



Algorithmic Immaturity and the Privatization of the Public Use of Reason

Micael Rosa Silva 

Professor of Philosophy in Basic, Technical, and Technological Education (EBTT), Federal Institute of Paraná (IFPR), Palmas Campus, Palmas, Brazil. E-mail: micael.silva@ifpr.edu.br

Article Info

Article type:

Research Article

Article history:

Received 10 September 2026

Received in revised form 15 September 2026

Accepted 27 September 2026

Published online 25 October 2026

Keywords:

Immanuel Kant, Artificial Intelligence, Algorithmic Immaturity, Public Use of Reason, Autonomy, Philosophy of Technology

ABSTRACT

Artificial intelligence revives Kant's problem of immaturity under conditions he could not have anticipated. Current debates typically locate the threat in dependence, cognitive offloading, manipulation, or the opacity of automated decisions. These diagnoses remain incomplete because they do not explain when technical assistance becomes tutelage, and they rarely connect delegated judgment to Kant's public use of reason. This article introduces algorithmic immaturity as a two-dimensional condition: At the agential level, recurrent delegation displaces the exercise of judgment; critical reappropriation may remain available but unused, or may itself be impaired. At the political level, algorithmic infrastructures personalize the visibility, salience, formulation, and circulation of reasons so thoroughly that ostensibly public speech is enclosed within privately ordered environments. Across both dimensions, the condition is self-incurred when critical reappropriation remains genuinely available but is persistently left unused. Algorithmic tutelage names cases in which technical or institutional arrangements impair or foreclose that capacity. The argument reconstructs Kant's 1784 essay on enlightenment as its central axis, in critical dialogue with the philosophy of technology and recent work on human autonomy and algorithmic governmentality. It distinguishes ordinary reliance on technical systems and expertise from relations of tutelage through a four-dimensional test of critical reversibility: detectability, intelligibility, contestability, and recoverability. Maturity in the algorithmic age requires more than individual vigilance: it depends on technical and political arrangements that preserve critical reappropriation and the reciprocal publicity of reasons.

Cite this article: Silva, R. M. (2026). Algorithmic Immaturity and the Privatization of the Public Use of Reason. *Journal of Philosophical Investigations*, 20(56), 317-336. <https://doi.org/10.22034/jpiut.2026.74670.4749>



© The Author(s).

Publisher: University of Tabriz.

Intruduction

The Unexpected Return of Immaturity

The promise attached to artificial intelligence is often expressed in the language of assistance. A system retrieves what one cannot remember, compares more cases than one could inspect, drafts what one has not yet formulated, and recommends a course of action amid excessive information. Each operation can enlarge human agency. The same operations can also settle quietly into the place from which a person understands a situation and answers for a judgment. This ambiguity gives Kant's short essay on enlightenment an unexpected contemporary force. Kant describes enlightenment as the emergence of the human being from self-incurred immaturity (*Unmündigkeit*; in Portuguese, *menoridade*).¹ Immaturity is an inability to use one's own understanding without another's direction, even though the relevant incapacity does not arise from a natural lack of understanding (Kant 1983, 41; AA 8:35)². His examples are strikingly mundane: a book that understands for its reader, a spiritual adviser who supplies a conscience, a doctor who prescribes a regimen. Guardianship succeeds through the relief it offers.

Artificial intelligence intensifies this old attraction. Recommendation systems arrange what deserves attention; predictive models sort persons and possible futures; large language models answer questions, summarize evidence, propose interpretations, and compose reasons in the user's voice. Their authority need not take the form of an order. It can emerge from the repeated experience that accepting a fluent result is easier than reconstructing how it was reached. The question is therefore less dramatic than whether machines will dominate humanity. It concerns the point at which assistance acquires the practical structure of tutelage.

Recent scholarship has brought this question close to Kant. Schwarting and Reihlen (2026) argue that pervasive delegation to artificial intelligence may produce a new self-imposed immaturity by eroding autonomy, responsibility, individuality, and dignity. Catena's (2026) review of the literature on AI and human autonomy likewise identifies cognitive outsourcing, dependence, deskilling, and the loss of control over automated decisions as recurring concerns. Her reconstruction also highlights the idea of "meta-autonomy," understood as the power to decide whether to decide, including the possibility of reversible delegation to AI (Floridi et al. 2018; Catena 2026). These contributions establish that the problem cannot be reduced to isolated errors or malicious applications. They leave two issues open. A concept is needed that marks the threshold at which delegation becomes tutelage without treating reliance itself as a failure. The individual capacity to choose and judge must also be connected to the social conditions under which reasons become public. The present account joins these issues by identifying critical reversibility as the threshold of tutelage and the privatization of the public use of reason as its political form.

This article develops that concept. Algorithmic immaturity names a relation in which recurrent algorithmic guidance displaces the exercise of one's own judgment. Recurrent reliance alone does not establish immaturity. Delegation remains compatible with maturity

when a person can understand the system's role, challenge or refuse its outputs, and reassume the delegated operation when circumstances require. The self-incurred form arises when those possibilities remain genuine, intelligible, and practicable, but the person persistently declines to use them and treats algorithmic guidance as a substitute for judgment. Algorithmic tutelage names a structurally produced condition in which technical or institutional arrangements erode or foreclose those possibilities, so that resolution alone cannot recover the delegated capacity. On the basis of Kant's account, this article reconstructs two modes of origin and imputation without attributing the distinction itself to Kant. The modes cut across the agential and political dimensions: they explain how the condition arose and to whom it is imputable; the dimensions identify whether it manifests in personal judgment or in the shared conditions of public reason.

Critical reappropriation provides the first part of the argument. Delegation remains critically reversible when those affected can recognize where algorithmic mediation has shaped a judgment, demand intelligible grounds appropriate to the stakes, contest and correct the result, refuse it without prohibitive penalty, and recover responsibility for the operation that was delegated. This standard does not require every user to inspect source code or reproduce a model's computation. It asks whether a path remains open from reliance back to answerable judgment. The criterion is practical and institutional as well as cognitive. A patient need not become a radiologist, for example. A consequential diagnosis must nevertheless remain reviewable by accountable professionals and challengeable through a human process.

