliveroo: Resistance in the New Economy*. Polity Press.

Crawford, K., & Paglen, T. (2021). Excavating AI: the politics of images in machine learning training sets. AI and Society, 36(4), 1105–1116.

Eiland, H., & Jennings, M. W. (2014). Walter Benjamin: A Critical Life. Belknap Press.

Glynn, R. (2020) 'Porosity and its Discontents: Approaching Naples in Critical Theory', *Cultural Critique*, 107(1), pp. 63–98. Available at: https://doi.org/10.1353/cul.2020.0012.

Jameson, F. (2020). The Benjamin Files. Verso.

Kellogg, K.C., Valentine, M.A. and Christin, A. (2020) Algorithms at Work: The New Contested Terrain of Control, *Academy of Management Annals*, 14(1), pp. 366–410.

Lilla, M. (1995) 'The Riddle of Walter Benjamin', *The New York Review of Books*, 25 May. Available at: https://www.nybooks.com/articles/1995/05/25/the-riddle-of-walter-benjamin/ (Accessed: 12 December 2024).

Marino, M.C. (2020) *Critical code studies*. The MIT Press (Software studies).

McGill, J. (2008) 'The Porous Coupling of Walter Benjamin and Asja Lacis', *Angelaki*, 13(2), pp. 59–72. Available at: https://doi.org/10.1080/09697250802432120.

Mody C. C. M., Bijker W. E., Carlson W. B., Pinch T. (2017) *The Long Arm of Moore's Law: Microelectronics and American Science*. MIT Press.

Rosenblat, A. (2019) *Uberland: How Algorithms Are Rewriting the Rules of Work*. University of California Press.

Smith, D. (2021) 'Porosity and the Transnational: Travelling Theory between Naples and Frankfurt (Walter Benjamin, Asja Lacis and Ernst Bloch)', *Forum for Modern Language Studies*, 57(2), pp. 240–259. Available at: https://doi.org/10.1093/fmls/cqab001.

Wizisla, E. (2016). Walter Benjamin and Bertolt Brecht The Story of a Friendship. Verso.

Zaporozhets, O. (2016) 'Subway And Digital Porosity Of The City'. Available at: https://www.academia.edu/29375359/SUBWAY_AND_DIGITAL_POROSITY_OF_THE_CITY.

The status quo: Urban AI and the deepening of technocentrism in urban management

Author: Leandry Jieutsa, University of Montreal, leandry.junior.jieutsa.nkuidje@umontreal.ca
DOI: https://doi.org/10.54337/aau.add.scai-11426

KEYWORDS

Urban AI, technocentrism, Urban governance, smart city

ABSTRACT

This paper analyzes the emergence of a latent technocentrism in urban governance in the age of urban AI. While the failures and criticisms of the smart city have highlighted the limits technocentric governance, the increasing integration of AI in cities does not mark a retreat from it, but rather its reconfiguration. By analyzing the case of Quebec, this paper shows that this regime is not based on an explicitly claimed ideology, but on a convergence of structural factors. These include the perceived benefits of AI for cities, the multi-level ecosystem dynamics, and the increasing embedding of algorithmic systems in urban management. The paper argues for a re-politicization of urban AI by strengthening participatory mechanisms (citizens' assemblies, consultations, co-design) and the internal capacities of municipalities.

Introduction

The first generations of smart cities have been the subject of sustained criticism from academic, political and citizen circles (Kempin Reuter, 2020). Technocentrism characterized by technological solutionism, the digital divide between territories, the dependence of cities on external actors, and disconnection from local realities have been widely denounced (Appio et al., 2019; Calvo, 2020; Kempin Reuter, 2020; Kitchin, 2016; León & Rosen, 2020; Mora et al., 2017). In response, there has been a reflexive shift in urban governance frameworks that now emphasize accountability, participation, transparency, and the protection of digital rights (Beckers & Mora, 2025; Calzada, 2021; Malek et al., 2021; UN-Habitat, 2023).

This development indicates that technocentrism has been surpassed. However, the rise of artificial intelligence (AI) poses a challenge to this shift (Sadek et al., 2024). Urban AI is gradually establishing itself as a central technology in urban governance (Ben Dhaou et al., 2024; Cugurullo, 2020; Son et al., 2023). It seems to be a practical tool, focused on optimization, efficiency, and complexity management (Yigitcanlar et al., 2024). Cities employ this tool to enhance service provision, foresee risks, maximize resource usage, and facilitate decision-making (Bibri et al., 2024; Jieutsa & Koseki, 2025; Yigitcanlar et al., 2024). Although it is utilized for optimal urban governance, it is not a neutral technology (Sanchez et al., 2024). Algorithmic biases can exacerbate and perpetuate social and territorial inequalities (Ferrara, 2023; Jieutsa, 2024; O'Neil, 2016). The opacity of systems ("black box") limits transparency and accountability (Carabantes, 2020; Sanchez et al., 2024; Yigitcanlar et al., 2020). AI also contributes to the datafication and surveillance of urban spaces, posing risks to fundamental rights and the right to the city (Cugurullo et al., 2022). From an environmental point of view, deploying this system involves high energy and material consumption (Bibri et al., 2024; Ren, 2024). Additionally, the growing reliance on technology companies raises concerns about digital autonomy and democratic oversight (Cugurullo et al., 2022; Sadek et al., 2024).

This paper thus puts forth the inquiry of whether the implementation processes of urban AI perpetuate the prevailing notion of technocentrism, rooted in the concept of a smart city. To do this, we examine the case of Quebec using a qualitative approach based on a thorough review of 77 relevant publications from 2020 to 2025 and 13 in-depth interviews with local government officials involved in AI deployment or governance. The results show that urban AI contributes to the emergence of latent technocentrism that is less explicit and less ideological than that of the first smart cities. It is just as structuring. This technocentrism is not based on an assumed political project, but on a series of organizational and technical decisions that guide them towards efficiency and competitiveness. Urban AI tends to become a long-term part of administrative routines, preconfiguring certain decisions and structuring the very definition of public problems. The structural dependence on external actors, the centrality of performance logics and the low involvement of citizens in deployment processes reinforce this dynamic.

This study offers a theoretical and empirical contribution to discussions on the governance of urban technologies, such as urban AI. On the theoretical level, we introduce the concept of latent technocentrism to describe an intermediate dynamic between the explicit technocentrism of the first smart cities and responsible governance approaches. This shows that AI does not perpetuate an ideological technocentrism, but is gradually becoming institutionalized through pragmatic logics of efficiency, performance, and complexity management. On an empirical level, this paper offers a systematic analysis of the deployment of AI in Quebec municipalities.

Research background

From technocentrism in smart cities to responsible governance

Technocentrism in smart cities refers to an approach that places digital technologies at the center of urban governance (Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017). From this perspective, urban challenges are mainly understood as technical problems requiring technological solutions, relegating the social, cultural, political and human dimensions of public action to the background (Hu et al., 2023; Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017; Yigitcanlar et al., 2019). The technocentric model is thus opposed to human-centered or socio-centric approaches that favor citizen participation, equity and anchoring in local contexts (Hu et al., 2023; Kempin Reuter, 2020; Mora et al., 2017). Although it has fostered efficiency and rapid modernization of urban services, this model has attracted much criticism, particularly for its potential to reinforce inequalities and to prioritize corporate and institutional interests over those of ordinary citizens (Calvo, 2020; Kempin Reuter, 2020; Kitchin, 2016; León & Rosen, 2020; Mora et al., 2017; Söderström et al., 2014; Willis, 2019). It is characterized mainly by technological solutionism, the digital divide, the dependence of cities on external actors, and disconnection from local realities (Calvo, 2020; Hu et al., 2023; Kempin Reuter, 2020; Kitchin, 2016; León & Rosen, 2020; Mora et al., 2017; Söderström et al., 2014; Willis, 2019).

Technocentrism is primarily defined by the belief that digital technologies can offer solutions to a multitude of urban issues. This technological solutionism ideology, as outlined by Hollands (2015), Kitchin et al. (2017) and León & Rosen (2020), holds that technological innovations possess the power to address various urban challenges. In this light, devices such as the Internet of Things (IoT), Big Data or digital platforms are mobilized to analyze, optimize and rationalize urban management (Mora et al., 2017). Urban issues are translated into data flows that are massively collected, centralized and processed by algorithmic systems (Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017). This tendency to datafication tends to make quantitative performance the main criterion for evaluating public policies (Artyushina, 2020; Calvo, 2020; Hollands, 2015; Kempin Reuter, 2020; Krivý, 2018). As a result, urban problems are reformulated as optimization and calculation issues that can be solved by algorithms rather than by participatory processes (Artyushina, 2020; Calvo, 2020; Hollands, 2015; Kempin Reuter, 2020; Krivý, 2018). However, this emphasis on performance is criticized for its tendency to reduce complex urban realities to technical variables (Kempin Reuter, 2020).

A second component of technocentrism is increased dependence on external actors (Hollands, 2015; León & Rosen, 2020; Söderström et al., 2014). Technological solutionism is part of a multi-level epistemic community (Kitchin et al., 2017) made up of technology companies, consulting firms and experts, which shape the smart city agenda (Hollands, 2015; León & Rosen, 2020; Söderström et al., 2014). This configuration leads to a structural dependence of municipalities on private suppliers and external actors (Cardon & Crépel, 2019; Courmont & Le Galès, 2019a; Gong & Sun, 2024; Jiang et al., 2022; Przeybilovicz & Cunha, 2024) (Cardon & Crépel, 2019; Hollands, 2015; Jiang et al., 2022; Przeybilovicz & Cunha, 2024). Technological companies often take the lead in promoting proprietary solutions that may not match local needs or values (Hu et al., 2023; Kempin Reuter, 2020; Mora et al., 2017). Public-private partnerships emerge as key drivers of technological decision-making, at the expense of community-oriented approaches (Hollands, 2015; León & Rosen, 2020; Söderström et al., 2014).

Technocentrism also manifests as a disconnection from local realities and the needs of populations. Several authors highlight the gap between the expectations of residents and the technological devices deployed (Hu et al., 2023; Kempin Reuter, 2020; Mora et al., 2017). Technocentric models tend to concentrate decision-making between technical experts and corporate actors, limiting citizen involvement (Hu et al., 2023; Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017). This configuration can lead to the exclusion of vulnerable groups and the exacerbation of existing inequalities (Calvo, 2020). Surveillance technologies, algorithmic management of urban services, or automated prioritization of public interventions have been associated with forms of systemic discrimination (Gohari et al., 2022; Kempin Reuter, 2020; O'Neil, 2016; Willis, 2019). Consequently, rather than consistently fostering more inclusive urban landscapes, some smart city projects have actually exacerbated existing disparities (Kempin Reuter, 2020; Mora et al., 2017; Yigitcanlar et al., 2019).

Finally, technocentrism is accompanied by a digital territorial divide (Yigitcanlar et al., 2019). Cities that are economically developed are better able to adopt technocentric policies, while less endowed regions may be marginalized or adopt these strategies at uneven paces and capacities (Mora et al., 2017). This differentiation reinforces interterritorial disparities, widening the innovation and digital governance gap.

In response to the accumulation of these criticisms, cities have gradually adopted more cautious and reflective postures towards digital technologies. The emergence of discussions on accountable governance, the ethics of urban technologies, and citizen engagement reflects an effort to move beyond a purely technocentric perspective (Abed, 2019; Almulhim & Yigitcanlar, 2025; Beckers & Mora, 2025; Calzada, 2021; Malek et al., 2021; UN-Habitat, 2020). The principles of participation, transparency, respect for fundamental rights and protection of freedoms now occupy a central place in the governance frameworks of urban technologies (Ataman et al., 2025; Farida et al., 2023; Kolotouchkina et al., 2022; Rasoulzadeh Aghdam et al., 2025; UN-Habitat, 2020). These approaches, described as human-centric or responsible are based on co-production, feedback loops, civic deliberation and indicators oriented towards collective well-being (Ben Dhaou et al., 2024; Bou Nassar et al., 2025; Hu et al., 2023; Kolotouchkina et al., 2022; Rasoulzadeh Aghdam et al., 2025; UN-Habitat, 2023; Yigitcanlar et al., 2019). They aim to reposition citizens as co-creators of public services rather than as simple data providers (UN-Habitat, 2020). However, this does not mean abandoning initial motivations for adoption, such as the efficiency and innovation promised.

