

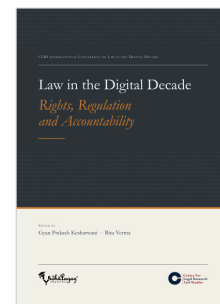
CHAPTER 1

Accountability in AI and Automated Decision-Making: From Explainability to Reviewable Administrative Governance

Tzu-Yu Chiou

in *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma (VidhiAagaz 2026), pp. 3–20

DOI <https://doi.org/10.63108/VAB.LDD.1.1>



CITATIONS

BLUEBOOK (21ST ED.)

Tzu-Yu Chiou, *Accountability in AI and Automated Decision-Making: From Explainability to Reviewable Administrative Governance*, in *LAW IN THE DIGITAL DECADE: RIGHTS, REGULATION AND ACCOUNTABILITY* 3 (Gyan Prakash Kesharwani & Ritu Verma eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Tzu-Yu Chiou, ‘Accountability in AI and Automated Decision-Making: From Explainability to Reviewable Administrative Governance’ in Gyan Prakash Kesharwani and Ritu Verma (eds), *Law in the Digital Decade: Rights, Regulation and Accountability* (VidhiAagaz 2026) 3.

CHICAGO (17TH ED.)

Chiou, Tzu-Yu. “Accountability in AI and Automated Decision-Making: From Explainability to Reviewable Administrative Governance.” In *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma, 3–20. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.1.1>.

APA (7TH ED.)

Chiou, T. (2026). Accountability in AI and automated decision-making: From explainability to reviewable administrative governance. In G. P. Kesharwani & R. Verma (Eds.), *Law in the digital decade: Rights, regulation and accountability* (pp. 3–20). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.1.1>

SOURCE

<https://doi.org/10.63108/VAB.LDD.1.1>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 1

Accountability in AI and Automated Decision-Making: From Explainability to Reviewable Administrative Governance

TZU-YU CHIOU*

This paper argues that the prevailing discourse on algorithmic accountability, which centers on technical explainability, offers an insufficient normative foundation for governing AI-driven automated decision-making (ADM) within public administration. Drawing on Lorenz von Stein's dialectical theory of state and society (Staat-Gesellschaft-Dialektik) and his conception of administration (Verwaltung) as the institutional mediator between state will and social interests, this paper reframes algorithmic governance as a contemporary iteration of Stein's core concern: preventing the capture of state functions by dominant social forces. Where Stein warned that unchecked class interests could colonize administrative power, this paper contends that concentrated ownership of data, models, and computational infrastructure by technology capital poses a structurally analogous risk to public administration today. Explainability-focused accountability regimes, by reducing legitimacy to procedural transparency of decision logic, mirror a narrowly technocratic conception of administration that Stein explicitly rejected. Instead, this paper proposes that administrative legitimacy in the age of AI-ADM requires institutionalized, continuous mechanisms of review—echoing Stein's insistence on the reciprocal feedback between administration and society (Selbstverwaltung)—that allow public institutions to intervene when automated governance drifts from the public interest. Using Taiwan's emerging AI governance framework as a case study, the paper examines how administrative law instruments (judicial review, procedural safeguards, oversight bodies) might be reconstructed to operationalize this Stein-inspired model of reviewable governance. The paper concludes that accountability for automated decision-making should be evaluated not merely by the interpretability of algorithmic outputs, but by the robustness of institutional channels through which state and society can jointly contest and correct administrative power exercised through AI systems.

Keywords: algorithmic accountability; automated decision-making; Lorenz von Stein; administrative law; AI governance; Taiwan

1 INTRODUCTION

The governance of artificial intelligence in public administration has, over the past decade, converged on a single dominant vocabulary: explainability. From the European Union's General Data Protection Regulation and its much-debated “right to explanation,” to the OECD's AI Principles, to the technical standards issued by national metrology institutes, the policy consensus holds that an automated decision becomes accountable to the extent that the reasoning behind it can be rendered intelligible to the person it affects. This paper contends that this consensus, however well-intentioned, rests on a category error. It treats accountability as a property of outputs—the legibility of a model's internal logic—rather than as a property of institutions: the durable channels through which a polity can contest, correct, and if necessary reverse the exercise of public power. Explainability asks whether a decision can be understood.

* Assistant Professor at Department of Law, Hsuan Chuang University, Hsinchu, Taiwan.

Accountability, in the older and more demanding sense the term carries in administrative law, asks whether a decision can be *undone*.

This distinction is not merely semantic. It reflects two different theories of where the danger of automated decision-making actually lies. The explainability paradigm locates the danger in opacity: the automated system is a black box, and the remedy is to open it. The reviewability paradigm advanced in this paper locates the danger elsewhere—in the concentration of the material and epistemic infrastructure of governance (data, compute, proprietary models) in private hands whose interests do not necessarily align with the public interest the administration is charged with executing. On this view, even a perfectly explainable algorithm remains a threat to administrative legitimacy if the institutional mechanisms for contesting and correcting its outputs have atrophied or were never built. Conversely, an opaque model embedded within a robust structure of continuous review, judicial oversight, and participatory correction may pose less of a legitimacy problem than a transparent one operating outside any such structure.

To develop this argument, the paper turns to a theorist largely absent from the algorithmic accountability literature: the nineteenth-century German jurist and social theorist Lorenz von Stein (1815–1890). Stein is remembered, where he is remembered at all in Anglophone scholarship, as a founder of comparative public administration and as an intellectual forerunner of the modern *Rechtsstaat*'s administrative apparatus. His more distinctive contribution, however, was a dialectical theory of the relationship between state (*Staat*) and society (*Gesellschaft*)—one in which administration (*Verwaltung*) occupies the structurally decisive position of mediating between the universal will of the state and the particular, often self-interested, claims of social classes. Stein's abiding worry, developed across his multi-volume *Verwaltungslehre*¹ and his earlier writings on socialism and the social question,² was that administration could be captured: that a dominant social class, through its command of economic resources, could hollow out the state's claim to represent the common good and redirect administrative power toward its own particular ends. Administration remains free, in Stein's framework, only for as long as it retains institutional mechanisms that allow it to sense and respond to the totality of social interests rather than the loudest or best-resourced among them.

This paper's central claim is that Stein's diagnosis, developed to explain the class conflicts of industrializing Europe, describes with uncanny precision the structural position that large technology firms now occupy vis-à-vis the administrative state in the age of AI. The firms that supply foundation models, cloud compute, and the data pipelines on which AI-ADM systems depend are not merely vendors; they are, in Stein's terms, a rising social power whose command of a scarce and decisive resource—computational and epistemic infrastructure—gives them the capacity to shape administrative outcomes in ways that elude ordinary channels of political and legal accountability. An accountability framework built solely around explainability treats this structural relationship as a technical footnote, a matter of interface design, rather than what it is: a question of who controls the means of administrative cognition, and whether the state has retained institutional capacity to correct that control when it drifts from the public interest.