The second part returns to a distinction within Kant's essay that is frequently overshadowed by *sapere aude*: the distinction between the private and public uses of reason. The public use occurs when a scholar addresses the reading world in their own person; it must remain free because public criticism is the medium through which enlightenment becomes possible (Kant 1983, 42; AA 8:37). Contemporary algorithmic systems mediate the conditions of that address. They rank visibility, infer audiences, individualize relevance, and optimize circulation according to objectives that are rarely open to the speakers and readers involved. Generative systems now enter an earlier stage as well, helping determine the vocabulary, evidence, and argumentative form through which a claim becomes sayable. The result can be a privatization of the public use of reason: speakers address a public, but proprietary, personalized, and behaviorally adaptive systems order the routes by which their claims appear and gain uptake.

The argument proceeds through conceptual reconstruction and critical comparison. The first section reconstructs Kantian immaturity as a relation among understanding, courage, and guardianship, then asks what changes when the guardian becomes an adaptive infrastructure. The second places calculation and judgment in dialogue with Simondon, Stiegler, Heidegger, Weizenbaum, and Dreyfus in order to formulate critical reversibility without imagining a human subject untouched by technics. The third shows how algorithmic mediation can deform the public use of reason and specifies the institutional conditions under which AI may instead support it. The resulting thesis is deliberately conditional. The use of AI and the delegation of cognitive work do not, by themselves, constitute algorithmic immaturity. The decisive

questions concern whether judgment remains reappropriable and whether reasons remain capable of entering a shared field of contestation.

1. Immaturity and Tutelage from Kant to the Algorithmic Guardian

Kant's definition joins an incapacity to a relation. The immature person cannot use understanding without direction, and the direction comes from another. The "another" need not coerce. Kant attributes self-incurred immaturity to a lack of resolution and courage, then observes that laziness and cowardice make it comfortable to remain a minor (Kant 1983, 41; AA 8:35). Guardians cultivate that comfort by taking over burdensome acts. Once they have domesticated those under their care, they present independent movement as dangerous. A learned dependency eventually appears as evidence that dependency was necessary from the beginning.

Kant's syntax makes the imputation conditional: "Selbstverschuldet ist diese Unmündigkeit, wenn die Ursache derselben nicht am Mangel des Verstandes, sondern der Entschließung und des Muthes liegt, sich seiner ohne Leitung eines andern zu bedienen" (Kant 1983, 41; AA 8:35)³. In this sentence, the conjunction states the condition under which immaturity is self-incurred: its cause must lie in a lack of resolution and courage rather than in a lack of understanding. The text therefore distinguishes Unmündigkeit from its imputation as selbstverschuldet, although Kant's essay takes self-incurred immaturity as its principal target. The self-incurred minor possesses the understanding required for independent thought and persistently avoids exercising it. Kant captures this practical structure in the sentence: "I need not think, if only I can pay: others will readily undertake the irksome work for me" (Kant 1983, 41; AA 8:35).

Kant also assigns the guardians (Vormünder) an active role in sustaining and deepening tutelage. They first reduce those in their care to a docile condition (Kant's term is Hausvieh, domestic livestock) and confine their movement to an apparatus of leading-strings (Gängelwagen) (Kant 1983, 41; AA 8:35–36). Only then do they present the passage to maturity (Mündigkeit) as perilous. The sequence matters: comfort and dependency are cultivated before danger is invoked, and the warning is addressed to a capacity that has already been prevented from practice.

Kant then describes a practical incapacity produced within this relation of guardianship. Immediately after the passage on the Vormünder, he observes:

It is difficult for any individual man to work himself out of the immaturity that has all but become his nature. He has even become fond of this state and for the time being is actually incapable of using his own understanding, for no one has ever allowed him to attempt it. (Vorderhand wirklich unfähig, sich seines eigenen Verstandes zu bedienen, weil man ihn niemals den Versuch davon machen ließ.) (Kant 1983, 41; AA 8:36)

The decisive clause is *weil man ihn niemals den Versuch davon machen ließ*: no one allowed the attempt. Kant's guardians do more than benefit from an existing refusal; they help produce a practical incapacity by withholding the exercise through which understanding becomes independently usable. This article calls that structurally produced condition algorithmic tutelage. The category is an analytical reconstruction from Kant's account of guardianship, not a category that Kant himself explicitly formulates⁴.

[Foucault \(1997\)](#) reads Kant's exit (*Ausgang*) from immaturity as "the permanent reactivation of an attitude—that is, of a philosophical ethos that could be described as a permanent critique of our historical era" ([Foucault 1997, 312](#)). Algorithmic tutelage has no single genealogy. It can emerge when initially voluntary delegations become normalized, embedded in technical standards and institutional routines, and converted into defaults that later users cannot meaningfully refuse. It can also be imposed from the outset by employers, schools, platforms, welfare agencies, or public authorities, without any prior act of delegation by those subjected to it. Both paths can impair critical reappropriation. Kant's warning that one age may not bind the next so as to make further enlightenment impossible ([Kant 1983, 43; AA 8:39](#)) gives inherited tutelage a specifically Kantian point of reference.

Kant's account joins three elements: the capacity to understand, the resolution to exercise it, and social conditions in which such exercise is possible. He imputes self-incurred immaturity to subjects who possess the capacity and decline to use it; he also describes guardians who cultivate dependency and obstruct the attempt at independent use. Responsibility must therefore follow effective power within the relation. The stronger a person's practical ability to recognize, contest, and reverse algorithmic guidance, the stronger the grounds for holding that person answerable for persistently avoiding judgment. As opacity, coercion, dependency, or the absence of viable alternatives weakens those abilities, the basis for self-imputation diminishes and responsibility shifts toward the actors able to configure, impose, audit, or change the arrangement. Collective enlightenment also requires freedom to address a public.