The emergence of Urban AI and the technocentrism statu quo

AI is now widely deployed across cities worldwide and is increasingly referred to in the literature as urban AI (Batty, 2023; Caprotti et al., 2024). Urban AI designates any "system integrating data from the urban environment, processed by algorithms, and whose results find useful applications in the socio-spatial fabric of the city" (Popelka et al., 2023, p. 14). It relies on the rapid, often real-time processing of large volumes of urban data generated by sensor networks and Internet of Things infrastructures, enabling enhanced urban management (Batty, 2023; Cugurullo, 2020; Luusua et al., 2022; Son et al., 2023).

Urban AI is applied to land-use planning, infrastructure management, mobility systems, housing, and public service delivery (Caprotti et al., 2024; Cugurullo & Xu, 2025; Heidari et al., 2022; Koseki et al., 2022; Son et al., 2023). While typologies vary, major categories include autonomous vehicles, robots, urban brains, and software agents (Cugurullo et al., 2024). Others classify AI according to planning phases with analytical, functional, textual, visual, and interactive AI (Othengrafen et al., 2025). Applications span mobility optimization, smart grids, environmental monitoring, infrastructure maintenance, and emerging generative AI simulations (Servicenow et al., 2025). AI thus permeates multiple urban domains, becoming a central component of contemporary urban management (Yigitcanlar et al., 2024).

The emergence of urban AI comes at a time when the first generations of smart cities are being questioned for their technocentrism (Kempin Reuter, 2020; Mora et al., 2017). More responsible governance frameworks are emerging (UN-Habitat, 2020). However, the implementation of urban AI challenges, if not contradicts, this aspiration for responsible governance. AI is not a neutral technology, as Sanchez et al. (2024) emphasize. Biases in databases and algorithmic models can reproduce and amplify urban inequalities, to the detriment of the most vulnerable populations (Cugurullo, 2020; Sanchez et al., 2024; Sherman, 2023). By strengthening surveillance devices in the public space, some AI applications can also infringe on fundamental rights and the right to the city (Cugurullo et al., 2022; Lefebvre, 1968; Yigitcanlar et al., 2020). In addition, the "black box" nature of many algorithmic systems makes their operation difficult to explain, limiting transparency and public understanding of their impacts (Carabantes, 2020; Sanchez et al., 2024; Sherman, 2023).

Urban AI may reinforce the tendencies of technocentrism, particularly through the use of solutionist logic. Complex challenges such as traffic congestion, security, or public service management are often reimagined as algorithmic optimization problems (Caprotti et al., 2024; Cugurullo & Xu, 2025; Heidari et al., 2022; Koseki et al., 2022; Son et al., 2023). Furthermore, Urban AI relies heavily on large amounts of data from sensors, platforms, and digital infrastructure, contributing to the growing phenomenon of datafication in public action (Batty, 2023; Caprotti et al., 2024; Popelka et al., 2023). The centrality of data could accentuate the dependence of municipalities on expensive technical infrastructure, energy-intensive data centers, and supply chains for critical material resources (Jin & Miles, 2025). It also reinforces the structural dependence on big tech companies, which design, deploy and maintain algorithms (Sadek et al., 2024; Sudmann, 2019). Despite the criticism leveled at the smart city model, the dominant position of private actors in the urban AI landscape remains largely unchanged, raising issues of digital sovereignty and democratic control. Moreover, the increasing role of AI systems in urban decision-making raises questions about the very foundations of responsible governance. When algorithms participate in prioritizing public interventions or allocating resources, questions of human control and accountability arise (Connecticut, 2023; Sanchez et al., 2024). Automated decision-making can create unpredictable or unfair effects that are difficult for citizens to challenge.

Contemporary discourse highlights a normative shift towards ethical and inclusive values (Rasoulzadeh Aghdam et al., 2025). Despite this, a persistent underlying technocentric structure exists. AI is increasingly central to urban management (Cugurullo et al., 2024). Cities are developing dedicated strategies and investing in improving algorithmic capabilities, while promoting the anticipated benefits of these technologies (Ben Dhaou et al., 2024; Servicenow et al., 2025). Municipal leaders' speeches, which are often imbued with technological optimism, help to legitimize the gradual integration of AI as an inevitable evolution of local public action. Therefore, the challenges of urban AI raise a fundamental question: do these technologies really make it possible to consolidate responsible governance? Or do they participate in a renewed reconfiguration of technocentrism in more sophisticated and data-intensive forms? AI doesn't just add a layer of technology; it redefines urban knowledge production, power relations and modes of participation. It tests the ability of cities to reconcile algorithmic innovation and democratic principles.

Method

This paper adopts an exploratory qualitative approach based on the case of Quebec (Creswell & Creswell, 2017; Yin, 2009). We chose this field because it has been recognized as having a great deal of dynamism in the field of AI (Conseil de l'innovation du Quebec, 2023). In particular, this region has a concentration of academic, entrepreneurial, and institutional players, which makes it a relevant laboratory for analyzing the dynamics of AI deployment in the region's municipalities.

The initial stage of the research involves a comprehensive examination of 77 pieces of literature, which were published between 2020 and 2025. The collection encompasses news articles, content from government websites, reports from technology companies, municipal records (strategic plans, press releases, internal policies), and publications from intermediary organizations. We have sourced these documents from municipal websites, AI-focused companies, as well as local and regional news outlets. The analysis of this corpus had two main objectives. First, to identify the actors involved in the urban AI ecosystem. Second, to identify the AI tools actually deployed, their fields of application and their implementation methods. This stage allowed for the mapping of relationships between actors and the identification of key technological tendencies (computer vision, supervised learning, generative AI, predictive systems, etc.).

The second phase is based on 13 semi-structured interviews with municipal actors directly involved in the deployment or governance of AI. The participants are 6 elected officials, 2 directors or senior managers in urban planning, 3 managers of information technology, AI or information systems, and 2 senior managers in municipal management. Each interview, which ranged from 30 to 60 minutes in length, explored various aspects. These include the existing applications of AI, the underlying drivers and decision-making processes behind these selections, the implementation strategies, the collaborations involved, and the anticipated consequences.

The data from the interviews were analyzed using a thematic approach. Coding was carried out in a hybrid way, combining an inductive and a deductive approach around the dimensions of technocentrism. This analytical strategy has made it possible to examine the extent to which municipal practices and discourses are part of a technocentric logic, or on the contrary, testify to an attempt to implement a more responsible governance of urban AI. By combining the documentary analysis (which provides a structural perspective on the ecosystem) with the interviews (which offer a situated viewpoint from the actors), a comprehensive understanding of the complexities of AI implementation in Quebec municipalities can be gained.

Results

Urban AI in Quebec municipalities

Canada has a strong reputation as a global hub for AI. In particular, there is a particularly high concentration in Quebec, particularly in Montreal. The country was the first to launch a dedicated national strategy (Pan-Canadian AI Strategy) in 2017. Between 2011 and 2022, the government invested nearly $385 million in over 4,000 AI research projects (Attema et al., 2025). The Quebec ecosystem is characterized by its close ties between universities, startups, incubators and public institutions, facilitated by initiatives such as Vitrine IA Québec. As a result, AI is used in various sectors.

At the municipal level, this technology is deployed in various fields. Municipalities are mobilizing a range of technologies from traditional Machine Learning (ML) to generative AI. This adoption affects infrastructure management, citizen services, mobility, the environment and internal administration. Cities are utilizing an array of technologies, from traditional machine learning to the latest generative AI. These technological approaches incorporate techniques such as supervised learning, computer vision, spatio-temporal modeling, natural language processing (NLP), and generative AI tools directly integrated into the services.

In the field of infrastructure, AI is mainly used for predictive maintenance and automated anomaly detection. Solutions such as City Rover use computer vision and Deep Learning (DL) models to detect cracks and potholes from images captured by cameras on municipal vehicles. For drinking water networks, CANN Forecast uses supervised learning algorithms and time-based statistical models to anticipate water consumption and prioritize interventions on risky behaviors. In winter service, Météo-Routes combines meteorological data, predictive models and decision-making support systems to optimise routes and the use of abrasives.

Citizen services are a second major field of application. Several municipalities have implemented chatbots, based on NLP and LLM techniques. These tools automate answers to frequently asked questions and guide users in their administrative procedures. Internally, various AI agents use generative AI to assist in the drafting of documents, the synthesis of information or regulatory analysis. In some cases, machine learning is also used for automated information extraction in invoice processing or call center support.

In mobility, municipalities use predictive models and real-time data analysis to optimize flows. Tools like Niosense rely on GPS data analysis and optimization algorithms to dynamically adjust the timing of traffic lights. Modelling platforms such as CIVILIA use geostatistics, simulation and machine learning to assess the impact of construction sites or plan cycling networks. Some smart bus stops also incorporate computer vision to measure ridership and detect anomalies. Cities like Quebec have partnered with big tech companies such as Google to optimize traffic management by synchronizing traffic lights under the Green Light Project.

Finally, in terms of the environment and climate resilience, AI supports territorial analysis and risk management. ACARA CLIMATE combines multi-source data and analytical models to assess climate vulnerability. Geospatial solutions like XEOS and K2 GEOSPATIAL leverage satellite image analysis, LiDAR, and DL to map heat islands and track the state of the urban canopy. Other tools use predictive modeling and scenario simulation to anticipate cascading effects during major crises.

Doing more with less: Combining performance and limited resources

The in-depth examination of the interviews uncovers a striking pattern. Urban AI is deployed in a setting characterized by a surge in citizen expectations, yet without a corresponding expansion in organizational capabilities. Several respondents explicitly describe this structural tension: "I have an increase of nearly 400% in citizen demand. But I haven't had more employees for 5 years, so I always have to do more with fewer resources." (ID0019). From this perspective, AI is perceived as a strategic lever for meeting the imperatives of efficiency and rationalization. As one manager puts it: "For us it makes all the difference. This is the lever we need. That's when I saw, in terms of resource optimization, how much we gain from working with AI" (ID007).

The tools deployed illustrate a reconfiguration of urban problems into objects that can be processed by algorithmic systems. Road defects are becoming an automated detection problem via devices such as computer vision and DL. The risks of water main breaks are reformulated as predictive issues through the ML. The dynamics of congestion or climate vulnerability are modelled by solutions such as Niosense or ACARA CLIMATE. This progressive algorithmization transforms urban uncertainty into a problem of data analysis, classification or optimization. Citizen requests are also transformed into tokens that are processed by NLPs and LLMs. Interactions between citizens and their municipalities are thus done through algorithms and generated responses.

However, municipalities do not present AI as a proprietary or deterministic solution. One respondent said: "When we use AI, it's because it was the most appropriate solution. That doesn't mean it was the only one we saw" (ID014). AI is described as one tool among others, mobilized in a pragmatic logic. It is designed as a support at work rather than a substitute: "The foundation behind it is to use this technology to increase the operational efficiency of the organization. […] So it becomes a support for the employee's work and it does not replace them » (ID009). This posture underlines an instrumental appropriation, oriented towards the improvement of processes.

Nevertheless, the analysis highlights a growing spread of AI as a near-default solution to save time and increase productivity. Chatbots, administrative assistants, or generative tools are now widely used, including in municipalities that do not have specialized tools. Respondents emphasize the substantial time gains. "If it takes you 3 days to make a nice presentation of 50 slides and then the other one next to you uses gamma and gives me back the 50 slides in 02h00, I could no longer afford for you to spend 3 days on it" (ID001). AI is even described as "the employee you don't need to hire" (ID007), reflecting the perception of a partial substitute for the shortage of human resources.

In addition, some actors emphasize its role in the preservation of organizational memory. "A city is a lot of information left and right [...] We lose a certain quality of organizational memory [...] And working with AI means that this memory doesn't disappear" (ID018). AI thus appears as a cognitive stabilization device in a context of increased staff mobility.