This argument situates the paper within a broader research program concerned with recovering classical and nineteenth-century theories of administration—Stein's account of *Verwaltung* prominent among them—as resources for evaluating the legitimacy of contemporary regulatory and administrative

¹ Lorenz von Stein, *Die Verwaltungslehre*, 8 vols. (Stuttgart: Cotta, 1865–1868).

² Lorenz von Stein, *Geschichte der sozialen Bewegung in Frankreich von 1789 bis auf unsere Tage* (Leipzig: Wigand, 1850).

arrangements that a purely proceduralist or technocratic vocabulary struggles to assess.³ This recovery project runs parallel to, and draws methodological encouragement from, recent efforts to retrieve classical legal-theoretical resources for the assessment of contemporary public law more generally.⁴ Where much of the algorithmic accountability literature treats the problem of AI-ADM as historically unprecedented, requiring correspondingly novel technical and legal instruments, this paper's method is deliberately genealogical: it treats the AI-ADM accountability problem as a contemporary instance of a much older and better-understood problem in the theory of administration, namely the risk that an ostensibly neutral administrative apparatus will be captured by whichever social force controls a scarce and decisive resource, and it looks to the administrative law traditions built to manage that older problem for guidance on how the newer one might be governed. This is not to deny that AI-ADM raises genuinely novel technical questions; it is to insist that the normative question of accountability—who may contest an exercise of public power, and through what institutional channel—is not one of them.

The paper proceeds in seven further parts. Part 2 surveys the explainability-centered discourse on algorithmic accountability and identifies its normative limitations, particularly its tendency to conflate transparency with legitimacy and its neglect of the institutional question of who may act on what is disclosed. Part 3 reconstructs Stein's dialectical theory of state, society, and administration, with particular attention to his concepts of *Selbstverwaltung* (self-administration) and the administrative state's vulnerability to class capture. Part 4 develops the paper's central analogy, arguing that concentrated technological capital reproduces the structural threat Stein associated with concentrated economic capital, and that explainability-based accountability regimes reproduce the narrowly technocratic conception of administration Stein rejected. Part 5 elaborates the alternative this paper proposes: a reviewability principle grounded in continuous, institutionalized feedback between administration and society, modeled on Stein's account of self-administration. Part 6 turns to Taiwan as a case study, examining the country's emerging AI governance architecture—including the newly enacted Artificial Intelligence Basic Act, the National Science and Technology Council's AI governance framework, and existing administrative law instruments—to assess how far current arrangements approximate, and how they might be reconstructed to better approximate, the reviewability model. Part 7 considers objections and limitations. Part 8 concludes.

2 THE EXPLAINABILITY PARADIGM AND ITS DISCONTENTS

The turn to explainability as the organizing principle of algorithmic accountability has a coherent intellectual history. Early critiques of automated decision-making in the 2010s focused on the “black box” problem:⁵ machine learning models, particularly deep neural networks, produce outputs through computational processes too complex for a human to trace step by step, even when the underlying code is fully disclosed. Commentators observed that this opacity was qualitatively different from the opacity of earlier bureaucratic decision-making, where a human official could always, in principle, be asked to state reasons. The proposed remedy followed naturally from the diagnosis: if the problem is that models cannot explain

³ See generally Lorenz von Stein, *Die Verwaltungslehre*, 8 vols. (Stuttgart: Cotta, 1865–1868).

⁴ Adrian Vermeule, *Common Good Constitutionalism: Recovering the Classical Legal Tradition* (Cambridge: Polity Press, 2022).

⁵ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015).

themselves, the solution is to build models that can, or to build supplementary tools that approximate an explanation after the fact.

This diagnosis produced an enormous and genuinely valuable technical research program—counterfactual explanations, feature-attribution methods, model distillation into interpretable surrogates, and so on—alongside a parallel legal literature debating whether instruments such as the GDPR’s Article 22 and Recital 71 in fact create an enforceable right to explanation, and if so, what kind of explanation would satisfy it.⁶ Administrative law scholars extended the inquiry to public-sector ADM specifically, asking whether existing doctrines of reasoned decision-making, developed for human bureaucrats, could or should be adapted to algorithmic ones.

Yet even sympathetic observers have noted the limits of this approach. Three limitations bear directly on the argument of this paper.

First, explainability is frequently unavailable or, when available, unhelpful, in exactly the cases where accountability matters most. High-performing models are often the least amenable to faithful, human-legible explanation, so that jurisdictions face a persistent trade-off between the accuracy of an automated decision and the honesty of the explanation offered for it. Where this trade-off is resolved in favor of interpretable-but-weaker models, one accountability problem is exchanged for a different one: administrations now systematically deploy less accurate tools in order to satisfy an explainability requirement whose connection to the public interest the requirement was meant to serve is, at best, indirect.

Second, explainability, even where technically achievable, does not by itself confer any capacity to act. A citizen who receives a lucid account of why an algorithm denied their benefits application, revoked their license, or flagged their tax return for audit is no better positioned to have that decision reversed unless some further institutional mechanism exists through which the explanation can be contested. Explainability is, in this sense, a necessary but radically insufficient condition for accountability in the administrative-law sense of the term⁷—a sense that has always centered not on whether reasons were given, but on whether an independent body could review those reasons and substitute its own judgment, or compel the original decision-maker to reconsider. The explainability literature has tended to treat this second, institutional question as separate from, and often secondary to, the technical question of interpretability. This paper treats the ordering as reversed: institutional reviewability is the primary accountability variable, and explainability is, at most, an input that makes reviewability more effective once the reviewing institution exists.

Third, and most importantly for this paper’s argument, the explainability paradigm has a specific and underappreciated ideological valence. It locates the site of accountability inside the algorithm itself—in its architecture, its features, its decision boundary—and correspondingly locates the remedy inside the technical apparatus that produced the algorithm: better documentation, better interface design, better disclosure. This has the effect, whether intended or not, of depoliticizing the governance question. It invites regulators to ask “can this system explain itself adequately?” rather than “who controls the infrastructure that makes this system possible, and does the public retain any leverage over them?” The former question can be answered, and often is answered, by the very firms that build and sell the systems in

⁶ Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,” *International Data Privacy Law* 7, no. 2 (2017): 76–99.

⁷ Danielle Keats Citron and Frank Pasquale, “The Scored Society: Due Process for Automated Predictions,” *Washington Law Review* 89, no. 1 (2014): 1–33; Danielle Keats Citron, “Technological Due Process,” *Washington University Law Review* 85, no. 6 (2008): 1249–1313.

question, through voluntary disclosure frameworks, model cards, and algorithmic impact assessments that they themselves author.⁸ The latter question cannot be answered by the firms at all; it can only be answered by public institutions willing to examine the structural position those firms occupy relative to the administrative state.