The algorithmic guardian occupies this relational position without resembling Kant's human examples in every respect. It is ambient rather than episodic. It accompanies search, navigation, consumption, education, employment, health, and political communication. It is adaptive because its outputs respond to behavioral traces and to the patterns learned from other users. It is also infrastructural: refusing one recommendation does not remove the ranking systems, data classifications, and institutional models that organize the environment in which the next choice occurs. A guardian of this kind can guide without presenting itself as an authority. Defaults, rankings, completions, risk scores, and fluent drafts make one path more available than its alternatives.

Calling a system a guardian attributes a function within a relation, not a mind or intention to the artifact. Simondon warned that culture oscillates between treating machines as empty utilities and imagining them as autonomous rivals. Both attitudes conceal the human reality crystallized in technical objects and the social relations through which they operate ([Simondon](#)

1980, 1–4). The relevant guardian may include designers, deployers, institutional routines, economic incentives, training data, interfaces, and a model’s statistical operations. Responsibility does not disappear simply because agency is distributed across this arrangement. The term “algorithmic” identifies the distinctive mediation through which direction is personalized, scaled, and rendered difficult to perceive.

This functional account clarifies why opacity alone does not define immaturity. Many legitimate forms of reliance involve processes that a person cannot fully reconstruct. Few passengers understand an aircraft’s avionics, and no reader verifies every link in the production of a scholarly book. Modern agency depends on testimony, institutions, expertise, and artifacts. People can remain mature in such relations when their reliance is responsive to reasons, bounded by role and context, and connected to procedures of correction. Conversely, a technically explainable system can exercise tutelage if its use has become compulsory, its categories cannot be challenged, or the user has surrendered the effort to judge what the explanation means.

Kant’s own distinction between public and private reason already rejects the fantasy that maturity entails deciding everything alone. The private use of reason is the use made within an entrusted civil post: an officer obeys a command while serving, a citizen pays an assessed tax, and a clergyman teaches under an ecclesiastical commission. Such coordination can require restrictions. The same persons must remain free, as scholars, to criticize the rules and doctrines governing their roles before the public (Kant 1983, 42–43; AA 8:37–38). Maturity is compatible with organized dependence when role-bound action does not swallow the capacity for public judgment. The problem begins when no standpoint remains from which the arrangement can be examined and revised.

Algorithmic systems can narrow that standpoint. A worker managed through performance metrics may be able to dispute a single score. The categories that define good performance can still remain inaccessible. An applicant may correct a data error and remain unable to challenge the model’s conception of risk. A student can ask a language model for alternative answers and gradually lose the conceptual command needed to assess any of them. In each case, local options coexist with a deeper asymmetry: the person acts within a field whose relevant distinctions have been established elsewhere. The appearance of choice can therefore survive the loss of authorship over the terms of choice.

The phrase “self-incurred” requires particular care. Kant’s attribution retains its force when a user can identify the delegation, understands its significance, has a viable alternative, and repeatedly chooses relief from judging. Repeatedly choosing convenience can become a practice of self-disqualification. Encounters required by employers, schools, insurers, public agencies, or dominant platforms alter those conditions because refusal may carry serious economic and social costs. Designers and institutions can further weaken agency by exploiting attentional limits, concealing alternatives, or converting occasional reliance into a durable

default. Such constraints reduce the grounds for self-imputation in proportion to their effect on the user's practical capacity to do otherwise.

Algorithmic immaturity is consequently a mixed condition in many real cases. Habit and unexamined trust may coexist with restricted avenues of appeal, compulsory workplace systems, market concentration, manipulative interface design, and an unequal distribution of technical knowledge. Attribution should track the power each participant has to create, recognize, contest, or withdraw from the relation. A user's share of responsibility increases with viable opportunities to understand the mediation and recover judgment. Institutional responsibility increases with the power to conceal, enforce, audit, and revise the arrangement. The same case may therefore contain unequal shares that change over time. Kant's demand for courage remains decisive where independent action is practicable. His account of the guardians directs attention to those who shape the conditions of that action.

The conceptual gain lies in treating immaturity as more specific than dependence. Dependence is an inescapable feature of social and technical existence. Tutelage arises when direction occupies the place of one's understanding and blocks the return from guided action to answerable judgment. An algorithmic system may be extraordinarily capable and still remain assistance. A modest system may become a guardian when its classifications are final in practice. The boundary turns on the organization of the relation, not on a general ranking of human and machine intelligence.

1. Between Calculation and Judgment When Assistance Becomes Tutelage

Any criterion for algorithmic immaturity must begin with the technical constitution of human agency. Stiegler's account of technics makes this point with particular force. Human memory and temporality are formed through exteriorization; inherited traces, tools, and writing preserve a past that no individual has lived and make it available for appropriation. The interior is constituted through exteriorization rather than existing as a complete faculty that later chooses to use tools ([Stiegler 1998, 140–142](#)). A demand for unmediated thought would therefore mistake the conditions of thought for obstacles to it. Writing, archives, diagrams, calculators, and digital systems can release attention, extend memory, and enable forms of reflection that unaided cognition could not sustain.

Simondon reaches a compatible conclusion through the mode of existence of technical objects. Alienation follows from a failure to integrate technical reality into culture and meaning. He associates technical perfection with openness and a margin of indetermination in which the human being remains an interpreter and coordinator among machines ([Simondon 1980, 3–5](#)). This image supplies an alternative to both mechanical submission and sovereign mastery. Mature technical agency develops through participation in an open ensemble whose operations can be interpreted, adjusted, and recomposed.

These accounts change the normative question. The issue cannot be how much thinking has been exteriorized, because exteriorization also carries thought across persons and generations.

What matters is the form in which an exteriorized operation returns to the subject. A notebook returns a trace that the writer can inspect and revise. A calculator returns a result within a rule-governed practice whose relevance the user may still judge. A language model can return a synthesis that opens inquiry, or a polished conclusion that hides uncertainty and discourages reconstruction. The artifact's technical complexity does not settle which relation has been formed.