Finally, despite this largely positive vision "We will optimize everything to the maximum" (ID007), municipalities remain aware of the issues associated with AI. This is particularly true in terms of responsible use and supervision of generative AI. Several have adopted internal policies aimed at guiding practices and ensuring human control over algorithmic productions. AI is viewed as a potential solution to the current challenges facing local public action. It is not considered a goal in itself, but rather as a preferred tool or even a default option for converting urban issues into technological solutions focused on efficiency, rapidity, and resource optimization. However, municipalities admit that the challenges posed by the technology would not limit their deployment because for them the city has pressing and constant needs that must be addressed and urban AI makes it possible to optimize their response.

Emerging deployment and structural dependence of cities

The analysis of the ecosystem shows that the deployment of urban AI concerns municipalities of various sizes, but remains quantitatively limited at the provincial level. Of all the municipalities in Quebec, 172 have been identified as using various AI tools, while about 32 have more advanced tools (computer vision, predictive modeling, simulation, integrated systems, etc.). Compared to the total number of Quebec municipalities (1123), this level of adoption remains relatively low. Urban AI therefore appears to be an emerging and still unevenly distributed phenomenon, concentrated in certain more dynamic territories, more specifically those with sufficient resources. However, the rollout cuts across the three main municipal categories: very small municipalities (less than 10,000 inhabitants), medium-sized municipalities (10,000–100,000 inhabitants) and large municipalities (more than 100,000 inhabitants). Medium-sized municipalities occupy a central position. They have sufficient financial and organizational resources to invest in technology solutions, but they do not have the full in-house teams of major cities. Larger municipalities, on the other hand, tend to develop in-house solutions or build structured AI strategies, while smaller ones remain more dependent on standardized tools.

Despite this diversity, municipalities are highly dependent on external actors for the identification, selection and deployment of AI solutions. Technology companies play a central role in the development of tools, but also in raising awareness and training. Several solutions have emerged following training offered to municipalities. As one participant pointed out, “It was at this training [...] that I was hooked” (ID010). Mainly local companies structure the available offer and guide uses. NPOs and municipal groups such as Regional County Municipality (RCMs) and metropolitan communities are also strategic intermediaries. They act as facilitators, coaches and especially NPOs sometimes as an “external innovation department” for municipalities without in-house expertise. One participant said that “Going through organizations allows us to sort out and achieve these goals, because we lack the resources to do it ourselves” (ID007). These intermediaries make it possible to circumvent the constraints of public procurement rules and reduce the risks associated with the technological choice. Their role goes beyond simple mediation because they influence the dissemination of solutions and contribute to structuring regional adoption dynamics.

Educational and research institutions are mainly involved in research and development, as well as in responsible governance. However, their specific involvement in municipal dynamics remains limited in view of the density of the Quebec AI ecosystem. Government agencies, on the other hand, exert indirect influence through general funding of innovation, without a policy specifically dedicated to municipal AI.

Discussion

The results suggest that the governance of urban AI in Quebec does not mechanically reproduce the technocentrism of the first generations of smart cities. Nevertheless, they reveal the existence of what we introduce as a latent technocentrism, which is more diffuse and institutionalized. This is gradually becoming part of administrative routines and municipal digital infrastructures.

Municipal stakeholders present urban AI as a pragmatic optimization tool, intended to improve the performance, predictability and efficiency of urban management. This vision is part of a technophile imaginary where AI is perceived as capable of “steering” the city at different scales, in a quasi-autonomous way (Cugurullo & Xu, 2025; Echeverría & Tabarés, 2017). The speeches highlight competitiveness, sustainability and the modernization of public action, thus helping to normalize the integration of this technology. These logics are similar to those identified by scholars criticizing the normalization of technologies in urban governance (Hollands, 2015; Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017). The vision of urban AI held by municipal stakeholders is not merely idealistic, but pragmatic. It contributes to fostering the perception that AI, despite its challenges, is a functional technology that enables significant gains. While certainly pragmatic, this current vision also contributes to the normalization of technology through peer pressure (Gong & Sun, 2024; Rodriguez Müller et al., 2025). This leads to a solutionism driven not by corporations, as in the case of smart cities, but by the municipal stakeholders themselves. These people are in favor of the time and productivity gains that the technology offers, especially since urban AI not only provides assistance but also operates in a near autonomous way (Cugurullo, 2020).

Indeed, the analysis shows that algorithms no longer only assist the decision but preconfigure it. They define relevant categories, prioritize, anticipate risks, and guide interventions. Public action then tends to become a process of algorithmic supervision, where the administration monitors and adjusts automated systems rather than collectively deliberating on the ends. It is similar to recent studies on anticipatory governance where AI is embedded in urban governance and transforming into an infrastructure of power, shaping the city continuously and often invisibly (Cugurullo, 2020; Cugurullo & Xu, 2025; Son et al., 2023).

Contrary to some assumptions of a territorial digital divide (Samsurijan et al., 2023; Yigitcanlar et al., 2019), the results show that municipalities of various sizes are deploying urban AI tools. Technological accessibility, especially through subscription models and turnkey solutions, limits the structural exclusion of small municipalities. However, adoption remains low compared to the total number of Quebec municipalities. This partial diffusion seems to be less related to access than to organizational and cultural factors. These include fear of job loss, resistance to change, technological skepticism or competing budget priorities (Yigitcanlar et al., 2023). Latent technocentrism is therefore not uniformly imposed but progresses where institutional, cultural and political conditions allow it.

A central result is the structural dependence of municipalities on external actors. The same companies, NPOs and intermediaries return to the identified projects, acting as external offices of the smart city. This is similar to previous smart cities governance practices where external actors frame urban policies (Hollands, 2015; Kitchin et al., 2017; Mora et al., 2017). In the case of Quebec, although many are NPOs, their distribution logics are similar to market dynamics. This configuration favors the emergence of an epistemic community of urban AI (Kitchin et al., 2017), committed to the promotion of performance-oriented and optimization-oriented solutions. Municipalities integrated into a multi-level ecosystem (states, international organizations, research centers, companies), become both beneficiaries and testing grounds. International discourses associating AI with sustainable development, innovation and modernization help to legitimize its local adoption (Gong & Sun, 2024; Rodriguez Müller et al., 2025). In this context, technocentrism is not the result of an explicit decision, but of an institutional trajectory and ecosystem dynamics in which municipalities operate.

Another factor contributing to this latent technocentrism is the gradual confinement of municipalities to approaches that limit participation. The results show that tools deployed respond to local challenges but remain mainly oriented towards internal performance. Citizen participation remains marginal in the lifecycle and this is an important critic on smart cities that is still present in AI governance practices (Hu et al., 2023). This can therefore reinforce technocentrism where municipalities are focused on optimization in service delivery through disruptive technologies such as AI without citizen-centric approaches.

Technocentrism can also be reinforced by defining urban problems solely through the lens of technology Kempin Reuter, 2020; León & Rosen, 2020; Mora et al., 2017). The increasing integration of AI into decision-making processes tends to redefine public problems themselves. When algorithms determine priorities or model risks, they implicitly influence what is considered a legitimate problem and an acceptable response (Batty, 2024). This dynamic can lead to a gradual disconnection from local needs and an institutionalization of technological solutionism. AI is touted as capable of reducing urban complexity, anticipating uncertainty, and achieving sustainability goals (Son et al., 2023). Despite a declared awareness of the risks (bias, opacity, environmental impacts, etc.), control mechanisms remain limited in the face of the “black box” nature of the systems. Cities don’t always have the expertise to fully understand the algorithms they use, increasing vendor lock-in.

Thus, the governance of urban AI is not based on an explicit and assumed technocentrism. Rather, it is part of an operational and latent form of technocentrism, where Urban AI is gradually becoming an ordinary, even unavoidable, component of urban management. This technocentrism is less ideological than pragmatic, less spectacular than that of smart cities, but just as structuring. Caught up in a multi-level dynamic of AI valorization, and practical benefits of the technology, municipal actors tend to perceive its integration as rational, necessary and inevitable. The challenge is therefore not only to regulate AI, but to politicize its uses again, by strengthening participatory mechanisms (citizens' assemblies, consultations, co-design) and the internal capacities of municipalities. Without these rebalancings, AI risks imposing itself not as a deliberate choice, but as an institutional trajectory that is difficult to contest, consolidating a diffuse technocentrism at the heart of contemporary urban governance.

Conclusion

The objective of this research was to examine whether the deployment of urban AI in Quebec municipalities reproduces technocentric governance. The literature review and interviews reveal a nuanced picture. AI is now integrated into municipalities of various sizes and mobilized mainly for the objectives of administrative optimization, infrastructure management and service improvement. It is perceived in a pragmatic way, as an effective tool to "do more with less". Cities do not see it as an autonomous end, and many demonstrate an awareness of the ethical and organizational issues associated with its use.

However, the results show that this gradual integration is accompanied with what we describe as latent technocentrism. AI tends to become a permanent part of administrative routines, to preconfigure certain decisions and to structure the very definition of public problems. The structural dependence on external actors, the centrality of performance logics and the low citizen involvement in the deployment processes reinforce this dynamic. Thus, AI does not reproduce an explicit and ideological technocentrism, but it contributes to reconfiguring its forms, in a more diffuse, pragmatic and institutionalized way.

However, there are some limitations. The 77 documents analyzed, although from a long watch (2020–2025), do not necessarily reflect the exhaustiveness of an ever-evolving ecosystem. To address this issue, we extended our observation period and conducted interviews with key municipal actors to expand our literature review and gain a more comprehensive understanding of ongoing dynamics. Nevertheless, some emerging or informal initiatives may have escaped analysis. Future research could extend the investigation to other territories to compare urban AI adoption and governance trajectories. A comparative approach would provide a better understanding of whether the latent technocentrism observed in Quebec is a generalized or contextual trend. In addition, it seems essential to explore more systematically the mechanisms likely to limit this technocentric drift: mechanisms for citizen participation, strengthening the internal capacities of municipalities, alternative models of technological development or more restrictive governance frameworks.

As AI becomes entrenched in urban practices and imposes itself as an ordinary management infrastructure, the question is no longer just whether it reproduces technocentrism, but how to orient its integration so that it remains compatible with democratic principles and the collective goals of local public action.

REFERENCES

Abed, A. (2019). Designing with informality: Towards an urban design framework for Yeoville's main street. In *POLITICS AND COMMUNITY-BASED RESEARCH: Perspectives from Yeoville Studio, Johannesburg* (p. 313). (WOS:000711854100023). https://doi.org/10.18772/22019103849

Almulhim, A. I., & Yigitcanlar, T. (2025). Understanding Smart Governance of Sustainable Cities: A Review and Multidimensional Framework. *Smart Cities, 8*(4), 113. https://doi.org/10.3390/smartcities8040113

Appio, F. P., Lima, M., & Paroutis, S. (2019). Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting and Social Change, 142*, 1–14. https://doi.org/10.1016/j.techfore.2018.12.018

Artyushina, A. (2020). Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics, 55*, 101456. https://doi.org/10.1016/j.tele.2020.101456

Ataman, C., Herthogs, P., Tunçer, B., & Perrault, S. (2025). From insight to action: An integrated assessment framework for digital citizen participation in data-centric urban practices. *Cities, 156*, 105545. https://doi.org/10.1016/j.cities.2024.105545

Attema, G., Mertz, M., Anawati, A., Austin, A., Bertrand, J., Jewett, R., & Cameron, E. (2025). *Public Funding for AI in Canada 2011–2022: An equity-focused environmental scan*. Scientific Communication and Education. https://doi.org/10.1101/2025.01.31.635792

Batty, M. (2023). *The emergence and evolution of urban AI*. (38), 1045–1048.

Batty, M. (2024). *The computable city: Histories, technologies, stories, predictions*. The MIT Press.