A further, closely related difficulty deserves separate mention. Much of the explainability literature implicitly assumes a bilateral relationship between the affected citizen and the administration that deploys the algorithm, on the model of the classical due process encounter between an individual and a bureaucrat. This assumption obscures the fact that the party best positioned to explain an AI-ADM system's behavior is frequently neither the citizen nor the administration, but the private vendor that designed, trained, and often continues to host and update the model. Administrations procuring AI systems under standard commercial licensing arrangements not infrequently lack full visibility into training data provenance, model weights, or the precise mechanism by which a given input produces a given output, particularly where the vendor treats these as trade secrets. In such arrangements, an explainability mandate addressed to "the algorithm" in the abstract risks becoming, in practice, a mandate addressed to whatever the vendor is willing to disclose, filtered through the vendor's own commercial incentives to protect proprietary methods and limit liability exposure. This triangulation—citizen, administration, vendor—is largely invisible in accountability frameworks that speak only of "the algorithm" and "the affected person," and it is precisely this triangulation that the Steinian framework developed below is designed to bring into view.

Comparative administrative law scholarship offers a further, complementary observation. Jurisdictions with strong traditions of reasoned administrative decision-making—Germany's *Begründungspflicht*, France's *motivation des actes administratifs*, Taiwan's own statutory requirement under Article 96 of the Administrative Procedure Act that administrative dispositions state their factual and legal basis—already possessed, prior to the emergence of AI-ADM, doctrinal resources considerably more demanding than anything the explainability literature has proposed as a novel remedy. The puzzle this raises, and one this paper returns to in Part 6, is why AI governance discourse has tended to treat explainability as a novel technical challenge requiring novel technical solutions, rather than as an instance of a much older administrative law problem—the adequacy of stated reasons—for which many legal systems already possess doctrine, albeit doctrine developed with human decision-makers in mind and in need of extension rather than wholesale replacement.

It is this third limitation that motivates the turn to Stein. Stein wrote at a moment when a strikingly analogous depoliticizing move was available and was, in fact, being made by contemporaries who treated the "social question" of industrial capitalism as a matter of amelioration—better wages, better working conditions, better relief for the poor—rather than as a structural question about who controlled the means of production and what that control implied for the state's capacity to govern in the name of the whole of society. Stein's intervention was to insist that amelioration without structural attention to control was self-defeating: that an administration which improved the treatment of subordinate classes without addressing the structural capacity of a dominant class to redirect state power toward its own ends had not solved the problem of legitimacy, but merely postponed its recognition. The parallel to contemporary AI governance, this paper argues, is exact. Improving the explanations automated systems provide

⁸ Cary Coglianese and David Lehr, "Transparency and Algorithmic Governance," *Administrative Law Review* 71, no. 1 (2019): 1–56.

without addressing the structural capacity of the firms that build those systems to shape administrative outcomes is amelioration without structural correction.

3 STEIN'S DIALECTICAL THEORY OF STATE AND SOCIETY

Lorenz von Stein's political and administrative theory is best understood as an extended argument about the conditions under which a state can be said to represent something more than the interests of whichever social group currently controls it. Stein's starting point, developed most fully in his early study of French socialist and communist movements and elaborated across the volumes of his later *Verwaltungslehre*, is a distinction between state (*Staat*) and society (*Gesellschaft*). Society, for Stein, is the sphere of particular interests: the realm in which individuals and classes pursue their own material advantage, structured by the prevailing distribution of property and labor. The state, by contrast, is or ought to be the sphere of the universal will—the institutional expression of the interest of the whole community as such, standing above and independent of any particular class interest within society.

The difficulty, which Stein treats as the central problem of modern politics, is that the state does not exist in a vacuum above society; it is staffed, financed, and to a significant degree captured by particular social actors, and above all by whichever class currently holds the dominant position within the prevailing economic order. Stein's analysis of nineteenth-century industrial society identified the emerging propertied bourgeoisie as the class most capable of translating its economic dominance into political dominance, precisely because economic power confers disproportionate access to the instruments of governance: to legislatures, to the press, and—crucially for this paper's purposes—to the administrative apparatus itself. Where a dominant class successfully colonizes the state, Stein argues, the state ceases to perform its distinctive function. It no longer stands above particular interests; it becomes an instrument of one particular interest against the rest of society, while continuing to claim the legitimacy that only representation of the universal interest can confer. Stein terms this condition, in various formulations across his work, a degeneration of the state into an instrument of class rule—a diagnosis that anticipates, and indeed directly influenced, aspects of Marx's own analysis of the state, even as Stein's proposed remedy diverges sharply from Marx's.

Stein's remedy is not revolution but administration. He assigns to *Verwaltung*—the administrative apparatus of the state, as distinct from both the legislature and the monarch or executive head—the decisive role of preventing class capture, a conception later elaborated within the German administrative law tradition itself.⁹ Administration, on Stein's account, is the institutional site at which the abstract universal will of the state is translated into concrete measures affecting particular persons and groups, and it is precisely because administration operates at this point of translation that it has the structural capacity either to correct for class bias or to entrench it. A well-constituted administration, continuously attentive to the actual condition of all social classes rather than only the dominant one, can act as a countervailing force: extending education, regulating labor conditions, redistributing opportunity, and in general using the instruments of public power to prevent any one class's particular interest from permanently capturing the state's claim to represent the whole. A poorly constituted administration—one that has itself been captured, whether through personnel, financing, or dependence on information supplied by the dominant

⁹ Ernst Forsthoff, *Lehrbuch des Verwaltungsrechts*, 10th ed. (Munich: C.H. Beck, 1973).

class—ceases to perform this corrective function and instead becomes an additional instrument of capture.

The concept that operationalizes this corrective function in Stein’s mature work is *Selbstverwaltung*, self-administration. Stein’s use of this term is broader than its later, narrower legal sense of local-government autonomy. For Stein, self-administration names the structural condition under which administration remains continuously responsive to the actual, changing configuration of social interests, rather than freezing at a moment of maximal influence for whichever class happened to be dominant when a given administrative arrangement was established. Self-administration requires institutional mechanisms of feedback: channels through which changes in social condition can register within the administrative apparatus and prompt revision of administrative practice, before those changes harden into the kind of grievance that Stein associated with revolutionary rupture. Administration, in other words, must be reviewable and revisable by the society it governs—not in the sense that society directly commands administration on a case-by-case basis, which would simply invert the capture Stein feared, but in the sense that durable institutional mechanisms exist through which administrative drift away from the universal interest can be detected and corrected before it consolidates into permanent class rule.