Weizenbaum's distinction between judgment and calculation helps identify the danger. He objected to the modern tendency to recast human dilemmas as formally specifiable problems, especially when computational success makes a procedure appear competent to determine the ends for which it should be used. Some acts of thought, he argued, "ought to be attempted only by humans" because their delegation would evade moral responsibility rather than merely improve a calculation (Weizenbaum, 1976, 13). His claim concerns the ends, interpretation, and responsibility attached to computational use. A successful output cannot derive those elements from computational power alone.

Dreyfus pursued the epistemic side of this distinction. Human intelligence encounters objects within a meaningful situation; relevance is usually grasped through embodied involvement before it is decomposed into explicit features. Formal systems, by contrast, require a representation in which certain elements have already been selected as significant (Dreyfus 1972, xv–xvii, 148–150). Contemporary AI has exceeded many empirical expectations of early critics, including some of Dreyfus's forecasts about performance. His durable question concerns the background that performance presupposes. A model's successful output does not show that the practical meaning of the situation, or responsibility for selecting what counts, has been exhausted by the representation.

Heidegger's analysis of modern technology broadens the point from individual tools to a mode of revealing. Modern technics orders beings so that they appear as resources available for calculation and use; humans themselves become implicated in this ordering (Heidegger 1977, 12–20). Algorithmic mediation inherits this power of disclosure. A prediction system makes a person appear as a risk score, a recommender makes culture appear as an ordered set of likely engagements, and a generative model makes language appear as a space of statistically available continuations. These disclosures can be illuminating. Their danger grows when they monopolize relevance and withdraw the framing operation from question.

Heidegger, Simondon, Stiegler, Weizenbaum, and Dreyfus do not share a common diagnosis of technology. Heidegger's account of enframing (*Gestell*) treats the calculative revealing of modern technics as a historical destiny that instrumental correction cannot simply repair; Simondon, by contrast, holds that technical culture, properly integrated, opens a margin of human interpretation rather than closing it. No synthesis of these positions is attempted here, nor would one be faithful to either thinker. What each supplies instead is a distinct, independently defensible constraint on critical reversibility. Stiegler and Simondon demonstrate that reversibility cannot require the impossible condition of unmediated cognition.

Weizenbaum and Dreyfus show that formal success does not settle relevance or responsibility, and Heidegger shows that the relevant danger can operate at the level of disclosure itself, prior to any particular error. The four dimensions below draw on these constraints severally, not on a shared ontology of technology that would need independent construction and defense.⁵

Critical reversibility names the practical capacity to keep algorithmic framing within judgment. It can be assessed along four dimensions. First, detectability: those affected must be able to recognize that mediation has occurred and identify where it bears on the judgment. Second, intelligibility: reasons and evidence must be available at a level appropriate to the decision and its consequences. Third, contestability: the grounds and outcome must remain open to contradiction, correction, and, where applicable, refusal without prohibitive penalty; when an immediate role-bound act cannot be refused, the rule authorizing it must remain publicly criticizable and institutionally revisable. Fourth, recoverability: a person or accountable human institution must possess both the competence and authority to reassume the delegated operation. The test is conjunctive at the threshold and graduated in attribution. Each dimension must reach a level adequate to the stakes, and a severe failure in one cannot be compensated by strength in the others. When all four corresponding capacities remain adequately available and a person persistently leaves them unused, Kant's criterion supports an attribution of self-incurred algorithmic immaturity. When structural arrangements place one or more dimensions below the practical threshold, algorithmic tutelage is present. A case is mixed when persistent non-use and structural impairment jointly produce the deficit; responsibility then tracks each participant's effective power over it.

The four dimensions should be understood as a contemporary operational reconstruction of Kantian concerns, not as criteria Kant himself formulated. Detectability makes visible the guidance by another that defines Unmündigkeit. Intelligibility translates the requirement of using one's own understanding into conditions for mediated judgment. Contestability extends Kant's demand that the rules and doctrines governing a role remain open to public criticism, even where particular role-bound acts performed under them do not; it is the arrangement, not each delegated act, that must remain answerable to a public. Recoverability reconstructs Kant's *Ausgang aus der Unmündigkeit* as the practical possibility of emerging from tutelage by reassuming the delegated judgment.

These dimensions specify, in practice, what the Kantian requirement of a reappropriable delegation demands: an operation must remain something the affected person can question, correct, and resume. A philosophical criterion arrived at this way can still find an echo in positive law: Article 14 of the European Union's Artificial Intelligence Act. For high-risk systems, it requires human overseers to understand relevant capacities and limitations, remain aware of automation bias, interpret outputs, and be able to disregard, override, reverse, or halt them ([European Union 2024, Art. 14](#)). These provisions correspond most closely to intelligibility and contestability. Critical reversibility also asks whether a person or institution can exercise those formal powers without prohibitive cost and retain the competence needed to

reopen the judgment. A nominal override does not satisfy the criterion when organizational pressure, deskilling, or lack of authority predetermines the human response.

Article 14 applies only to systems formally classified as high-risk under a narrow regulatory taxonomy — biometric identification, employment, credit scoring, law enforcement, and similar categories. Critical reversibility is not so restricted: it applies equally to the everyday, formally unregulated uses — a travel-planning suggestion, a drafting assistant, a recommended course of action — through which algorithmic tutelage can also develop.

This criterion extends the idea of reversible delegation found in discussions of meta-autonomy (Floridi et al. 2018; Catena 2026). Although the power to decide whether to delegate is indispensable, a visible on/off switch offers too thin a measure. A person may technically refuse a model. The surrounding institution, however, may offer no viable way to act on that refusal. One may request an explanation that merely redescribes an output, or assume formal control after the relevant skill has atrophied. Critical reversibility therefore concerns the cultivated capacities, organizational procedures, and material options that make the resumption of judgment real.