Beckers, D., & Mora, L. (2025). Overcoming the Smart City Governance Challenge: An Innovation Management Perspective. *Journal of Urban Technology, 32*(2), 85–106. https://doi.org/10.1080/10630732.2025.2461983

Ben Dhaou, S., Isagah, T., Distor, C., Ruas, I. C., & Programme (UN-Habitat), U. N. H. S. (2024). *Global Assessment of Responsible Artificial Intelligence in Cities: Research and recommendations to leverage AI for people-centred smart cities*. United Nations Human Settlements Programme (UN-Habitat). https://unhabitat.org/sites/default/files/2024/08/global_assessment_of_responsible_ai_in_cities_21082024.pdf

Bibri, S. E., Huang, J., Jagatheesaperumal, S. K., & Krogstie, J. (2024). The synergistic interplay of artificial intelligence and digital twin in environmentally planning sustainable smart cities: A comprehensive systematic review. *Environmental Science and Ecotechnology, 20*, 100433. https://doi.org/10.1016/j.ese.2024.100433

Bou Nassar, J., Calleja-López, A., Sharp, D., Anwar, M., Bartram, L., & Goodwin, S. (2025). Characterising and reassessing people-centred data governance in cities. *Frontiers in Sustainable Cities, 6*, 1518618. https://doi.org/10.3389/frsc.2024.1518618

Calvo, P. (2020). The ethics of Smart City (EoSC): Moral implications of hyperconnectivity, algorithmization and the datafication of urban digital society. *Ethics and Information Technology, 22*(2), 141–149. https://doi.org/10.1007/s10676-019-09523-0

Calzada, I. (2021). The right to have digital rights in smart cities. *Sustainability (Switzerland), 13*(20). https://doi.org/10.3390/su132011438

Caprotti, F., Cugurullo, F., Cook, M., Karvonen, A., Marvin, S., McGuirk, P., & Valdez, A.-M. (2024). Why does urban Artificial Intelligence (AI) matter for urban studies? Developing research directions in urban AI research. *Urban Geography, 45*(5), 883–894. https://doi.org/10.1080/02723638.2024.2329401

Carabantes, M. (2020). Black-box artificial intelligence: An epistemological and critical analysis. *AI & Society, 35*(2), 309–317.

Cardon, D., & Crépel, M. (2019). Les algorithmes et la régularisation des territoires. In A. Courmont & P. Le *Galès, Gouverner la ville numérique*. la Vie des idées.fr PUF.

Connecticut, S. of. (2023, July). *Act concerning artificial intelligence, automated decision-making and personal data privacy* (pp. 1–10). Governor of Connecticut.

Conseil de l'innovation du Quebec. (2023). *État de situation: Les impacts sociétaux de l'IA sur la démocratie, l'environnement et les arts et la culture*. Conseil de l'innovation du Quebec. https://conseilinnovation.quebec/wp-content/uploads/2023/10/CIQ_Impacts_societe_IA_EDS-1.pdf

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cugurullo, F. (2020). Urban Artificial Intelligence: From Automation to Autonomy in the Smart City. *Frontiers in Sustainable Cities, 2*. https://doi.org/10.3389/frsc.2020.00038

Cugurullo, F., Caprotti, F., Cook, M., Karvonen, A., McGuirk, P., & Marvin, S. (2024). The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence. *Urban Studies, 61*(6), 1168–1182. https://doi.org/10.1177/00420980231203386

Cugurullo, F., McCarroll, C., Valdez, M., & smarvin, S. M. (2022). *Urban AI in China: Social control or hyper-capitalist development in the post-smart city?*

Cugurullo, F., & Xu, Y. (2025). When AIs become oracles: Generative artificial intelligence, anticipatory urban governance, and the future of cities. *Policy and Society, 44*(1), 98–115. https://doi.org/10.1093/polsoc/puae025

Echeverría, J., & Tabarés, R. (2017). Artificial Intelligence, Cybercities and Technosocieties. *Minds and Machines, 27*(3), 473–493. https://doi.org/10.1007/s11023-016-9412-3

Farida, I., Ningsih, W., Lutfiani, N., Aini, Q., & Harahap, E. P. (2023). Responsible Urban Innovation Working with Local Authorities a Framework for Artificial Intelligence (AI). *Scientific Journal of Informatics, 10*(2), 121–126. https://doi.org/10.15294/sji.v10i2.43818

Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), 3.

Gohari, A., Ahmad, A. B., Rahim, R. B. A., Supa'at, A. S. M., Abd Razak, S., & Gismalla, M. S. M. (2022). Involvement of Surveillance Drones in Smart Cities: A Systematic Review. *IEEE Access, 10*, 56611–56628. https://doi.org/10.1109/ACCESS.2022.3177904

Gong, Y., & Sun, H. (2024). Exploring Factors Influencing Artificial Intelligence Adoption in Smart Cities. *Proceedings of the 2024 15th International Conference on E-Business, Management and Economics*, 343–349. https://doi.org/10.1145/3691422.3691435

Heidari, A., Navimipour, N. J., & Unal, M. (2022). Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review. *Sustainable Cities and Society, 85*, 104089. https://doi.org/10.1016/j.scs.2022.104089

Hollands, R. G. (2015). Critical interventions into the corporate smart city. *Cambridge Journal of Regions, Economy and Society, 8*(1), 61–77. https://doi.org/10.1093/cjres/rsu011

Hu, W., Wang, S., & Zhai, W. (2023). Human-centric vs. technology-centric approaches in a top-down smart city development regime: Evidence from 341 Chinese cities. *Cities, 137*, 104271. https://doi.org/10.1016/j.cities.2023.104271

Jiang, H., Geertman, S., & Witte, P. (2022). Smart urban governance: An alternative to technocratic "smartness." *GeoJournal, 87*(3), 1639–1655. https://doi.org/10.1007/s10708-020-10326-w

Jieutsa, L. (2024). Fairness and non-discrimination in urban AI. In M. G. Vercher & A. V. D'oleo (Eds.), *Ethical urban AI in practice: Policy mechanisms to establish local governance frameworks* (CIDON edicions, Vol. 89, pp. 39–48). https://www.cidob.org/en/publications/ethical-urban-ai-practice-policy-mechanisms-establish-local-governance-frameworks

Jieutsa, L., & Koseki, S. (2025). Multilevel Collaborative Governance to Foster Responsible AI Deployment in Cities: Case Studies of eThekwini in South Africa and Thérèse-De Blainville in Quebec. In P. Hacker (Ed.), *Oxford Intersections: AI in Society* (1st ed.). Oxford University Press-Oxford. https://doi.org/10.1093/9780198945215.003.0034

Jin, N., & Miles, I. (2025). The Dynamics of AI Innovation Ecosystems: A Case Study of Greater Manchester. In R. Zimmermann, J. C. Rodrigues, A. Simoes, & G. Dalmarco (Eds.), *Human-Centred Technology Management for a Sustainable Future* (pp. 161–173). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72494-7_17

Kempin Reuter, T. (2020). *Smart City Visions and Human Rights: Do They Go Together?*

Kitchin, R. (2016). The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2083), 20160115. https://doi.org/10.1098/rsta.2016.0115

Kitchin, R., Coletta, C., Evans, L., Heaphy, L., & MacDonncha, D. (2017). *Smart cities, urban technocrats, epistemic communities and advocacy coalitions*. https://doi.org/10.31235/osf.io/rxk4r

Kolotouchkina, O., Barroso, C. L., & Sánchez, J. L. M. (2022). Smart cities, the digital divide, and people with disabilities. *Cities, 123*, 103613. https://doi.org/10.1016/j.cities.2022.103613

Koseki, S., Farnadi, G., David, R., Catherine, R., Amanda, L., Bezenac, C. de, Occhini, G., Lefebvre, H., Gallego-Posada, J., Chehbouni, K., Molamohammadi, M., Sefala, R., Salganik, R., Yahaya, S., Téhinian, S., Kizhakkethottam, C., Lahoud, C., Pienaar, H. J., Thung, I., … Prud, B. (2022). *AI and Cities: Risks, applications and governance* (pp. 1–124). UN-Habitat. https://unhabitat.org/ai-cities-risks-applications-and-governance

Krivý, M. (2018). Towards a critique of cybernetic urbanism: The smart city and the society of control. *Planning Theory, 17*(1), 8–30. https://doi.org/10.1177/1473095216645631

Lefebvre, H. (1968). *The Right to the City*. https://theanarchistlibrary.org/library/henri-lefebvre-right-to-the-city.lt.pdf

León, L. F. A., & Rosen, J. (2020). Technology as Ideology in Urban Governance. *Annals of the American Association of Geographers, 110*(2), 497–506. https://doi.org/10.1080/24694452.2019.1660139

Luusua, A., Ylipulli, J., Foth, M., & Aurigi, A. (2022). Urban AI: understanding the emerging role of artificial intelligence in smart cities. *AI and Society*, (0123456789). https://doi.org/10.1007/s00146-022-01537-5

Malek, J. A., Lim, S. B., & Yigitcanlar, T. (2021). Social Inclusion Indicators for Building Citizen-Centric Smart Cities: A Systematic Literature Review. *Sustainability, 13*(1), 376. https://doi.org/10.3390/su13010376

Mora, L., Bolici, R., & Deakin, M. (2017). The First Two Decades of Smart-City Research: A Bibliometric Analysis. *Journal of Urban Technology, 24*(1), 3–27. https://doi.org/10.1080/10630732.2017.1285123

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (First edition). Crown.

Othengrafen, F., Sievers, L., & Reinecke, E. (2025). From Vision to Reality: The Use of Artificial Intelligence in Different Urban Planning Phases. *Urban Planning, 10*, 8576. https://doi.org/10.17645/up.8576

Popelka, S., Zertuche, L. N., & Beroche, H. (2023). *Urban AI Guide*. Urban AI. https://urbanai.fr/our-works/urban-ai-guide/

Przeybilovicz, E., & Cunha, M. A. (2024). Governing in the digital age: The emergence of dynamic smart urban governance modes. *Government Information Quarterly, 41*(1), 101907. https://doi.org/10.1016/j.giq.2023.101907

Rasoulzadeh Aghdam, S., Bababeimorad, B., Ghasemzadeh, B., Irani, M., & Huovila, A. (2025). Social smart city research: Interconnections between participatory governance, data privacy, artificial intelligence and ethical sustainable development. *Frontiers in Sustainable Cities, 6*, 1514040. https://doi.org/10.3389/frsc.2024.1514040

Ren, S. (2024, July). *The uneven distribution of AI's environmental impacts*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts

Rodriguez Müller, A. P., Tangi, L., & Lerusse, A. (2025). Understanding the Adoption of Artificial Intelligence in Local Government Decision-Making: The Influence of Institutional Pressures and Managerial Perceptions. *Public Administration,* padm.70033. https://doi.org/10.1111/padm.70033

Sadek, M., Kallina, E., Bohné, T., Mougenot, C., Calvo, R. A., & Cave, S. (2024). Challenges of responsible AI in practice: Scoping review and recommended actions. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01880-9

Samsurijan, M. S., Ebekozien, A., Azazi, N. A. N., Shaed, M. M., & Badaruddin, R. F. R. (2023). Artificial intelligence in urban services in Malaysia: A review. *PSU Research Review*. https://doi.org/10.1108/PRR-07-2021-0034

Sanchez, T. W., Brenman, M., & Ye, X. (2024). The Ethical Concerns of Artificial Intelligence in Urban Planning. *Journal of the American Planning Association, 0*(0).