Stein’s account of *Selbstverwaltung* is best approached through the distinction he draws between three moments of state activity: the constitutional moment, in which the fundamental form of the state and the relationship among its organs is fixed; the legislative moment, in which the state’s universal will is articulated in general rules; and the administrative moment, in which those general rules are applied to the particular, constantly changing circumstances of individual life. Stein’s insight, distinctive among nineteenth-century state theorists, is that only the third of these moments is structurally capable of tracking the ongoing transformation of social conditions in real time. Constitutions are amended rarely and legislation revised episodically, but administration operates continuously, encountering daily the concrete situations—a family’s poverty, a worker’s injury, a merchant’s failed enterprise—in which the abstract promise of the universal interest either is or is not made good. It follows, for Stein, that administration is not merely the executive tail of the legislative dog, mechanically implementing rules fixed elsewhere, but the organ through which the state’s claim to represent society as a whole is continuously tested against social reality. This is why Stein devotes such disproportionate attention, relative to his contemporaries among state theorists, to the mundane machinery of administration—poor relief, factory inspection, public health, education administration—rather than to constitutional design narrowly conceived: it is in this machinery, not in the constitutional text, that the fate of the state’s claim to universality is actually decided.

This account carries a further implication that bears directly on the present paper’s argument. Because administration is, for Stein, the site at which the universal interest is continuously tested against social reality, the health of administration cannot be assessed by inspecting its formal legal authorization alone. A ministry may be entirely lawful in its formal constitution, entirely faithful to the letter of the legislation it implements, and still have degenerated, in Stein’s sense, into an instrument of a particular class, if the informational and material channels through which it perceives and responds to social reality have themselves been captured. This is precisely the diagnostic move this paper transposes to AI-ADM: a public administration may comply fully with data protection law, procurement law, and whatever AI-specific disclosure statute is in force, and its use of automated decision-making may still have degenerated, in the relevant sense, if the informational and material channels through which the underlying AI system perceives the social field—training data, feature selection, model architecture, update cadence—have been

captured by a vendor whose commercial interests do not track the public interest the administration is charged with serving. Formal legality, on Stein's account as on this paper's extension of it, is a necessary but insufficient condition of administrative legitimacy; the sufficient condition additionally requires that the channels of perception themselves remain open to correction from outside the interest that currently dominates them.

Three features of this framework are worth isolating for the argument that follows. First, Stein's analysis is structural rather than moralistic: the danger he identifies does not depend on any individual administrator or capitalist acting in bad faith. Capture can occur, and typically does occur, through the ordinary operation of institutional dependency—an administration that relies on a dominant class for information, personnel, or resources will tend to reproduce that class's perspective even where every individual official involved believes themselves to be acting impartially. Second, the remedy Stein proposes is institutional rather than technical: it does not consist in better disclosure of what administration is doing, but in durable structures that allow administration's outputs to be contested and revised. Third, and most distinctively, Stein locates the danger of capture not in the formal separation of powers—the classical liberal concern that the executive might usurp legislative or judicial authority—but in the material dependency of administration on the resources controlled by a particular social class, however constitutionally proper the formal channels of administrative action might appear. It is this third feature, in particular, that gives Stein's framework its purchase on the AI governance question, since the risk of capture posed by concentrated technological infrastructure operates precisely through material and epistemic dependency rather than through any formal usurpation of administrative authority.

4 FROM CLASS CAPTURE TO DATA CAPTURE: REFRAMING ALGORITHMIC GOVERNANCE

The central move of this paper is to argue that the structural position Stein assigned to the capitalist class of industrializing Europe is today occupied, with respect to public administration's use of artificial intelligence, by the small number of firms that control the foundation models, cloud computing infrastructure, and large-scale data pipelines on which most AI-ADM systems depend. This is not a claim that technology firms are malicious in the way Stein sometimes suggested industrial capital could be; it is, in keeping with Stein's own structural method, a claim about institutional dependency that holds regardless of the intentions of any particular actor.

Consider the ordinary lifecycle of an AI-ADM system deployed within public administration—for instance, a system used to triage tax audits, allocate social welfare inspections, or support permitting decisions. Few public administrations possess in-house the computational infrastructure, the proprietary training data, or the specialized technical personnel required to build such systems from first principles. The overwhelming tendency, documented across jurisdictions with widely varying administrative traditions, is for public administration to procure these capabilities from a small number of firms capable of supplying them at scale: firms that operate the major cloud platforms, that have developed and trained the leading foundation models, and that increasingly bundle the technical, legal, and even ethical-compliance infrastructure (model documentation, bias audits, explainability tooling) that accompanies deployment.¹⁰ Administration, on this arrangement, becomes structurally dependent on a small set of private actors not merely for a discrete input, as it might depend on a stationery supplier, but for the very capacity

¹⁰ David S. Rubenstein, "Acquiring Ethical AI," *Florida Law Review* 73, no. 4 (2021): 747–819.

to exercise a core function of contemporary governance: the sorting, prioritizing, and adjudicating of claims upon the state.

This dependency reproduces, with striking fidelity, the structure Stein worried about. It is a dependency of *capacity*, not merely of *resources* in the ordinary sense: administration does not simply purchase a tool from these firms, the way a nineteenth-century ministry might purchase paper or ink; it purchases a cognitive apparatus—the very means by which it perceives and processes the claims of the population it governs. Where Stein worried that a dominant economic class could shape the administrative apparatus because that class controlled the material conditions on which the state’s ordinary operation depended (financing, information networks, and eventually direct administrative personnel drawn from its own ranks), the contemporary technology sector occupies an analogous position because it controls the computational and epistemic conditions on which the state’s algorithmic operation depends. The parallel extends to Stein’s account of how capture typically proceeds: not through crude bribery or overt political capture, but through the ordinary, seemingly neutral operation of technical dependency, in which the vendor’s design choices, training data selections, and default configurations come to shape administrative outcomes in ways that are difficult for the administration itself, let alone the public, fully to perceive or interrogate.

Against this backdrop, the limitations of the explainability paradigm identified in Part 2 take on a sharper political meaning. An accountability framework that asks only whether the algorithm can explain its own outputs leaves entirely untouched the antecedent question of who designed the algorithm’s objective function, who selected and curated its training data, and who retains the practical capacity to modify the system once deployed. It is entirely possible—indeed, on current market structures, it is the ordinary case—for a system to satisfy every prevailing explainability standard while the actual power to alter that system’s behavior, to update its weights, or to withdraw it from service rests with a private firm whose contractual relationship with the public administration was negotiated with limited public administrative bargaining power and even less public participation. Explainability, on this reading, functions analogously to the “amelioration without structural correction” that Stein criticized in his contemporaries’ response to the social question: it treats the symptom (the citizen’s inability to understand a given decision) while leaving the underlying structure of dependency (the administration’s inability to control the conditions of its own decision-making capacity) entirely intact.