An advocate of extensive delegation can press a stronger objection. Time and attention are scarce, and an autonomous person may allocate both by deciding what to delegate. Assigning search, translation, drafting, or diagnosis to a more reliable system can free a person for ends judged more important. Maintaining direct competence in every operation may waste resources without improving the final judgment. Much of this objection is sound. Critical reversibility asks whether the choice to delegate remains exercisable over time and under the conditions created by delegation. When users can assess the stakes, seek independent advice, refuse the output at tolerable cost, and recover responsibility, delegation may enlarge autonomy. The objection loses force when immediate efficiency erodes the capacities required for later scrutiny, makes institutional exit prohibitively costly, or binds people who never chose the arrangement. Any calculation of opportunity cost must also include the future value of retained judgment. Once a system has displaced the relevant skill or authority, reclaiming it may become impossible precisely when disagreement or failure makes it necessary.

Two further literatures reinforce this defense of delegation. The extended mind thesis (Clark & Chalmers 1998) holds that a cognitive process can be legitimately realized outside the biological organism when an external resource is reliably available, trusted, and easily incorporated into ongoing action — a notebook or a device can be constitutive of memory, not merely an aid to it. Clark and Chalmers do not themselves specify detectability, reliability-tracking, contestability, or costless exit; their thesis explains why external dependence is not inherently alienating, while critical reversibility supplies the further, normative criterion that the thesis alone does not provide. Social epistemology's account of testimony similarly holds that accepting a claim on another's authority, without independently verifying it, is not thereby irrational or immature; it is how most human knowledge is acquired and transmitted (Coady 1992). What distinguishes a well-functioning extended cognitive system, or a well-functioning

practice of testimony, from algorithmic tutelage is exactly what the four dimensions specify: the external resource or testifier can be identified as a source, its reliability can be tracked and revised in light of evidence, and the relation can be renegotiated or abandoned without prohibitive cost. Kant's allowance for role-bound duties shows that mature agents may participate in institutional coordination while retaining the freedom to criticize its governing rules publicly. The criterion developed here preserves that structure. Extended cognition and testimonial reliance are not competitors to autonomous judgment; they become immaturity only when they compromise the reflexive governance of the dependency itself.

[Schwartz and Reihlen \(2026\)](#) distinguish effects on autonomy, responsibility, individuality, and dignity, and they show how opaque delegation can erode the capacities needed to reclaim autonomy. Their account gathers voluntary reliance, structural constraints, and intergenerational effects under the rubric of self-imposed immaturity. The present contribution reconstructs from Kant's conditional attribution and his account of the guardians a distinction between self-incurred algorithmic immaturity and structurally produced algorithmic tutelage. That distinction allows responsibility to be distributed rather than presuming a single locus of fault. The argument also connects the agential problem to Kant's public use of reason and operationalizes critical reversibility as a conjunctive, risk-sensitive test. The amount of delegation remains an indirect indicator. Tutelage is present when delegation closes the route by which reasons can be inspected, responsibility resumed, and the operative frame contested. The relevant unit of analysis may therefore be a public institution rather than an isolated user.

Consider three boundary cases. A researcher uses a language model to generate objections to a thesis, checks each objection against the cited literature, rewrites the argument, and can explain why the final claims survive. Delegation here furnishes resistance that the researcher reappropriates. A second researcher accepts a fluent review, cites sources that have not been read, and becomes unable to defend the synthesis without reopening the model. The same class of tool now occupies the place of understanding. In a high-stakes administrative case, an official may scrutinize a model's recommendation and still lack authority to depart from it without career risk. Formal human oversight then disguises an institutionally irreversible delegation.

The temporal dimension matters because immaturity can develop before it becomes visible. Carr's account of internet use linked habits of rapid switching and fragmented attention to difficulty sustaining the reflective concentration cultivated by deep reading ([Carr 2010, 5–7, 115–118](#)). His evidence concerns the broader internet rather than current generative systems, so it cannot establish an effect specific to AI. [Gerlich's \(2025\)](#) mixed-method study analyzes 666 valid cases from a United Kingdom sample and reports a negative association between frequent AI-tool use and critical-thinking measures, with cognitive offloading as a mediator. Participants were recruited through a combination of convenience and purposive sampling; this non-probability recruitment, the cross-sectional design, and partial reliance on self-report prevent a causal conclusion. Taken cautiously, these works support a narrower philosophical

point: repeated patterns of mediation can alter the capacities through which future outputs are evaluated. Reversibility has to be maintained as a practice; an exit option offered after prolonged deskilling may arrive too late.

The practical weight of the four dimensions depends on the stakes. A person need not preserve manual competence for every routine calculation. The required degree of traceability, challenge, and retained skill rises with a decision's consequences and with the difficulty of correcting it later. A music recommendation and a criminal risk assessment do not demand the same safeguards. Collective systems of expertise can preserve mature agency when their authority is accountable, their conclusions remain contestable, and affected persons retain standing to demand reasons. The Kantian reconstruction proposed here can accommodate an unavoidable division of cognitive labor, provided that it does not harden into unanswerable guardianship.

This yields a practical distinction. Assistance expands or reorganizes agency while leaving the assisted person connected to the reasons, responsibilities, and revision of the act. Tutelage supplies direction in a manner that weakens or forecloses that connection. The difference can change over time and across institutions, even with identical software. Design choices that expose uncertainty, preserve sources and intermediate decisions, invite counterargument, and support meaningful refusal can strengthen reappropriation. Interfaces optimized for frictionless acceptance, institutions that reward conformity to scores, and systems that make appeal costly push the relation toward immaturity. The normative object is the human-technical arrangement in which judgment occurs.

3. Algorithmic Immaturity and the Public Use of Reason

The preceding account would remain incomplete if maturity were treated as a private possession. Kant's essay assigns a distinctive role to the public use of reason. A person uses reason publicly when speaking as a scholar before the reading world, in their own name, and submitting claims to unrestricted examination. Kant allows discipline within an office because the officeholder can still criticize the governing rule as a member of the public. Only the free public use of reason can produce enlightenment among people, since errors become corrigible through the circulation and confrontation of reasons (Kant 1983, 42; AA 8:37). Maturity thus depends on a communicative relation, not solely on an inward faculty.