Servicenow, Deloitte, Nvidia, & Thought Lab. (2025). *AI-powered cities of the future.* https://www.deloitte.com/global/en/Industries/government-public/research/ai-powered-cities-of-future.html

Sherman, S. (2023). The Polyopticon: A diagram for urban artificial intelligences. *AI and Society, 38*(3), 1209–1222. https://doi.org/10.1007/s00146-022-01501-3

Söderström, O., Paasche, T., & Klauser, F. (2014). Smart cities as corporate storytelling. *City, 18*(3), 307–320. https://doi.org/10.1080/13604813.2014.906716

Son, T. H., Weedon, Z., Yigitcanlar, T., Sanchez, T., Corchado, J. M., & Mehmood, R. (2023). Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. *Sustainable Cities and Society, 94*, 104562. https://doi.org/10.1016/j.scs.2023.104562

Sudmann, A. (Ed.). (2019). Deep Learning's Governmentality The Other Black Box. In *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms* (1st ed., Vol. 1). transcript Verlag. https://doi.org/10.14361/9783839447192

UN-Habitat. (2020). *Centering people in smart cities: A playbook for local and regional governments.* United Nations Human Settlements Programme. www.unhabitat.org

UN-Habitat. (2023). *Human rights in the digital era* (pp. 1–56). United Nations Human Settlements Programme. https://unhabitat.org/human-rights-in-the-digital-era-learnings-on-local-governance-from-pilots-in-europe

Willis, K. S. (2019). Whose Right to the Smart City? *The Right to the Smart City,* 27–41. https://doi.org/10.1108/978-1-78769-139-120191002

Yigitcanlar, T., Agdas, D., & Degirmenci, K. (2023). Artificial intelligence in local governments: Perceptions of city managers on prospects, constraints and choices. *AI and Society, 38*(3), 1135–1150. https://doi.org/10.1007/s00146-022-01450-x

Yigitcanlar, T., David, A., Li, W., Fookes, C., Bibri, S. E., & Ye, X. (2024). Unlocking Artificial Intelligence Adoption in Local Governments: Best Practice Lessons from Real-World Implementations. *Smart Cities, 7*(4), 1576–1625. https://doi.org/10.3390/smartcities7040064

Yigitcanlar, T., Desouza, K. C., Butler, L., & Roozkhosh, F. (2020). *Contributions and Risks of Artificial Intelligence (AI) in Building Smarter Cities: Insights from a Systematic Review of the Literature.*

Yigitcanlar, T., Kamruzzaman, Md., Foth, M., Sabatini-Marques, J., Da Costa, E., & Ioppolo, G. (2019). Can cities become smart without being sustainable? A systematic review of the literature. *Sustainable Cities and Society, 45*, 348–365. https://doi.org/10.1016/j.scs.2018.11.033

Yin, R. K. (2009). *Case study research: Design and methods* (K. deMarrais & S. D. Lapan, Eds.; 4. ed., [Nachdr.]). Sage.

What Employees Expect from AI: Characteristics and directionality within a plurality of AI expectations

Authors: Otto Hedenmo, Jönköping University, otto.hedenmo@ju.se
Maria Riveiro, Jönköping University, maria.riveiro@ju.se
Annika Engström, Jönköping University, annika.engstrom@ju.se
Nina Edh, Jönköping University, nina.edh@ju.se
Carla Gonçalves Machado, Jönköping University, carla.goncalvesmachado@ju.se
Daniel Pittino, Jönköping University, daniel.pittino@ju.se
DOI: https://doi.org/10.54337/aau.add.scai-11427

ABSTRACT

AI is viewed as the next major technological breakthrough for organizations. The range of areas, professions, and practices that can be improved with AI assistance or automation is overwhelming. However, this wide array of possibilities also brings a variety of expectations about how AI will change organizations and employees' everyday work. Considering that voiced expectations influence adoption processes by both reflecting and shaping certain relational, belief-driven dynamics, we can learn a great deal about AI adoption by studying the organizational plurality of AI expectations. Therefore, this study examined AI expectations held by employees in three organizations currently adopting AI for use in the workplace. The study is based on a thematic analysis of empirical material form 15 focus groups in three Swedish AI-adopting organizations and shows how AI expectations shape the following: (1) a growing desire to move from exploring AI to establishing AI routines and regulations, (2) emerging dilemmas related to both the violation and fulfillment of AI promises, and (3) how dynamics and unpredictability in the AI field require organizations to adapt to shifting trends and innovations.

Introduction

Few have missed the buzz surrounding AI, an umbrella term defined here as encompassing numerous data-driven, self-learning, automated, and digital technological innovations that have quickly become part of organizational processes and structures (Lindgren, 2024). The increasing use of AI is transforming organizations, affecting work processes from employee recruitment and product manufacturing (Jackson et al., 2024) to how communication flows coordinate and organize people (Gulbrandsen & Just, 2024). These technologies are being rapidly and widely adopted by organizations, requiring a diverse range of professionals to understand what to *expect* from these technologies. Inevitably, multiple understandings have emerged. While these offer opportunities for innovative approaches, they also carry risks—such as organizational disagreements, conflicts, and inertia—stemming from the absence of a common direction.

In technological adoption processes, voicing expectations allows organizations to make sense of the new technology (Engström et al., 2024 and to "develop and enact projects, plans and strategies"

(Beckert, 2016, p. 14) that also reflect and inform collective beliefs. Most likely, this is also true for *AI* adoption processes. Characteristics of AI expectations, therefore, bring insights into the social formations that AI, as a phenomenon, invites and enforces. In this study, we highlight the presence of AI expectations in organizations and explore how they inform current adoption processes in order to contribute to knowledge about these expectations and the organizational environments that need to address them.

Expectations here are defined as individual understandings shaped by social constructs regarding how the future will unfold. We use a broad definition that includes aspects ranging from professional estimations to hopes and fears about AI. Two elements in AI adoption processes make this broad definition necessary and the organizational management of AI expectations essential: 1) AI adoption is happening virtually everywhere in both private and professional spheres and involving multiple departments, professions, and stakeholders with diverse needs, ideas, and expectations; and 2) the technology is not new as a social construct and, therefore, is already connected to personal beliefs, experiences, and societal discourses about its potentials and dangers. As such, what people expect from AI is highly multifaceted and somewhat "fuzzy" (Willim, 2024, p. 57; see also Lomborg & Kapsch, 2020).

As an organizational change process, introducing AI can be viewed as non-linear, dynamic, relational, and co-created within a surrounding organizational context (Lewis, 2019). In other words, adopting new tools and technologies involves personal experiences, public dialogues, myths about other organizations' adoption processes, and more: all contribute to shaping expectations and informing how change processes evolve. The two elements noted consequently lessen organizations' control of AI adoption processes, placing organizations more in a position wherein knowledge about employees' evolving expectations becomes key to managing the introduction of AI. Indeed, if organizations are to coordinate and legitimize a particular AI transformation, they must adhere to the transboundary, cross-sectoral, and cross-professional nature of AI (van Lente, 2012). Therefore, knowledge about the variety and characteristics of AI expectations in organizations brings valuable insights for research on the dynamic topic of sociotechnical AI integration and for organizations actively maneuvering these complex processes.

Thus, this study was intended to deepen our understanding of the plurality and features of organizational AI expectations and how they influence AI adoption processes. We qualitatively examined AI expectations within three Swedish companies in the industry/technology sector, each in the process of adopting AI technologies. By thematically analyzing empirical material from 15 focus group discussions with company employees regarding their expectations of AI, we examined the nuances and relationships of voiced AI expectations and explored how their characteristics provide direction for organizational AI adoption processes.

Next, we review previous research on AI expectations and how it informed the current study. Then, a theoretical discussion on expectations is presented, followed by an outline of the investigation method. Finally, the results are presented, discussed, and concluded.

Previous studies on AI expectations

People's expectations about the future have long been a concern across scientific fields. Research on our thoughts about tomorrow and beyond ranges from local, action-oriented studies to broader societal discourses (Beckert, 2016). This includes ideas about AI.

At a microlevel, tech-focused studies, such as those based on expectation-confirmation theory (ECT) or the technology acceptance model (TAM), have examined how AI use is influenced by user expectations about "what will happen during and after the interaction [with AI]" (Grimes et al., 2021, p. 1). The goal has been to assess perceived AI performance in terms of usefulness and satisfaction (e.g., Li & Wang, 2025). Using this approach, they connect individuals' expectations of AI capabilities with how people practically use AI (Grimes et al., 2021), as well as with people's experiences and actions after engaging with AI (Jenssen et al., 2025; Li & Wang, 2025).

While these studies ontologically differ from our research, they demonstrate how AI use informs practice and connects to social processes. An example provided by Nguyen et al. (2025) showed how employees' expectations of AI are linked to social consequences, including whether people discuss their expectations and experiences with AI. Personal expectations and experiences, thus, generate relational consequences that inevitably influence others' expectations of AI. However, while these studies have illustrated that expectations socially matter, they lack qualitative understanding of the actual expectations they examined: that is, the "nuances" (Nguyen et al., 2025, p. 814) that differentiate people's expectations and how (and what) "socially accepted perceptions" (Jensen et al., 2025, p. 29) regarding motives, barriers, and context inform their expectations.

AI expectations can be seen as an umbrella term that covers various related theoretical areas, such as imaginaries (Christensen, 2025), dreams (Stevens et al., 2022), promises (Hirsch-Kreinsen, 2023), and anticipation (Hautala & Ahlqvist, 2024). In various ways, these studies have contributed to our knowledge about what people expect from AI. For example, sociotechnical imaginaries, the "collectively imagined forms of social life and social order" (Jasanoff & Kim, 2009, p. 120), have explored how broad societal ideas about AI link to social and political discussions that frame AI as a phenomenon (e.g., Bareis & Katzenbach, 2022). Additionally, they examined how public interest surrounding AI fosters ideas of technological optimism, societal progress, and competitive discourses regarding adoption of AI before others; they also examined how some ideas (e.g., AI-related security, responsibility, and ethics) are culturally specific and, therefore, dependent on context (Bareis & Katzenbach, 2022; Hirsch-Kreinsen, 2023).

These overarching ideas about AI have implications for organizations, including how AI strategies and action programs at various levels are intended to guide organizations in a specific direction (Kannelønning, 2023; Stevens et al., 2022). However, when general top-down expectations about AI intersect with employees' expectations, the adoption process becomes more complex, conflicted, and requires negotiation (Hautala & Ahlqvist, 2024; Lomborg & Kapsch, 2020). For example, while organizations' strategic expectations of AI tend to focus on progress and efficiency, employees' expectations include a wide range of anticipations, such as hopes for decolonial processes (Seto, 2025) and affectionate 'dreams' of becoming better professionals (Stevens et al., 2022). Prior research has indicated the multifaceted nature of AI expectations and how these perceptions connect not only to 'desired' futures (Borup et al., 2006) but also to unwanted changes and 'nightmares' to avoid (Stevens et al., 2022).

Given that the variety of employees' AI expectations tends to be complex and evolving and includes ideas about hope, ethics, and concerns (Corsini et al., 2025), the constitution of AI expectations raises questions that the organizational AI adoption process must seriously address—such as regarding work ethics and corporate responsibility. However, while many organizations establish mechanisms through which employees can critique AI, studies indicate their ability to influence the process remains limited (e.g., Kannelønning, 2023), leaving employees feeling they have "no voice in the process" (Seto, 2025, p. 13).

Our reading of the literature on AI expectations revealed how employees' AI expectations are both more complex than what the dominating number of tech-oriented studies address and more reflective than what is usually covered in broad and over-the-top aspirational strategies to nurture the potential from AI adoption. Employees' expectations are reflected not only in technical skills but also in deeper professional and cultural values about how and why AI should be used. Our aim was, therefore, to contribute to a more profound understanding of the plurality of AI expectations that exist within the organizational context and how they inform certain social formations among employees. Understanding the organizational variation in AI expectations enabled us to illuminate certain alignments of ideas, but also friction and conflicts, that underpin today's organizational AI adoption processes.

How we understand AI expectations

As mentioned, we recognize expectations broadly and as socially shaped individual understandings regarding how the future will unfold. As Beckert (2016) asserted, expectations are "interpretive frames that structure situations through imaginaries of future states of the world and of causal relations" (p. 9). To theoretically position expectations in the organizational realm, we referenced Weick's (1995) idea of expectations as "starting points" (p. 148) for organizations' belief-driven processes. While expectations always exist in organizing, Weick maintained they flourish during periods of instability, change, or "unfinished business" (see also van Lente, 2012). Voiced expectations, whether expressing desire and longing or risks and nightmares, then become starting points for developing "orderly interaction around *expecting*," which, by becoming meaningful, creates a "potent force in their own validation" (Weick, 1995, p. 134). Through employees' interpersonal relations, organizations develop dominating expectations on a particular matter that inform certain organizational doings.

Since the future is illusory for everyone, the concept of expectations closely relates to perceptions of uncertainty (Beckert, 2016). Expectations can be considered understandings of the future imagined more certain, or plausible, than others. Coordinating expectations, therefore, essentially reflects a collective method aimed at reducing perceptions of uncertainty. However, while expressed expectations can appear far-fetched, irregular, and chaotic (e.g., Willim, 2024), they are inevitably directed toward specific organizational or contextual elements. This *directionality* of uncertainties (van Lente, 2012, p. 774) is important to understanding how expectations are grounded in a potential future and to characterizing AI expectations and discussing their influence for organizational formation.