This structural dependency is compounded by an epistemic dimension that has no close analogue in Stein’s own account and that, if anything, sharpens the concern. Nineteenth-century capital, however dominant, did not control the categories through which administration perceived the social field it governed; ministries retained their own statistical bureaus, their own inspectors, their own independently gathered census and survey data, even where they depended on capital for financing and even where dominant classes contested the interpretation of that data. Contemporary AI-ADM systems, by contrast, frequently supply not only the mechanism of decision but also the very features and categories through which an administrative problem is represented: which variables are treated as predictive of tax fraud, which patterns are treated as indicative of benefit ineligibility, which combinations of factors are treated as elevating a permit application’s risk profile. Where these representational choices are embedded in a proprietary model whose training data and feature engineering the administration cannot independently reconstruct, the dependency Stein worried about extends from the material means of governance to its cognitive preconditions: administration comes to see the social field it governs through categories it did

not choose and often cannot fully inspect. This is, in effect, a deeper form of the capture Stein described, since it operates not merely on the resources available to implement a policy, but on the antecedent perception of what the policy problem consists in.

It is important to be precise about the scope of this claim. The argument is not that technology firms occupy a position identical in every respect to the industrial bourgeoisie Stein analyzed, nor that the remedy must therefore be identical. Stein's own framework was explicitly historical and comparative; he traced how the specific form taken by class capture, and the specific administrative remedies available to counteract it, varied across the different national and constitutional contexts he studied in England, France, and the German states. The claim advanced here is at the level of structure, not of surface detail: that wherever a scarce and decisive input to administrative capacity becomes concentrated in the hands of a narrow set of private actors, the risk Stein identified—administration ceasing to represent the universal interest and instead reproducing the particular interest of the class that controls the decisive input—recurs, and that the appropriate remedy is accordingly structural rather than technical. This reframing has direct implications for how accountability ought to be defined and pursued, developed in the following part.

5 SELBSTVERWALTUNG AND THE REVIEWABILITY PRINCIPLE

If the risk posed by AI-ADM is best understood, following Stein, as a risk of structural capture rather than a risk of insufficient technical transparency, the appropriate remedy follows Stein's own prescription: not better disclosure, but durable institutional mechanisms of self-administration—continuous, reciprocal feedback between administration and the society it governs, capable of detecting and correcting drift away from the universal interest before that drift consolidates into permanent capture. This paper terms the administrative-law operationalization of this idea the *reviewability principle*.

The reviewability principle can be stated in general terms as follows: an AI-ADM system deployed within public administration is accountable not to the extent that its outputs can be explained, but to the extent that its outputs, its training and deployment conditions, and the contractual and institutional relationships underlying its procurement remain subject to continuous, independent, and effective review by institutions capable of compelling correction. This formulation deliberately shifts the locus of accountability from the algorithm to the surrounding institutional architecture, in keeping with Stein's insistence that the health of administration is a function of its institutional relationship to society, not a function of the internal transparency of any single administrative act.

Four institutional components follow from this reformulation, each with a rough correlate in Stein's own account of self-administration and each translatable, as Part 6 will argue, into concrete instruments of administrative law.

The first component is *contestability at the point of decision*: affected persons must retain a meaningful, accessible avenue to challenge an individual automated decision, not merely to receive an explanation of it. This corresponds to Stein's insistence that self-administration requires channels through which particular grievances can register within the administrative apparatus. In administrative-law terms, this maps onto rights of appeal, rights to a human re-determination, and rights to have an automated decision suspended pending review—instruments considerably more demanding than a bare right to explanation,

since they entail an obligation on the part of administration to actually reconsider, not merely to account for, the decision reached.

The second component is *structural auditability independent of the vendor*: administration, or an independent body acting on its behalf, must retain the practical and legal capacity to audit an AI-ADM system's training data, decision logic, and performance across affected populations, without that capacity being mediated or gatekept by the private firm that supplied the system. This corresponds most directly to Stein's structural diagnosis: since the danger identified is one of dependency, the remedy must include institutional capacities that do not themselves depend on the cooperation of the party whose power is being checked. In practice, this requires procurement law to mandate audit rights, data portability, and, where feasible, in-house or third-party technical capacity within administration or its oversight bodies, sufficient to conduct meaningful review without relying exclusively on documentation the vendor itself authors.

The third component is *institutionalized, periodic re-evaluation independent of any individual complaint*: because Stein's concern is with structural drift that may not register as an individually cognizable grievance until it has already consolidated, self-administration requires mechanisms of review that operate on a scheduled or triggered basis, independent of whether any particular affected person has filed a challenge. This corresponds to instruments such as mandatory algorithmic impact assessments conducted at defined intervals, sunset clauses requiring active re-authorization of deployed systems, and standing oversight bodies charged with monitoring aggregate outcomes (disparate impact across demographic or regional lines, drift in system performance over time, and so on) rather than only individual disputes.

The fourth component is *judicial or quasi-judicial capacity to compel correction, not merely to require disclosure*: the reviewing institutions contemplated above must possess remedial powers adequate to Stein's demand that administration remain genuinely revisable, not merely observable. A court or oversight body that can require an administration to explain a system's operation, but cannot order the suspension, modification, or withdrawal of that system when it is found to depart from the public interest, replicates rather than solves the depoliticized, disclosure-centered model this paper criticizes. Genuine reviewability requires that the ultimate finding of a reviewing body carry binding, corrective force.

A brief illustration may make the interaction of these four components more concrete. Suppose a municipal administration deploys a machine-learning system to prioritize building-safety inspections, trained on historical inspection and violation data supplied by the vendor together with the system. Under a purely explainability-centered regime, accountability is satisfied once the vendor publishes a model card describing the features used (building age, prior violations, neighborhood density) and the administration provides, on request, a feature-attribution summary explaining why a given building was flagged. Under the reviewability principle, this disclosure would be a starting point rather than an endpoint. Contestability at the point of decision would require that a building owner who disputes an inspection priority ranking be entitled to request human reconsideration that genuinely revisits the ranking, not merely restates the algorithmic rationale. Structural auditability would require that the municipal administration, or an independent auditor acting on its behalf, be able to examine whether the historical violation data used to train the system itself reflects prior discriminatory enforcement patterns—a question the vendor's own model card, however detailed, has no particular incentive to surface, since it implicates the quality of the vendor's own training data rather than merely the technical fidelity of its model. Periodic independent re-evaluation would require that aggregate inspection outcomes be reviewed on a fixed schedule for disparate impact

across neighborhoods, independent of whether any individual building owner has filed a complaint, since the harm of a systematically skewed inspection regime may never register as a cognizable individual grievance even as it redistributes the burdens and protections of building-safety regulation in ways that track, rather than correct, prior patterns of neglect. Binding corrective remedy, finally, would require that findings from either the individual reconsideration process or the periodic structural review be capable of triggering an enforceable requirement that the system be retrained, reweighted, or withdrawn, rather than merely a public report noting the disparity for future policy consideration. Each component addresses a distinct failure mode that the others leave untouched, which is why the reviewability principle is specified as a conjunction of four elements rather than a single master criterion: a regime strong on contestability but weak on structural auditability will catch individually salient errors while leaving systemic bias in training data undetected; a regime strong on periodic review but weak on binding remedy will detect systemic bias without being able to correct it; and so on.