Kant's controlling contrast is between *privater Vernunftgebrauch* and *öffentlicher Vernunftgebrauch*. The private use names role-bound reasoning within an entrusted civil post; it does not mean communication under private ownership. The public use names the scholar's address in their own person to the reading world, rather than state ownership, mass visibility, or a physical venue. O'Neill (1986) clarifies its normative structure through publicizability: reasons must remain communicable and answerable to an unrestricted audience. Corporate ownership of a feed therefore does not transform speech into Kantian private use. Algorithmic

mediation can instead obstruct the publicity of public reason by governing visibility, reach, and the practical possibility of reply.

[Laursen \(1986\)](#) establishes the political and historically contestable force of Kant's vocabulary: public criticism can exceed assigned roles and unsettle the authorities that organize them. [Borghi \(2006\)](#) develops a different, metaphysical reading, and this article does not adopt his description of public reason as an "empiric free zone." His critique nonetheless supplies one precise insight for the present argument. Architecturally open media and widespread knowledge sharing can coexist with tutelage when pre-formed judgments and standardized perspectives organize what counts as public exchange ([Borghi 2006, 184–186](#)). The privatization of the public use of reason names the contemporary enclosure of the material and epistemic conditions through which reasons become visible, meet objections, and remain open to unrestricted scrutiny. It does not rename Kant's private use.

Digital platforms complicate this relation because they combine publication with individualized distribution. A user can address everyone in principle. In practice, a ranking system decides which fragments become visible to which inferred profiles. Sunstein's "Daily Me" captures the political attraction and cost of an architecture of control. Personalization can serve real interests and can also reduce unplanned encounters and the shared experiences on which democratic deliberation depends ([Sunstein 2017, 1–9](#)). [Kozyreva et al. \(2020, 104–109\)](#) show more generally that online environments are choice architectures. Their adaptive design regulates access to information and competes for attention under limited transparency and public oversight.

Empirical research does not support a uniform picture of sealed filter bubbles. In a large study of Facebook, [Bakshy et al. \(2015, 1130–1132\)](#) found that algorithmic ranking reduced exposure to cross-cutting material, but network composition and individual choices also played substantial, sometimes larger, roles. [Dubois and Blank \(2018, 729–745\)](#) likewise argue that echo chambers are often overstated once political interest and diverse media repertoires are considered. These findings qualify any simple causal claim about personalization and ideological isolation. The argument here does not depend on informational closure. Even users who encounter heterogeneous views may remain unable to know or contest the privately governed criteria that allocate salience, sequence, and audience.

Rouvroy and Berns offer the most direct account of what happens to subjectivity in this arrangement. Algorithmic governmentality collects dispersed traces, produces correlations that appear to arise directly from reality, and acts preemptively on possible behavior. It addresses statistical profiles without calling a reflexive subject to give an account, thereby reducing occasions for critique and recalcitrance ([Rouvroy and Berns 2013, 167–176](#)). They describe personalization as "a colonization of public space by a hypertrophied private sphere" ([Rouvroy and Berns 2013, 165](#)). The formulation is especially relevant to Kant: the computationally tailored environment can absorb public address into a sequence of private orientations before a common object of dispute takes shape.

Here, privatization is an expressly contemporary term for the enclosure of the conditions of Kantian public use. It has three related layers. Proprietary privatization places communicative infrastructure under corporate ownership and rules. Epistemic privatization makes each person's informational environment a function of inferences that others cannot inspect. Political privatization removes decisions about visibility, relevance, and audience formation from public justification. A state-operated system may produce comparable administrative de-publicization when it individualizes information and preempts contestation. The common problem is insulation from public answerability, not private ownership by itself.

Surveillance capitalism supplies strong incentives for the first pattern. Zuboff describes an economic logic in which behavioral data are converted into prediction products and increasingly used to “tune, herd, and condition” action through ubiquitous computational environments (Zuboff 2019, 8, 376–377). Heck (2024, 9–33) traces related forms of behavioral modulation through microtargeting, automated management, and the asymmetries produced by Big Tech infrastructures. Under these conditions, communication is valued for its capacity to predict and modify behavior. A reason becomes a signal of engagement, preference, risk, or susceptibility. The system can operate efficiently without recognizing the speaker as someone who offers a claim for judgment.

Recommendation systems primarily affect publicity by ordering visibility and reach. Generative AI can also affect publicizability because language models intervene during the formation of an utterance. They propose questions, summaries, examples, and tones for anticipated readers. Used reflectively, these capacities can help a person test an intuition, cross linguistic boundaries, and encounter objections. Under passive acceptance, they pre-format reason before public scrutiny. The user remains the nominal speaker but may no longer command the path by which the claim was formed.

This loss is not adequately described as plagiarism or inauthenticity. A text may be original in a legal sense and factually reliable while still exemplifying immaturity. The decisive issue is whether the speaker can answer for the reasons presented, revise them under criticism, and recognize the model's contribution to their form. Public reason needs authorship in this practical sense of answerability. When fluent generation severs assertion from the labor of judgment, public exchange fills with statements that have identifiable speakers without responsible advocates. Criticism then returns to the model for another response, and a conversation that appears active can proceed without anyone reassuming the contested judgment.

The public dimension also reveals why individual media literacy is insufficient. Even a highly skilled user who verifies sources and resists recommendations may participate in an environment that fragments audiences and conceals its allocation of attention. Critical reversibility at the personal level does not by itself restore common objects of dispute. Conversely, a publicly accountable platform cannot guarantee maturity if participants habitually outsource their understanding. Algorithmic maturity requires a conjunction:

delegated judgments must remain reappropriable by agents, and the infrastructure must preserve the reciprocal publicity through which reasons are tested.