Previous studies have theorized different forms of such directionalities. In developing our framework, we take inspiration from the conceptualization of uncertainties in change processes from

Bordia, Hobman et al., (2004), see also Bordia, Hunt et al. (2004). These scholars argue that employees primarily direct their uncertainties through three approaches: *strategic* ("reasons for change, planning and future direction of the organization"), *structural* (changes to such aspects as "reporting structures and functions of different work-units"), and *job-related* ("job security, promotion opportunities, changes to the job role, and so forth") (Bordia, Hobman et al., 2004, pp. 510–511). However, while these categories focus on the employee–organization relationship, change processes are not isolated but involve elements that are somewhat out of the organization's control (Lewis, 2014, 2019), particularly regarding transboundary questions, such as those related to AI (see Lomborg & Kapsch, 2020). Informed by the literature review, we therefore propose that to grasp AI expectations within an organization, a *contextual* direction of uncertainties (toward changes outside the organization) must be added.

Nevertheless, while the strategic, structural, job-related, and contextual categories characterize the direction of expectations, they give little guidance on organizational consequences. As expectations are voiced, developed, and empowered in interpersonal interactions, expectations scale up and move through organizations in ways that change organizing. The directionality of expectations can indicate where these changes may manifest. Processes related to how expectations impact and evolve on a societal dimension are theorized in research on the sociology of expectations (Beckert, 2016; van Lente, 2012). We take inspiration from those ideas to explore the movement of expectations within the organizational realm. Specifically, we base our approach on Borup et al. (2006), who theorized how expectations of technological change influence social formations in vertical, horizontal, and temporal ways.

Here we used the vertical, horizontal, and temporal dimensions as a framework to explore and discuss how the directionality of AI expectations carry potential consequences for organizations. *Vertically*, expectations mediate different scales of an organization. For example, expectations expressed in small talk can find ways of becoming or opposing superior and overarching expectations, and vice versa. Expectations thereby vertically align and coordinate with (or transform and scatter) a particular direction. (For an excellent empirical example, see Christensen, 2025.) *Horizontally*, expectations perform by resonating, or not resonating, across organizational boundaries, such as departments, professions, and other collectives, albeit through different ways and consequences of performing in each grouping. *Temporality* reflects how the 'potent force' of expectations evolves over time from "a new set of organized cues that have become meaningful" (Weick, 1995, p. 147), whereby new conditions and problems necessitate adaptation.

By understanding AI expectations as socially constructed, this study differs from dominant rationalistic perspectives that treat expectations as numeric scales from true/false or rational/irrational (e.g., TAM, ECT) and, instead, focuses on exploring nuances, alignments, and contradictions of AI expectations. By understanding expectations as social, we look beyond the "accuracy" of expectations (Buschmeyer et al., 2022, p. 447) and focus on their constitutive functionality in organizational change practices. Through this framework, we contribute with a novel understanding of how AI expectations shape organizations through their character and direction.

Method of analysis

This study was developed from an interactive research approach in which the research purpose was created and the collection of empirical material conducted through collaboration between academic and industry professionals. Interactive research is "characterized by recurrent interactions and joint learning activities between researchers and practitioners in commonly agreed upon efforts to study change and innovation in organizations" (Ellström et al., 2020, p. 1520). The project to which this study adhered investigated the managerial processes of integrating AI in organizations. To gain knowledge in this context, the project group (comprising industry partners and academic researchers) had been working together for five years to actively identifying methods for how to incorporate AI into their organizations. In this work, the project group identified the nuances and directions of AI expectations as a relevant phenomenon to investigate and contribute to research.

As we view expectations as ideas of the future, a qualitative method (Howitt, 2019) was necessary to gain descriptions of work–life conditions and hopes and concerns related to the AI adoption process and AI as a phenomenon. The empirical material comprised transcripts from 15 focus group sessions conducted across the three partner companies, five sessions at each. Focus groups are valuable for exploring phenomena wherein people possess a variety of understandings (Gustafsson, 2014) while allowing participants to introduce, reflect on, and challenge each other's expectations. As Weick (1995) argued, as employees "dwell on what might happen, people's expectations become better articulated" (p. 134). The focus group sessions began with a researcher facilitating a company-specific focus group and then training company representatives to lead their own groups. In this way, 12 of the 15 sessions were conducted by the companies themselves. The research team had positive experiences with this setup in previous work and had conducted similar focus groups, in terms of companies and questions, five years earlier (Engström et al., 2024, 2025). Participating companies, located in Sweden, is presented in Table 1.

Table 1: The companies that participated in the collaborative, interactive research project.

	Company A	Company B	Company C
Year of establishment	1945	1977	1992
Product	lighting product manufacturer	C-parts manufacturer/ supplier	technology consultancy service
Revenue Thousands Euro	124,000 (2024)	465,254 (2020)	386,848 (2024)
Number of employees in Sweden	~500 employees	~1,300 employees	~2,100 employees

White-collar employees from various departments and professions, including analytics, human resources, engineering, information technology, product development, communications, and others, participated in the focus group discussions, which did not diverge much company-wise, probably due to participants' office-centered work. The most obvious divergences were the company-specific examples raised by participants (e.g., lighting solutions or C-parts). All sessions were conducted in Swedish, lasted about 60 minutes, involved around 60 employees, and were recorded and transcribed verbatim.

The focus groups were organized around six questions: (1) What is your spontaneous reaction when you hear the concept of AI? (2) What do you think AI is in practice? (3) In what applications can AI be applied within your organization? (4) What can AI be used for near your own work area? (5) What consequences can AI have for your organization/work? (6) How can your organization prepare itself and you in the best way for the application of AI? While none of the questions explicitly mentioned "expectations," they were designed to encourage participants to share their ideas about the future with AI by directing discussions toward what could happen. The discussions started broadly and then focused increasingly on work-related aspects. This approach generated rich insights, with participants offering a range of ideas about how AI might influence strategic, structural, job-related, and contextual elements of the organization.

Analysis

To explore the diversity, character, and relationships of AI expectations, we used thematic analysis, a simple yet "foundational method for qualitative analysis" (Braun & Clarke, 2006, p. 78). This approach gave us the flexibility to interpret the material while staying grounded in our theoretical framework. Our analysis started with familiarization with the material and then initial coding, followed by systematically developing categories and themes through ongoing review, refinement, and questioning.

We approached the material in an open and exploratory way (Saldaña, 2015, p. 213) by coding excerpts in which participants shared ideas about how AI will influence the future. Codes were described as, for example, "AI as administrative support" or "AI as efficiency." Throughout the process, we continuously revised the codes, checking whether similar codes already existed. The goal was, however, to keep the labels as nuanced as possible. For example, codes labeled "AI ambivalence" included keywords (e.g., "efficiency/replacement") to capture the subtleties of ambiguity in the quotations. In total, 467 codes were created.

Next, we colored the codes by interpreting the relation of the expectations of AI to the four directions mentioned (Bordia, Hobman et al., 2004; Lewis, 2019):

- Strategic: Related to future directions for the organization
- Structural: Related to structural changes for the organization
- Job-related: Related to changes to the job role
- Contextual: Related to change outside the organization

The color coding provided us with an overview of the codes' organizational location and direction. For example, it showed that job-related and contextual expectations were dominant, while strategic and structural codes were less prominent.

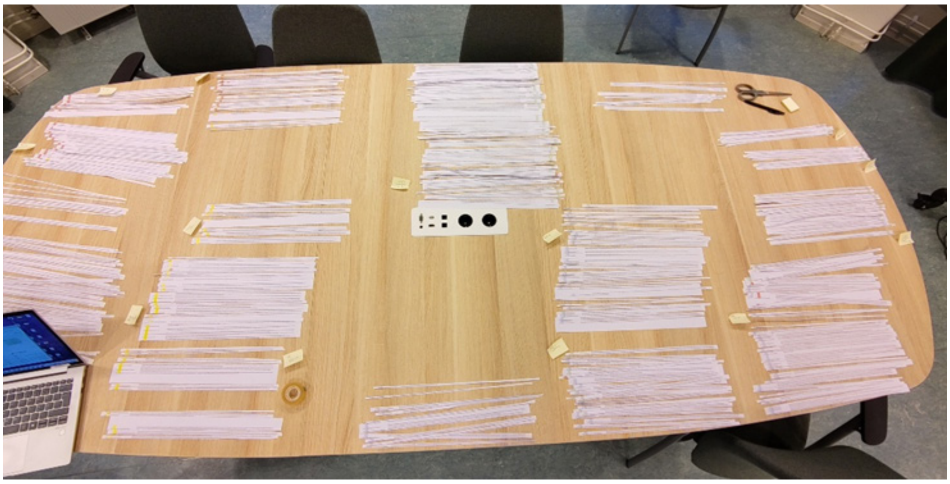


Figure 1: Grouping of codes

Next, we grouped the codes based on how they reflected certain hopes, fears, and ideas about what to expect from AI (see Figure 1). We developed the groups inductively, resulting in labels such as "tech-optimism," "AI as admin/text support," and "AI needs a plan." The color coding visually highlighted how some expectations clearly related to particular directions, while others were more scattered.

In the final step, we identified four themes of statements that "expand[ed] on the major ideas" (Saldaña, 2015, p. 260) about AI expectations. We focused on understanding specific relationships by examining how the groups aligned with, contradicted, or simply coexisted with each other within the material. See Table 2 for the themes, groups, and directions that dominated.

Table 2: Focus group analysis: themes (top row), groups (below themes), and dominating directions (left column)

	Echoes and conflicts of tech optimism	**Expectations of strategy**	**Perceived necessity for structured AI readiness**	**Hopes of leaving modern organizing**
Strategic		• AI as strategic risk • AI needs planning		
Structural	• AI as structural risk • AI for prognosing future		• AI demands a learning organization	• AI as auditor/reviewer
Job-related	• AI for analysis • Improved customer relationships		• AI as competence/job security	• AI as admin/text support
Contextual	• Tech optimism • AI critique • AI ambivalence	• AI as inevitable		

In the following sections, we present the four themes and discuss them in relation to the vertical, horizontal, and temporal dimensions introduced previously. All excepts were translated into English by the researchers.

Themes of AI expectations

In this section, we introduce our four themes: "echoes and conflicts of tech optimism," "expectations of strategy," "perceived necessity for structured AI readiness," and "hopes of leaving modern organizing." The themes are arranged from the most overarching, general expectations to the most specific and closest to work, with quotations from the focus group participants included to support our analysis.

While this investigation focused on employees' expectations about AI, a variety of understandings about "what AI *is*" emerged. Some participants conveyed a conceptually open attitude toward AI by inquiring, for example, as to what AI application was being discussed, but most equated the term with predictive or generative AI (i.e., large language models [LLMs], such as ChatGPT or Copilot). One reason these types of AI were prominent was the text-intensive nature of the participants' professions. Regardless of whether they were using AI for coding, benchmarking, or recruiting new employees, most were using it to develop or process text. We interpret this dominating understanding of AI as an imaginary focused on envisioning the text-assisting and bureaucratic potentials of AI, encouraged by the tsunami of LLMs washing over organizations and individuals in the last three years.

Echoes and conflicts of tech optimism

The focus group discussions typically began with participants sharing broad ideas about AI, expressing optimism, criticism, and feelings of ambivalence. The predefined questions prompted the groups to begin with wider questions about "societal impact," which helped guide the flow of the conversations. Our analysis, therefore, easily indicated how overarching ideas about AI, coexisting within and across the focus groups, created a delicate balance between views of AI as introducing potential improvements and possible disasters.

Optimistic ideas about AI shared similarities with national and tech-enthusiastic discourses that highlight AI's potential (Bareis & Katzenbach, 2022). Participants echoed these common ideas by connecting AI to concepts like efficiency, productivity, and societal progress. Common in this regard was viewing AI as limitless. One Company A participant explained: "Well, it is actually just our fantasies that set the limit. If the challenge is processing the data, then there is no limit. The development is that fast." Comments like these, which stress AI as unlimited, inevitable, and a force of nature, reflect optimistic and unquestioning beliefs linking AI to overall improvements, both quantitative (e.g., optimization) and qualitative (e.g., error reduction). Expectations from accepting tech enthusiasts considered AI as just another tool while omitting details about how, why, and with what consequences.