It bears emphasis that the reviewability principle does not dispense with explainability; it relocates it. Explainability remains valuable, on this account, primarily as an evidentiary input that makes the four components above more effective—an audit is easier to conduct, a periodic re-evaluation more meaningful, and a judicial remedy more precisely targeted, when the system under review can articulate its own reasoning. The paper’s claim is one of priority and sequencing, not of substitution: explainability without institutionalized reviewability is inert, while reviewability substantially strengthens whatever explainability a system can offer. The following part examines how far Taiwan’s emerging AI governance framework approximates this model, and what a Stein-inspired reconstruction of its administrative law instruments might look like.

6 TAIWAN’S AI GOVERNANCE FRAMEWORK: A CASE STUDY

Taiwan offers a particularly instructive case study for the reviewability model developed above, for three reasons. First, Taiwan’s administrative law tradition, drawing heavily on German and Japanese administrative law scholarship, already possesses many of the doctrinal building blocks—judicial review of administrative disposition (行政處分), procedural rights under the Administrative Procedure Act (行政程序法), and a specialized administrative court system—on which a reviewability-centered AI governance regime could be constructed without wholesale institutional innovation.¹¹ Second, Taiwan’s AI policy is at a formative stage: Taiwan has enacted its first dedicated AI statute, the Artificial Intelligence Basic Act, passed by the Legislative Yuan on 23 December 2025 and promulgated on 14 January 2026,¹² which designates the National Science and Technology Council (NSTC), successor to the former Ministry of Science and Technology, as the central competent authority for AI governance and calls for an internationally aligned AI risk classification framework. Because the Act is deliberately framed as a skeleton of

¹¹ Administrative Procedure Act [行政程序法] (Taiwan), amended Jan. 20, 2021, <https://law.moj.gov.tw/ENG/Law-Class/LawAll.aspx?pcode=A0030055>; Administrative Appeal Act [訴願法] (Taiwan), amended June 27, 2012, <https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?pcode=A0030020>; Administrative Litigation Act [行政訴訟法] (Taiwan), <https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=A0030154>, all accessed September 17, 2026.

¹² Artificial Intelligence Basic Act [人工智慧基本法] (Taiwan), promulgated Jan. 14, 2026, <https://law.nstc.gov.tw/LawContent.aspx?id=GL000592>; “The Artificial Intelligence Basic Act Is Promulgated and Enters into Force,” Lexology, January 15, 2026, <https://www.lexology.com/library/detail.aspx?g=6cc0ad64-a156-4bb1-9e1e-0bc3d961894f>; “Taiwan’s AI Basic Act: A Different Logic of Democratic AI Governance,” MediaLaws, July 6, 2026, <https://www.medialaws.eu/taiwans-ai-basic-act-a-different-logic-of-democratic-ai-governance/>.

seven governance principles rather than a set of binding obligations, and delegates risk-classification and enforcement detail to sector-specific competent authorities over a two-year implementation window, the choice between an explainability-centered and a reviewability-centered model of implementation remains genuinely open at the level that matters for this paper's argument, rather than already locked in by legacy regulation. Third, Taiwan's economic position within the global semiconductor and AI supply chain gives the dependency dynamic described in Part 4 a distinctive local inflection: Taiwanese public administration is simultaneously a downstream consumer of foundation models and cloud infrastructure supplied predominantly by non-domestic firms, and an upstream supplier, through its semiconductor industry, of the computational hardware on which those same firms depend—a structural position that complicates, without dissolving, the capture dynamic this paper has described.

Taiwan's evolving AI governance architecture has, to date, leaned heavily on instruments recognizable from the explainability paradigm. Policy documents and draft legislative frameworks emphasize risk-tiered classification of AI systems (echoing the EU AI Act's structure), disclosure obligations for high-risk applications, and the establishment of an AI safety or governance institute tasked with technical evaluation of models. These instruments are valuable and this paper does not counsel their abandonment. But considered against the four components of the reviewability principle set out in Part 5, current and proposed Taiwanese arrangements remain heavily weighted toward disclosure and technical evaluation, and comparatively underdeveloped with respect to contestability, independent auditability, periodic structural re-evaluation, and binding corrective remedy specific to public-sector AI-ADM.

Consider contestability at the point of decision. Taiwan's Administrative Procedure Act already establishes a general right to a statement of reasons for administrative dispositions and a general framework for administrative appeal (訴願) followed by administrative litigation before the administrative courts.¹³ These general instruments extend, in principle, to dispositions informed or produced by automated systems, since Taiwanese administrative law does not currently exempt automated decisions from the ordinary requirements governing administrative dispositions. The practical difficulty is that these general instruments were designed for a world in which the reasons an administrator gives for a decision are also the reasons that actually produced it—a correspondence that AI-ADM systems, particularly those built on complex machine learning models procured from external vendors, cannot be assumed to preserve. A reviewability-consistent reconstruction would require Taiwan's administrative procedure framework to be supplemented with AI-specific provisions clarifying that a citizen's right to a statement of reasons under Article 96 of the Administrative Procedure Act extends to a right to know that a decision was substantially informed by an automated system, coupled with a right to request human re-determination that is not satisfied merely by an official's post hoc ratification of the algorithmic output.¹⁴

Consider next structural auditability independent of the vendor. Here the gap between current arrangements and the reviewability model is most pronounced. Public procurement of AI systems in Taiwan, as in most jurisdictions, proceeds through the Government Procurement Act, whose provisions were drafted with conventional goods and services in mind and do not, absent specific amendment or

¹³ Administrative Appeal Act [訴願法] (Taiwan), <https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?pcode=A0030020>; Administrative Litigation Act [行政訴訟法] (Taiwan), art. 4, <https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=A0030154>, both accessed September 17, 2026.

¹⁴ Administrative Procedure Act [行政程序法] (Taiwan), art. 96, <https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=A0030055>, accessed September 17, 2026.

contractual innovation, mandate the kind of ongoing audit access, training-data disclosure, or model documentation rights that structural auditability requires.¹⁵ A Stein-inspired reconstruction would treat procurement law itself as a central site of administrative-law reform: government contracts for AI-ADM systems should, as a matter of mandatory procurement policy rather than case-by-case negotiation, include standing audit rights exercisable by an independent oversight body (discussed further below), data and model portability provisions that prevent vendor lock-in from becoming a de facto veto over administrative correction, and contractual terms that subordinate vendor intellectual property claims to the administration's overriding need to demonstrate, to a reviewing court, the actual basis on which a contested decision was reached.