Reciprocal publicity does not demand identical information for all citizens. Such uniformity would suppress legitimate plurality and reproduce the gatekeeping of older mass media. Personalized systems may expose people to specialized knowledge and voices previously excluded from public attention. The relevant threshold is reached when personalization becomes unobservable and unanswerable to those whose communicative relations it organizes. Participants must be able to know that selection is occurring, understand its governing purposes at a meaningful level, compare alternative orderings, and contest rules that systematically structure public visibility. Some spaces also need deliberate exposure to common matters and unchosen perspectives, because a public cannot deliberate about objects that never become mutually perceptible.

Kant's distinction suggests, by analogy, a diagnostic of functional subordination. Algorithmic profiles are not civil posts in Kant's sense. They are predicted identities—consumer, borrower, voter, employee, or student—through which systems organize visibility and action toward selected outcomes. When public speech is continuously routed through such profiles, a functional identity precedes the claim made in the speaker's own name. The space for speaking beyond the assigned function becomes narrower even where no sentence is censored. The analogy concerns this functional narrowing; it does not identify algorithmic profiles with Kantian private use.

The problem is therefore architectural. Freedom of expression remains necessary, though formal permission to speak cannot secure the public use of reason when infrastructures determine salience and uptake through hidden individualized criteria. A mature informational order must make room for reasons to exceed the profiles through which speakers are anticipated. It must preserve intervals in which prediction does not immediately become intervention, and in which a subject can appear through a response that the profile did not contain. [Rouvroy and Berns \(2013\)](#) describe such interruption as a condition for subjectification: reflexive suspension and the encounter among disparate elements allow a person to become more than a statistical double.

These requirements can be translated into institutional commitments without pretending that philosophy can specify a complete technical design. Systems that mediate consequential judgments should preserve records of relevant inputs, sources, changes, and human decisions so that responsibility can be reconstructed. Affected persons need a genuine right to challenge outcomes before accountable human agents who possess authority to change them. Educational and professional settings should cultivate the capacity to perform and explain core judgments even when AI routinely assists them. Public communication platforms should provide intelligible control over ranking, enable comparison with non-personalized or differently ordered views, disclose the objectives shaping recommendation, and support access to matters

of shared concern. Independent research and auditing require access sufficient to examine systemic effects on visibility and public discourse.

Kozyreva and colleagues (2020) frame many interventions as cognitive tools, including boosts that develop user competences and technocognitive changes to information environments. That approach is compatible with Kant when it enlarges the conditions for reflective agency. Designers can introduce friction when speed favors unexamined transmission, make provenance visible so that users can reconnect claims to evidence, and request reasons at moments when a recommendation might otherwise pass directly into action. Such an environment allows users to trust outputs without losing the means to question, refuse, and publicly revise a judgment.

People can also use artificial intelligence in ways that serve the public use of reason. They can translate texts across linguistic boundaries, make specialized arguments more accessible, trace policy claims across large documentary records, and expose latent assumptions through counterarguments or alternative formulations. Such uses contribute to enlightenment when the system's assistance returns users to sources, disagreement, and responsibility. Their value is measured by the quality of reappropriation and public contestation they make possible, rather than by the friction they remove.

The difference between an algorithmic guardian and an intellectual interlocutor lies here. A guardian resolves uncertainty on the user's behalf and quietly organizes the field in which further questions will occur. An interlocutor introduces material that the user can appropriate, dispute, and carry into relations with others. The same model can occupy either position. Its role emerges from interface, institutional setting, economic purpose, cultivated practice, and the user's disposition. This is why algorithmic immaturity is neither a property of software nor a moral defect located wholly within individuals. It is a relation whose political form becomes visible in the fate of public reason.

Conclusion

Sapere Aude in the Algorithmic Age

Kant's account of enlightenment remains useful because it places the burden of maturity at the meeting point of capacity, courage, guardianship, and freedom. Artificial intelligence changes each term. It extends cognitive capacities. Some capacities may be exercised less often as a result. It offers relief from uncertainty and effort, which can weaken the disposition to judge. Its guardianship is distributed across models, interfaces, institutions, and economic incentives. Its mediation reaches the public conditions in which claims become visible and answerable.

Algorithmic immaturity names a relation in which algorithmic guidance displaces the exercise of judgment. At the agential level, the condition ranges from the persistent non-use of an available capacity for critical reappropriation to structural conditions in which delegated operations cease to be reversible. At the political level, the circulation and formulation of reasons are enclosed within personalized systems whose ordering cannot itself be publicly

challenged. These dimensions can separate in particular cases. Maturity in an AI-shaped society requires both critical reappropriation and reciprocal publicity.

Self-incurred algorithmic immaturity and algorithmic tutelage have distinct genealogies that can intersect. The first arises when people persistently exchange judgment for convenience while a genuine and understood alternative remains available; this is the condition of self-imputation. Kant's *wenn* marks precisely that conditional relation. Tutelage may emerge through sedimentation as voluntary delegations become institutional defaults. It can also be imposed directly or inherited by people who never participated in those choices. Kant's guardians (*Vormünder*) illuminate the latter relation because they organize conditions in which independent movement is never attempted: *weil man ihn niemals den Versuch davon machen ließ*. The demand for courage governs the self-incurred case. In structurally produced tutelage, responsibility rests primarily with actors who control the technical and institutional conditions of judgment, with residual responsibility allocated according to each person's effective power to contest or change the arrangement. A formal right to question does not restore maturity when the question cannot alter an outcome or refusal excludes a person from education, work, health care, or civic participation.

The proposed criterion avoids making maturity synonymous with cognitive isolation. Human thought has always developed through technical exteriorization and social dependence. AI can deepen inquiry, widen access, and give public reason new reach. Its contribution remains emancipatory when assistance is organized so that users return from the system with greater command of reasons and a stronger capacity to face disagreement. The design of explainable outputs is part of this task; so are education, professional responsibility, human appeal, plural information orders, and public oversight of communicative infrastructures.