Hopes of reducing errors supported AI as another tool to improve work tasks and replace other tools. AI-driven improvements to reduce flaws were often linked to specific tasks, such as deepening customer knowledge or developing analytical processes, but also framed AI as a solution to myriad problems. As one Company A participant stated, "I think it [AI] could also ensure quality similar to human intelligence. However, it doesn't experience fatigue or other human factors, so it can both maintain and improve quality while increasing efficiency." This excerpt exemplifies tech-optimistic reasoning in which replacing human intelligence can lead to increased quality and efficiency. This can be interpreted as part of a solutionist reasoning (Lindgren, 2024, pp. 61–62) in which arguments for an AI solution omit questions of social complexity while focusing on benefits such as "quality" and "efficiency." Expressions like these echoed progressive ideas commonly found in AI strategies and action programs (Bareis & Katzenbach, 2022).

Conversely, participants also expressed fears and anxieties when criticizing how AI could change society. These concerns included various perceptions of AI weaknesses, such as unreliability, misuse, replacing people, or a general sense of fear.

> *I feel a certain fear and worry for what it [AI] can mean for ... our society, for us. I understand that there might be a lot of upsides to it, but it feels like—at least for my part—I am not informed enough to say anything. But … I feel fear about what it could mean to us. If that is because I have watched too many movies or read too much, I don't know, … but I can feel more worried than joyful about it. (Company A Participant)*

Through analysis, we observed how AI-optimistic discourses coexisted with these concerns about how the technology will impact the environment outside the organization. Although they recognized the benefits (the "upsides to it,") AI as a social phenomenon was perceived as unreliable and erratic. However, while the unreflective nature of tech optimism was built on rational ideas and specifics, critical voices emphasized vague feelings of "fear" and "worry," highlighting an underlying uncertainty that follows the current AI trend in which expectations of AI are based not only on rational beliefs but also on myths, feelings, and hunches.

Amid both optimistic and unfavorable views, employees also expressed instant ambivalence by balancing expectations for improvement with fears of disaster. "I believe it's a huge opportunity but also a small threat, if we are to be honest" (Company A Participant). This AI ambivalence was rarely thoughtless or unclear; instead, it emerged from specific expressions that raised opportunities and concerns. As such, comments about AI ambivalence involved perceived positives *and* negatives, such as efficiency versus stupefying, increased human contact versus uncontrolled data sharing, and simplification versus security threat. Together, these expressions reflected a larger theme of mixed expectations, whereby AI adoption was seen as not entirely good or bad but, instead, as uncertain and in need of a cautious, balanced approach.

> *Both positive and negative: that's my spontaneous reaction today. It's cool to see what AI can do and cool to … experience … all the improvements that happen within the company. But I also believe it is a bit scary when it comes to replacing people within … some areas and to what extent. (Company B Participant)*

Coexisting expectations that reflect optimism, critique, and ambivalence illustrate how employees' views on AI's societal and contextual impact are not fixed to a certain imaginary but, as Stevens et al. (2022) also emphasized, open to negotiation and change. By showing how tech optimism coexisted with employees questioning the reasons and methods for adopting AI, mixes of unreflective certainty, perceptions of fear, and ambivalence reflected the overall notion of unpredictability regarding what will change when adopting AI.

Expectations of strategy

While the previous theme highlighted how employees perceived unpredictability and uncertainty about how AI will change the organizational context, uncertainty was also prominent in the theme "expectations of strategy" but related to expectations about how to organizationally navigate this uncharted territory. The theme "expectations of strategy" underscores a dominating expectation on leadership and management to make strategic decisions, especially to avoid unwanted risks.

Although optimism, critique, and ambivalence about AI coexisted among the focus group participants, the decision of whether to adopt AI was never up for debate. AI was perceived as an unavoidable innovation to which the organization needed to adapt. As a Company C participant said, "We need to have mature organizations for this [AI], but we also need to question our own maturity for it.… We cannot choose not to.… We need to be on the wagon or else it [AI adoption] won't work."

As an organizational response, employees directed their expectations toward management for guidance. Participants conveyed expectations for the long-term vision of the organizational AI initiative ("Where we are in three years and where we are in … whenever that might be"), as well as for help in minimizing unwanted risks among staff.

> *I mean, it is not dangerous to say, "yes" [to] Chat-GPT [and] "how to use it." There must be these … like, I am thinking, from the company … how to use it and for what parts. What is fine and what is absolutely not fine to do. (Company B Participant)*

We interpreted that employees hoped to gain relief from general unpredictability and, sometimes, fears about misusing AI from management strategies. Employees' discomfort with improper use did not seem linked to any past misuse within the companies but was, instead, tied to myths, media stories, and anecdotes about AI that employees had heard elsewhere, such as concerns about data sharing, cheating, and fake content. The perceived need for guidance highlights challenges pertaining to AI being adopted at work, at home, and in life simultaneously. Here, experiences, myths, and examples from the professional and personal realms contributed to making expectations of AI less clear and more speculative.

The perceived need for strategy and direction was interpreted as an expectation, or hope, to reduce uncertainties, harness potential, and manage risks. However, along with these strategic expectations, structural expectations also emerged about how the organization would reduce the uncertainties.

The expected necessity for structured AI readiness

The idea that AI (or technology in general) will "take our jobs" is common in the history of technological change. Predictably, this idea was raised during the focus group discussions. However, the participants' lack of concern about this was surprising. Participants considered the loss of jobs to be more likely to affect others than themselves.

> *Some are like "yeah, AI will take your job." I don't really believe that, but I think we will have to change many work tasks, or parts of work tasks, and optimize many processes. But what I do think is that if you are opposing it and definitely do not want to learn anything, completely reluctant, then I think there is a risk that you get replaced with someone who actually gets it. (Company C Participant)*

This excerpt illustrates how the need for AI skills among employees is expected to grow, while the fear of being replaced was essentially absent. This diverges from previous studies that suggest fear of job loss was a central theme in AI adoption (e.g., Willcocks, 2020). Our focus group participants did not view AI as a threat, instead emphasizing the need for humans to ensure its proper functioning. This skepticism toward AI has been highlighted elsewhere (e.g. Sondern et al., 2025). However, prominent in our material was skepticism about AI's ability to handle social qualities (e.g., human-to-human interactions), as well as a desire for someone to manage the AI.

> *What kind of AI is it? Who will own the AI? Who will train the AI? Who will understand the way AI makes decisions? These are also competencies you need to secure in one way or another. If you dare to outsource everything, hand out all information, or to have the competence in-house, these questions you need to address, too. (Company A Participant)*

Employees expressed an expectation that developing the workforce in an AI-ready manner was inevitable but questioned who would prepare the organization. While similar studies suggest learning is expected at the team level (Hautala & Ahlqvist, 2024), our analysis indicates that employees directed the expectation to learn toward the organizational structure. That is, to be AI-ready, opportunities to learn AI need to be embedded in organizational routines and workflows.

> *We need to build knowledge here and now to understand this [AI] and get very … clear leadership about how to get … the human to keep up and wanting to keep up in this transformation, instead of forcing them into something. (Company B Participant)*

The expectation was, thus, not a perceived need to "go back to school" but, rather, that the organization would, out of necessity, provide opportunities for employees to learn how to adopt AI solutions or, as a participant from Company A phrased it, to "educate" the workforce about AI.

> *I just wanted to say that education, a basic education, about what it [AI] is, what clear pitfalls or clear … positive stuff we can use from it, is really, really basic, so that everyone has some kind of basic level [knowledge] about it. I believe that is a really great start.*

Therefore, structures were expected to educate the workforce on how to use AI effectively while also ensuring competent human oversight of AI systems. Although our analysis uncovered employees' awareness that they need to learn, any specific expectations about *what* they had to learn remained vague (see also Engström et al., 2025). Participants described AI learning initiatives as similar to "some kind of education or knowledge about it. It doesn't need to be that detailed or in-depth, but to understand how it works in general" (Company A Participant). Again, the uncertainty of what to expect from AI rose to the surface. In the background of the dominating expectation that "AI will change everything," unpredictability about what AI will actually bring to the organization was percolating. However, employees became more specific when discussions shifted to how AI might change their daily work practices.

Hopes of leaving modern organizing

Overall, the fourth theme, "hopes of leaving modern organizing," connects expectations that were closely related to employees' work tasks; consequently, detailed reflections about participants' immediate work environment dominated. Due to participants' diverse professional backgrounds, the work tasks discussed varied. However, two overriding expectations emerged across the groups: hopes for administrative relief and structured auditing. Together, these expectations formed a theme that we interpreted as reflecting hopes of moving away from work tasks associated with modern organizing, where reporting, documentation, and auditing are central.

Participants viewed AI as a tool that could support text processing and help manage administrative tasks, mostly associated with various LLMs. These expectations involved specific activities, such as summarizing meeting notes, creating PowerPoint slides, and reporting work hours. For the participants, this administrative relief was expected to give time to focus on human-to-human tasks, which were viewed as more creative, qualitative, and enjoyable.

> *Sometimes there's a lot of repetitive, and just a lot of handling with old systems. ... [A] dream scenario might be to have a small "AI buddy" to keep you company while working. It knows everything you know, too, or everything you have written about, and it could handle all these boring tasks because it works really fast, giving you more time for what is critical. I think that would make it a bit more fun, too, ... in some way. (Company C Participant)*

This quote suggests that AI with efficiency and autonomy could handle administrative tasks to clear time for more important creative and high-quality work. Essentially, these expectations reflected an everyday work environment that employees did not consider fulfilling or meaningful. Administrative tasks were seen as what prevented them from making work more valuable. "The benefits [with AI] come in the form of increased efficiency and maybe an opportunity to make time for something else, and more valuable" (Company A Participant).

Employees argued that "death tasks" or "boring tasks" dominated their workdays. The urge to make work more meaningful is well-supported in organizational research and commonly connected to aversions toward rationalistic ideas of modern organizing (e.g., Alvesson, 2019; Bornemark, 2018). However, in our analysis, these aversions did not manifest in ideas of removing these tasks but in allocating them to AI. In participants' reflections, this involved both developing text and auditing their own work.

> *Yes, but I believe that AI could also act as some kind of "gatekeeper" for a quality label. To understand that—okay, now we are approaching the launch of a product—is this product ready? Do you have all the numbers in the right places? Do we have our data files in the right place? Does it have all the pictures? It [AI] can just validate that everything is okay for launch. (Company A Participant)*

Essentially, regardless of whether AI expectations referred to an AI buddy or a background validator, the hope was for AI to ensure that operations were running according to plan. What manifested in the analysis was the idea of AI as a framework for maintaining the organization to function as it already does while making daily work practices more meaningful.

These expectations for identifying approaches to make organizing more meaningful have been highlighted in AI studies before (Hautala & Heino, 2023; Stevens et al., 2022). The hopes for relief from administrative duties can reflect a deeper desire to escape certain aspects of modern work, evoked by AI's affordances to assist or audit work practices—such as reviewing data, ensuring standards are met, and maintaining quality. However, the ways AI affordances would provide a more meaningful work environment were vague and primarily linked to time efficiency. This was further emphasized in the unclear directionality of AI as administrative relief, where we observed a lack of clarity regarding whether employees would introduce the relief themselves or if that relief was supposed to result from a structural change within the organization.

Next, we elaborate on what the four themes reveal about AI adoption processes.

Discussion: Expectations and directionality in AI adoption processes

In the introduction, we argued that employee expectations reflect, on the one hand, the current state of their organizations and how closely they are intertwined with AI. On the other hand, the themes also allowed us to interpret and discuss how AI expectations shape and enforce certain organizational formations. By departing from the vertical, horizontal, and temporal dimensions presented by Borup et al. (2006), in this section, we explore what our four themes say about AI adoption in organizations based on the directionality of actions that AI expectations invite or enforce. We present three main insights that are key to characterizing today's AI adoption processes: (1) transitions toward regulation, (2) conflicts over meaningfulness, and (3) dynamics and unpredictability.