Consider periodic, independent re-evaluation. Taiwan's draft AI governance instruments have moved some distance toward this component through proposals for risk classification and periodic technical assessment of high-risk systems, plausibly to be housed within a National Institute of Cyber Security or an analogous technical body.¹⁶ The reviewability model developed here would extend this instinct specifically to public-sector AI-ADM by proposing a standing administrative body—analogue in function, though not necessarily in form, to the Control Yuan's existing oversight jurisdiction over administrative maladministration—charged with conducting mandatory, scheduled reviews of deployed AI-ADM systems' aggregate outcomes, independent of individual complaints, with findings reportable to the Legislative Yuan and directly actionable by the administrative courts.

Consider, finally, binding corrective remedy. Taiwan's administrative courts already possess, under the Administrative Litigation Act, the power to annul unlawful administrative dispositions and, in appropriate cases, to order administration to make a specific decision or to reconsider within a specified framework. These remedial powers are, in principle, adequate to the reviewability model's fourth component; the reconstruction required here is less a matter of new remedial authority than of ensuring that the courts' existing remedial powers are not frustrated in practice by the technical opacity or vendor-controlled documentation that currently surrounds AI-ADM systems—which returns the analysis to the auditability component discussed above. In other words, Taiwan's existing administrative law doctrine supplies much of the remedial architecture the reviewability principle requires; what is missing is the evidentiary and structural infrastructure—independent audit capacity, procurement-law-mandated transparency to the reviewing court rather than merely to the administration—that would allow that existing architecture to function against AI-ADM systems as effectively as it functions against conventional administrative action.

It is also worth situating Taiwan's approach comparatively against the European Union's AI Act,¹⁷ which has functioned as the de facto international template for much recent AI legislation, including elements of Taiwan's own newly enacted Artificial Intelligence Basic Act. The EU AI Act's risk-tiered structure—prohibiting certain uses outright, subjecting “high-risk” systems to conformity assessment, documentation, and human oversight requirements, and leaving lower-risk systems to voluntary codes of

¹⁵ Government Procurement Act [政府採購法] (Taiwan), promulgated May 27, 1998, <https://law.moj.gov.tw/ENG/Law-Class/LawAll.aspx?pcode=A0030057>, accessed September 17, 2026.

¹⁶ National Institute of Cyber Security, “About Us,” <https://www.nics.nat.gov.tw/en/about/introduction/>, accessed September 17, 2026; Alex Myslinski, “The Role of Taiwan's National Institute of Cyber Security,” *Taiwan Business TOPICS*, February 10, 2025, <https://topics.amcham.com.tw/2025/02/the-role-of-taiwans-national-institute-of-cyber-security/>.

¹⁷ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, 2024 O.J. (L 1689).

conduct—is, in the vocabulary developed in this paper, a paradigmatically explainability- and disclosure-centered instrument. Conformity assessment asks whether a system meets a technical standard and is documented as doing so; it does not, of its own force, create the individualized contestability, vendor-independent auditability, periodic re-evaluation, or binding corrective remedy that the reviewability principle requires, although national implementing legislation in EU member states may supply some of these elements through pre-existing administrative law doctrine, much as this paper argues Taiwan's existing doctrine could be extended to do. The risk for Taiwan, in adopting elements of the EU AI Act's structure without a parallel, deliberate integration with its own administrative law tradition, is that the country ends up with a technically sophisticated risk-classification and disclosure regime running alongside, rather than through, the administrative law instruments—the Administrative Procedure Act, the Government Procurement Act, the Control Yuan, the administrative courts—that Stein's framework suggests are the more decisive site of accountability. A reviewability-conscious approach to implementing Taiwan's now-enacted Artificial Intelligence Basic Act would treat the EU-style risk classification called for under the Act's implementing framework as a floor for technical documentation, while making explicit, in the subordinate regulations that sector-specific competent authorities are to issue within the Act's two-year implementation window, that classification as high-risk triggers not only disclosure and conformity-assessment obligations but also the specific contestability, auditability, and remedial mechanisms set out in Part 5, cross-referenced directly to the relevant provisions of the Administrative Procedure Act, the Government Procurement Act, and the Administrative Litigation Act, so that the AI-specific statute functions as an integrating amendment to existing administrative law rather than a freestanding parallel track.

The institutional design of any standing oversight body also merits closer attention than existing proposals have generally given it. A body modeled loosely on the Control Yuan's investigatory jurisdiction, but purpose-built for AI-ADM oversight, would need, to satisfy the reviewability principle's third component, statutory authority to initiate review of a deployed system independent of any individual complaint, a technical staff or a standing arrangement with independent technical auditors not employed by or contracted through the system's vendor, and a statutory obligation to report findings in a form usable by the administrative courts in subsequent litigation under the Administrative Litigation Act. Absent this last feature, an oversight body's findings risk remaining advisory in the same sense that vendor-authored algorithmic impact assessments are currently advisory: informative, but without the binding corrective force the reviewability principle's fourth component demands. Taiwan's Control Yuan already possesses, under the Constitution and the Control Act, corrective measure (糾正) and impeachment powers directed at maladministration; extending an analogous, AI-specific corrective measure power—paired with a formal channel connecting Control Yuan findings to standing to sue in the administrative courts—would substantially close the gap between Taiwan's existing oversight architecture and the reviewability model without requiring a wholly new constitutional organ.¹⁸

Taken together, this case study suggests that the reconstruction Taiwan's AI governance framework requires is less a matter of importing entirely new doctrines than of extending, deliberately and self-consciously, existing administrative law instruments—the Administrative Procedure Act's reasons requirement, the Government Procurement Act's contractual architecture, the Control Yuan's oversight

¹⁸ Constitution of the Republic of China, art. 97; Control Act [監察法] (Taiwan); Control Yuan, "Impeachment," <https://www.cy.gov.tw/en/cp.aspx?n=241>, accessed September 17, 2026.

jurisdiction, and the Administrative Litigation Act's remedial powers—into a domain (AI-ADM) that current drafting has left to a parallel, disclosure-centered track modeled on the EU AI Act. The Stein-inspired argument of this paper suggests that this parallel track, however useful as a technical risk-management framework, cannot substitute for the extension of ordinary administrative law's contestability and remedial machinery to automated decisions, precisely because it is that ordinary machinery, and not technical disclosure standards, that historically has performed the function Stein assigned to self-administration: keeping administrative power answerable to the whole of the society it governs, rather than to the particular interests, however well-documented, of whichever actor currently controls the means of administrative cognition.