Sapere aude now asks for the courage to interrupt fluent answers, reopen classifications, and take responsibility for judgments that a system can produce more quickly. It also asks societies to build the material conditions under which such interruption matters. Enlightenment in the algorithmic age will depend on whether artificial intelligence becomes an infrastructure for shared inquiry or a comfortable arrangement of guardianship. That outcome is still open, provided that the power to reclaim judgment and the public space of reasons remain open with it.

Endnotes

1. English translations commonly render *Unmündigkeit* as "immaturity," although "minority" and "nonage" also occur. The concept proposed here was originally formulated in Portuguese as *menoridade algorítmica*. The Portuguese *menoridade* names the legal condition of a person who has not attained full civil capacity and remains subject to guardianship. Here it identifies a relation of tutelage, not a psychological diagnosis of childishness. The term therefore preserves a juridical and political resonance that the English "immaturity" can obscure.

2. AA abbreviates Kants gesammelte Schriften (Akademie-Ausgabe); thus, “AA 8:35” refers to volume 8, page 35. Citations to Kant’s essay pair the page number in Humphrey’s translation (Hackett, 1983) with the Academy pagination, allowing passages to be located across different translations.
3. Kant’s wenn introduces the condition under which Unmündigkeit is self-incurred. Humphrey renders it as “when” (Kant 1983: 41); Gregor also uses “when,” while translating Unmündigkeit as “minority” (Kant 1996: 17); Nisbet makes the conditional force explicit with “if” (Kant 1991: 54). The distinction therefore belongs to Kant’s syntax rather than being created by a translation.
4. Kant did not formulate a separate theory of structurally imposed immaturity. He nevertheless describes someone as temporarily wirklich unfähig because man ihn niemals den Versuch davon machen ließ—no one allowed the attempt—and a public previously placed under the guardians’ yoke (AA 8:35–36). O’Neill accordingly calls an unqualified attribution of self-incurrence a “questionable exaggeration” (1986: 536–537). These passages ground the article’s distinction among self-incurred, structurally produced, and mixed cases.
5. The article draws on Heidegger, Simondon, Stiegler, Weizenbaum, and Dreyfus for limited, independent claims rather than treating them as a unified school. Stiegler shows how technical exteriorization participates in memory and anticipation; Simondon shows how technical mediation can preserve a margin of interpretation within open ensembles. Weizenbaum distinguishes judgment from calculation, and Dreyfus shows how formal performance presupposes embodied, situational relevance. Heidegger locates danger at the level of disclosure, prior to any discrete error. Critical reversibility draws on these claims separately and departs from a strict reading of Gestell by retaining the possibility of technical and institutional redirection.

References

- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Borghi, M. (2006). The public use of reason: A philosophical understanding of knowledge sharing. *The International Journal of the Humanities*, 3(4), 179–186. <https://doi.org/10.18848/1447-9508/CGP/v03i04/41626>
- Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. W. W. Norton.
- Catena, E. (2026). AI and human autonomy: A literature review. *AI and Ethics*, 6, Article 126. <https://doi.org/10.1007/s43681-025-00958-4>
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- Coady, C. A. J. (1992). *Testimony: A philosophical study*. Clarendon Press.
- Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. Harper & Row.

- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745. <https://doi.org/10.1080/1369118X.2018.1428656>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*, L 2024/1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schäfer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Foucault, M. (1997). What is enlightenment? (C. Porter, Trans.). In P. Rabinow (Ed.), *Ethics: Subjectivity and truth* (Essential Works of Foucault 1954–1984, Vol. 1, pp. 303–319). The New Press. (Original work published 1984)
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), Article 6. <https://doi.org/10.3390/soc15010006>
- Heck, R. L. (2024). Algoritmos que nos governam: Reflexões filosóficas sobre propriedades críticas da atual estrutura da revolução informacional [Algorithms that govern us: Philosophical reflections on critical properties of the current structure of the information revolution]. *Philosophos*, 29(1), 1–45. <https://doi.org/10.5216/phi.v29i1.78173>
- Heidegger, M. (1977). The question concerning technology. In *The question concerning technology and other essays* (W. Lovitt, Trans., pp. 3–35). Harper & Row.
- Kant, I. (1900–). *Gesammelte Schriften* [Akademie-Ausgabe]. Königlich Preußische Akademie der Wissenschaften (later Walter de Gruyter).
- Kant, I. (1983). An answer to the question: What is enlightenment? (T. Humphrey, Trans.). In *Perpetual peace and other essays on politics, history, and morals* (pp. 41–48). Hackett Publishing. (Original work published 1784)
- Kant, I. (1991). *Political writings* (H. S. Reiss, Ed.; H. B. Nisbet, Trans.; 2nd enlarged ed.). Cambridge University Press.
- Kant, I. (1996). *Practical philosophy* (M. J. Gregor, Ed. & Trans.). Cambridge University Press.
- Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the Internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, 21(3), 103–156. <https://doi.org/10.1177/1529100620946707>
- Laursen, J. C. (1986). The subversive Kant: The vocabulary of "public" and "publicity." *Political Theory*, 14(4), 584–603. <https://doi.org/10.1177/0090591786014004003>
- O'Neill, O. (1986). The public use of reason. *Political Theory*, 14(4), 523–551. <https://doi.org/10.1177/0090591786014004001>
- Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships? (L. Carey-Libbrecht, Trans.). *Réseaux*, 177(1), 163–196. <https://doi.org/10.3917/res.177.0163>
- Schwartz, D., & Reihlen, M. (2026). Self-imposed immaturity in the age of artificial intelligence. *Philosophy of Management*. Advance online publication. <https://doi.org/10.1007/s40926-026-00386-2>

- Simondon, G. (1980). *On the mode of existence of technical objects: Part I* (N. Mellamphy, Trans.). University of Western Ontario. (Original work published 1958)
- Stiegler, B. (1998). *Technics and time, 1: The fault of Epimetheus* (R. Beardsworth & G. Collins, Trans.). Stanford University Press. (Original work published 1994)
- Sunstein, C. R. (2017). *Republic: Divided democracy in the age of social media*. Princeton University Press.
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.