Vertically, the analysis indicated employees' AI expectations were directed toward specific organizational levels, mainly upward and toward management, for instance, in expecting managerial AI strategies and structured AI learning initiatives. These findings suggest a shift in organizational adoption processes from an acceptance of the chaotic, playful, and exploratory mindset that previously shaped these processes (e.g., Engström et al., 2024) toward expectations that focus more on strategizing, education, and regulation. The findings also suggest that the expectation for regulations is linked to an overall perception of uncertainty, reflected in employees' vagueness and lack of clarity about what to explicitly strategize and learn regarding AI. This uncertainty was reflected in the many myths, hearsay, and anecdotes about AI misuse that employees mentioned. In our analysis, we connect this uncertainty–stability relationship to Weick's (1995) argument that in times of uncertainty, "what people need is some form of stability" (p. 153) and that perceptions of stability are shaped by the myths we believe in (McPhee & Zaug, 2001) and the routines we follow (March, 1981). Thus, our analysis contributes to showing how employees today seek AI stability by looking upward with expectations for directives and structures to balance mythology with routines and regulations.

In the horizontal dimension, hopes for changing work practices dominated, manifesting in the desire to reduce textual administration and implement automated auditing processes. Employees expressed that their workdays comprised parallel situations, with one side of work, such as completing administrative tasks like documenting and calculating, considered repetitive and meaningless, while the other side was viewed as meaningful and consisting of core practices, that is, working creatively and human-to-human (also discussed by Bornemark, 2018). The hope for AI to amputate meaningless functions and make work life more substantive is logical and understandable. However, employees struggled to explain *what* more consequential work tasks would entail. An overall uncertainty regarding what AI would bring also lingered in employees' expectations about work life, making expectations about what to move away from considerably more specific than expectations about what to move toward. This ambiguity becomes paradoxical for companies adopting AI and puts them in a certain dilemma: to either violate hopes of meaningfulness through strategic decisions, technological limitations, or inadequate implementation *or* to succeed in meeting expectations, only to find that meaningfulness is unevenly distributed and that employees struggle to identify activities that make work more purposeful. Regardless of what path organizations end up following, this dilemma carries risks of resentment and criticism within the workforce.

Certain temporal characteristics distinguish the research period and reflect the dynamics and unpredictability of organizational AI adoption processes. As mentioned, the research team previously conducted similar focus groups five years earlier involving the same questions and companies (Engström et al., 2024, 2025). Comparing the two periods, we observed a notable increase in more vivid expectations related to administrative affordances and less toward the overarching societal consequences of AI. This change may be linked to the recent widespread adoption of LLMs, such as ChatGPT, Copilot, and Gemini, which have introduced new work expectations that were less prominent previously. Similar fluctuations of AI expectations have been shown to exist at societal levels for decades (Hirsch-Kreinsen, 2023). Through this study, we hope to bring insight into how these temporal dynamics, flows, and unpredictabilities shape AI innovations at the organizational level, prompting companies to adapt accordingly.

Conclusion

By exploring AI expectations among employees in three organizations within the industry/technology sector, we identified four themes that highlight the current diversity of expectations in AI-adopting organizations and how these expectations link to specific directions that socially inform organizations. Our analysis shows how these directions shape (1) a growing desire to move from exploring AI to establishing routines and regulations, (2) emerging conflicts related to the violation and fulfillment of AI promises about meaningfulness, and (3) how the dynamics and unpredictability in AI development require organizations to adapt to shifting trends and innovations.

Based on an interactive research approach, this study provides insights into questions perceived as important for AI-adopting organizations and, consequently, adds value to their adoption processes. Furthermore, by developing a framework that links ideas from organizational and social sciences to explore the presence, relationality, plurality, and dynamics in the AI expectations phenomenon, the study also theoretically and practically contributes to sociotechnical AI research. We hope this spurs further studies on the constantly changing nature of AI expectations.

Limitations

This study's main limitation is the lack of firsthand experience with the empirical material. The researchers had access to the material only through anonymized transcriptions, which meant they lacked insight into who said what in the focus groups. Accordingly, we lacked insights into how workers' expectations were reflected in management or vice versa. Additionally, the material was not analyzed in a longitudinal way, preventing any deeper analysis of how expectations develop over time. Both limitations offer valuable directions for future research.

Acknowledgements

The authors acknowledge the Knowledge Foundation, Jönköping University, and the industrial partners for funding the research and education environment for Knowledge Intensive Product Realization SPARK at Jönköping University, Sweden (Project: AFAIR; No: 20200223). Furthermore, the authors also wish to acknowledge the constructive and fruitful comments given by the two reviewers.

REFERENCES

Alvesson, M. (2019). *Extra allt!: när samhälls- och människoförbättrandet slår tillbaka*. Fri Tanke.

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, *47*(5), 855–881. DOI:10.1177/01622439211030007

Beckert, J. (2016). *Imagined futures: Fictional expectations and capitalist dynamics*. Harvard University Press.

Bordia, P., Hobman, E., Jones, E., Gallios, C., & Callan, V. (2004). Uncertainty during organizational change: Types, consequences, and management strategies. *Journal of Business and Psychology*, *18*(4). https://doi.org/10.1023/B:JOBU.0000028449.99127.f7

Bordia, P., Hunt, E., Paulsen, N., Tourish, D., & DiFonzo, N. (2004). Uncertainty during organizational change: Is it all about control? *European Journal of Work and Organizational Psychology*, *13*(3), 345–365. https://doi.org/10.1080/13594320444000128

Bornemark, J. (2018). *Det omätbaras renässans: en uppgörelse med pedanternas världsherravälde*. Volante

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, *18*(3–4), 285–298. DOI: 10.1080/09537320600777002

Buschmeyer, K., Hatfield, S., Heine, I., Jahn, S. & Markus, A. L. (2022) Expectation management in AI implementation projects: a case study. *EuroMed Journal of Business*, Vol. 18 No. 3, pp. 441–451. DOI 10.1108/EMJB-10-2021-0161

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Christensen, E. (2025). Organizing expectations: The co-constitutive dynamics of AI imaginaries. *Human Communication Research*, *51*(4), 215–227. https://doi.org/10.1093/hcr/hqaf014

Corsini, F., Annesi, N., & Frey, M. (2025). The role of AI in museums' journey towards sustainable development: Socio-technical imaginaries of a cultural and organizational transformation. *Technovation*, *146*, Article 103280. https://doi.org/10.1016/j.technovation.2025.103280

Ellström, P.-E., Elg, M., Wallo, A., Berglund, M., & Kock, H. (2020). Interactive research: Concepts, contributions and challenges. *Journal of Manufacturing Technology Management*, *31*(8), 1517–1537. DOI 10.1108/JMTM-09-2018-0304

Engström, A., Pittino, D., Mohlin, A., Edh, N., & Johansson, A. (2025). A paradox perspective on early AI adoption: Understanding temporal and relational tensions. *Journal of Organizational Change Management*, *38*(7), 1145–1171. DOI:10.1108/JOCM-02-2025-0086

Engström, A., Pittino, D., Mohlin, A., Johansson, A., & Edh Mirzaei, N. (2024). Artificial intelligence and work transformations: Integrating sensemaking and workplace learning perspectives. *Information Technology & People*, *37*(7), 2441–2461. DOI 10.1108/ITP-01-2023-0048

Grimes, M. G., Schuetzler, R. M., & Giboney, J. C. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, Article 113515. https://doi.org/10.1016/j.dss.2021.113515

Gulbrandsen, I. T., & Just, S. N. (2024). Artificial intelligence in organizational communication: Challenges, opportunities, and implications. In M. N. Ndlela (Ed.), *Organizational communication in the digital era: Examining the impact of AI, chatbots, and Covid-19*. Springer International Publishing AG.

Gustafsson, N. (2014). Traditionella och virtuella fokusgrupper [Traditional and virtual focus groups]. In J. Eksell & Å. Thelander (Eds.), *Kvalitativa metoder i strategisk kommunikation*. Studentlitteratur.

Hautala, J., & Ahlqvist, T. (2024). Integrating futures imaginaries, expectations and anticipatory practices: Practitioners of artificial intelligence between now and future. *Technology Analysis & Strategic Management*, *36*(9), 2100–2112. DOI: 10.1080/09537325.2022.2130041

Hautala, J. & Heino, H., (2023. Spectrum of AI futures imaginaries by AI practitioners in Finland and Singapore: The unimagined speed of AI progress. *Futures*, 153 (2023), 1–11. https://doi.org/10.1016/j.futures.2023.103247

Hirsch-Kreinsen, H. (2023). Artificial intelligence: A "promising technology." *AI & Society*, *39*, 1641–1652. https://doi.org/10.1007/s00146-023-01629-w

Howitt, D. (2019). *Introduction to qualitative research methods in psychology: Putting theory into practice*. Pearson.

Jackson, I., Ivanov, D., Dolgui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management: A capability-based framework for analysis and implementation. *International Journal of Production Research*, *62*(17), 6120–6145. DOI: 10.1080/00207543.2024.2309309

Jasanoff, S., & Kim, S-H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear regulation in the U.S. and South Korea. *Minerva*, *47*(2), 119–146. https://www.jstor.org/stable/41821489

Jensen, V. V., Alami, A., Rysholt Bruun, A., & Stouby Persson, J. (2025). Managing expectations towards AI tools for software development: A multiple-case study. *Information Systems and e-Business Management*, *23*, 869–901. https://doi.org/10.1007/s10257-025-00704-7

Kannelønning, M. S. (2023). Contesting futures of artificial intelligence (AI) in healthcare: Formal expectations meet informal anticipations. *Technology Analysis & Strategic Management*, *36*(11), 3845–3856. DOI: 10.1080/09537325.2023.2226243

Lewis, L. K. (2014). Organizational change and innovation. In L. L. Putnam & D. K. Mumby (Eds.), *The SAGE handbook of organizational communication: Advances in theory, research, and methods.* SAGE.

Lewis, L. K. (2019). *Organizational change: Creating change through strategic communication*. John Wiley & Sons.

Li, F., & Wang, H. (2025). Unveiling the mechanics of AI adoption in journalism: A multi-factorial exploration of expectation confirmation, knowledge management, and sustainable use. *Journalism and Media*, *6*(2), 65. https://doi.org/10.3390/journalmedia6020065

Lindgren, S. (2024). *Critical theory of AI*. Polity Press.

Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society*, *42*(5), 745–761. DOI: 10.1177/0163443719855301

March, J. G. (1981). Footnotes to organizational change. *Administrative Science Quarterly*, *26*(4), 563–577. https://www.jstor.org/stable/2392340?seq=1&cid=pdfreference#references_tab_contents

McPhee, R. D., & Zaug, P. (2001). Organizational theory, organizational communication, organizational knowledge, and problematic integration. *Journal of Communication*, *51*(3), 574–591. https://doi.org/10.1111/j.1460-2466.2001.tb02897.x

Nguyen, M., Mehrotra, A., Malik, A., & Pandey, R. (2025). Nexus between generative AI engagement, quality and expectation formation: An application of expectation–confirmation theory. *Journal of Enterprise Information Management*, *38*(3), 798–820. https://doi.org/10.1108/JEIM-08-2024-0452

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage Publications.

Seto, K. S. (2025). AI from the South: Artificial intelligence in Latin America through the sociotechnical imaginaries of Brazilian tech workers. *Globalizations*. DOI:10.1080/14747731.2025.2465166

Sondern, D., Arnholz, N., & Hertel, G. (2025). Employment Negotiations With an Algorithm? How AI as Negotiation Counterpart Would Affect Negotiators' Trust and Subjective Value Expectations. *Conflict Resolution Quarterly,* 2025, 0, 1–9. https://doi.org/10.1002/crq.21472

Stevens, M., Wehrens, R., Kostenzer, J., Weggelaar-Jansen, A. M., & De Bont, A. (2022). Why personal dreams matter: How professionals affectively engage with the promises surrounding data-driven healthcare in Europe. *Big Data & Society*, *9*(1), 1–13. DOI: 10.1177/20539517211070698

van Lente, H. (2012). Navigating foresight in a sea of expectations: Lessons from the sociology of expectations. *Technology Analysis & Strategic Management*, *24*(8), 769–782. DOI: 10.1080/09537325.2012.715478

Weick, K.E. (1995). *Sensemaking in organizations*. Sage.

Willcocks, L. (2020). Robo-apocalypse cancelled? Reframing the automation and future of work debate. *Journal of Information Technology*, *35*(4), 286–302. https://doi.org/10.1177/0268396220925830

Willim, R. (2024). *Mundania: How and where technologies are made ordinary*. Bristol University Press.