7 OBJECTIONS AND LIMITATIONS

Four objections merit consideration. The first is that the analogy between technology firms and Stein's industrial bourgeoisie is overdrawn: technology firms, unlike a nineteenth-century propertied class, do not seek direct political representation and are, in many jurisdictions, subject to competition law, data protection law, and sector-specific regulation that Stein's industrial capitalists never faced. This objection has force but does not undermine the paper's structural claim. Stein's own analysis did not require that a capturing class seek direct political representation; the danger he identified operated equally, and in his view more insidiously, through administration's ordinary dependency on that class's resources, independent of any deliberate political strategy on the class's part. The existence of competition and data protection law is precisely analogous to the "amelioration" Stein criticized: valuable correctives to particular abuses that leave the underlying structural dependency intact, since neither competition law nor data protection law is designed to, or does, guarantee public administration independent capacity to audit and correct the AI systems it procures.

The second objection concerns feasibility: the reviewability principle's demand for independent, in-house or third-party technical audit capacity may exceed the fiscal and human-capital resources of many public administrations, particularly outside the small number of jurisdictions with substantial domestic technical sectors. This is a genuine limitation, and the paper does not claim that reviewability can be achieved costlessly. It suggests, however, that the appropriate response is regional or international pooling of independent audit capacity—an approach with precedent in other domains of technical regulatory capacity, such as pharmaceutical and nuclear safety regulation¹⁹—rather than acceptance of vendor-authored disclosure as a substitute for independent review. Taiwan's position within the global semiconductor supply chain, noted in Part 6, may in fact afford it comparative technical capacity to develop or participate in such pooled review infrastructure, a possibility beyond this paper's scope but worth flagging for future research.

The third objection is theoretical rather than practical: critics sympathetic to Stein's broader corpus may object that this paper's use of *Selbstverwaltung* strips the concept of its original institutional referent (nineteenth-century local self-government) and repurposes it as a general metaphor for institutional feedback, risking anachronism. The paper accepts that its use of Stein is reconstructive rather than exegetical:

¹⁹ International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, "Mission," <https://www.ich.org>, accessed September 17, 2026; International Atomic Energy Agency, "Nuclear Harmonization and Standardization Initiative," <https://nucleus.iaea.org/sites/smr/SitePages/Nuclear-Harmonization-and-Standardization-Initiative.aspx>, accessed September 17, 2026.

it does not claim that Stein anticipated artificial intelligence, but that the structural logic of his diagnosis—administration’s vulnerability to capture by whichever social force controls a scarce and decisive resource, and the remedy of institutionalized, continuous reviewability—retains explanatory and normative purchase when that scarce resource is computational and epistemic rather than industrial and financial. This reconstructive use of historical theory to illuminate a contemporary institutional problem follows a well-established mode of argument in comparative administrative law and public law theory more broadly, and the paper’s claims should be assessed on the strength of the structural analogy, not on fidelity to every detail of Stein’s nineteenth-century institutional referents.

A fourth objection concerns the risk of regulatory overreach: demanding vendor-independent auditability and binding corrective remedy for all AI-ADM systems, however minor their impact, could impose compliance burdens disproportionate to the interests at stake and could deter beneficial adoption of AI tools that improve administrative efficiency without materially affecting individual rights. This objection is well taken as a caution against indiscriminate application, but it does not tell against the reviewability principle as such; it tells in favor of calibrating the principle’s four components to the stakes of the decision at issue, much as existing administrative law already calibrates procedural protection to the severity of the interest affected—compare, for instance, the more exacting procedural requirements Taiwanese administrative law attaches to dispositions affecting personal liberty or livelihood against the lighter requirements attaching to routine administrative convenience. A calibrated version of the reviewability principle would reserve the full complement of contestability, independent auditability, periodic re-evaluation, and binding remedy for AI-ADM systems materially affecting individual rights or the equitable distribution of public benefits and burdens—systems of exactly the kind the EU AI Act designates as “high-risk,” and that the risk-classification framework mandated under Taiwan’s newly enacted Artificial Intelligence Basic Act is intended to capture²⁰—while permitting lighter-touch, primarily disclosure-based governance for lower-stakes administrative automation. This calibration is fully consistent with Stein’s own method, which nowhere suggests that every administrative act requires the same intensity of self-administrative feedback, only that the intensity of feedback should track the potential for a given administrative function to consolidate capture by a particular interest against the interest of the whole.

8 CONCLUSION

This paper has argued that the accountability of AI-driven automated decision-making in public administration should be evaluated not by the interpretability of algorithmic outputs but by the robustness of the institutional channels through which state and society can jointly contest and correct administrative power exercised through AI systems. Drawing on Lorenz von Stein’s dialectical theory of state and society, the paper has argued that the concentration of data, models, and computational infrastructure in the hands of a small number of technology firms reproduces the structural risk Stein associated with class capture of the administrative state, and that the explainability paradigm currently dominant in AI governance discourse reproduces the narrowly technocratic, disclosure-centered response to that risk that Stein explicitly rejected in favor of institutionalized self-administration. The paper has proposed a reviewability principle, comprising contestability at the point of decision, structural auditability independent of the vendor, periodic independent re-evaluation, and binding corrective remedy, as the administrative-law

²⁰ Artificial Intelligence Basic Act [人工智慧基本法] (Taiwan), art. 16; Regulation (EU) 2024/1689 (cited above at note 17).

operationalization of Stein’s demand for continuous, institutionalized feedback between administration and society. Taiwan’s emerging AI governance framework, examined as a case study, illustrates both the promise and the current limitations of this approach: substantial existing administrative law infrastructure—the Administrative Procedure Act, the Government Procurement Act, the Control Yuan’s oversight jurisdiction, and the Administrative Litigation Act—supplies much of what a reviewability-centered regime requires, but current AI-specific policy has, to date, developed largely along a parallel, disclosure-centered track that has not yet been deliberately integrated with this existing architecture.

The broader implication for algorithmic accountability scholarship is that the field’s heavy investment in the technical and legal apparatus of explainability, while not misguided, has been incomplete in a specific and remediable way: it has addressed the question of whether automated decisions can be understood while leaving comparatively underdeveloped the older and more demanding administrative-law question of whether they can be undone, and by whom. Recovering this second question, and the institutional resources of administrative law traditions—Taiwan’s among them—that were built to answer it, offers a more secure foundation for algorithmic accountability than any refinement of explanation-generating technique is likely to provide on its own. Future work might extend this framework comparatively, examining whether jurisdictions with differently structured administrative law traditions and differently structured relationships to global technology infrastructure arrive at correspondingly different institutional solutions to the capture risk this paper has identified, while confronting the same underlying structural problem Stein diagnosed a century and a half ago in a very different technological register.

* * *