
This is the **published version** of the bachelor thesis:

García Villagrasa, Pablo. *Artificial Intelligence for Public Policy Decision-Making : An Analysis of Catalan Rental Subsidies*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Grau en Intel·ligència Artificial)

This version is available at <https://ddd.uab.cat/record/336317>

under the terms of the  license.

Artificial Intelligence for Public Policy Decision-Making: An Analysis of Catalan Rental Subsidies

Pablo García Villagrasa

Abstract

The automation of public policies using algorithmic systems, from point-based rules to Machine Learning models, has become a defining characteristic of modern governance. This work analyses how those systems are automated in the Spanish and Catalan context and how they can be audited, combining a literature review of Automated Decision-Making, three key informant interviews and an embedded case study of the scoring system used by the Agència de l'Habitatge de Catalunya to allocate a youth rental grant, applying an adversarial algorithmic auditing posture adapted from the Eticas Foundation guide. It argues that public ADMs must be studied case by case and that the main governance gap lies before deployment. Eligibility and not the score was the decisive threshold of the case and four minor undocumented discrepancies between published rules and implementation were found, with no significant gaps across income or origin groups.

Keywords: Automated decision-making, public administration, algorithmic auditing, AI Adoption, deterministic systems, housing policy, ecological inference.

1 INTRODUCTION

Public policies' automation through algorithmic systems has consolidated as a structural characteristic of modern governance, including a wide range of tools that go from simple point-based mechanisms to complex Machine Learning (ML) approaches [1], particularly in the form of last half century welfare state [2]. Spanish and Catalan public administration are not an exemption, and different automation cases and problematics derived from them have surfaced to the public opinion last years. Such is the case of BOSCO, a system designed for eligibility for the electric social bonus that was revealed it had flaws [3]; VioGén, that estimates risk levels in gender violence cases and has caused controversy regarding its efficiency and transparency [4] or RisCanvi, used in the Catalan prisons to assess recidivism risk [5].

These systems are usually presented as a more objective, consistent and efficient alternative to human judgement. However, recent criticism has raised concerns from Science and Technology Studies (STS), social players and civil society organizations. Among other diverse algorithm audits, for example of social platforms, whose recommendation systems have themselves caused controversy from

the amplification of harmful content [6] to addictive design [7], literature has documented how the introduction of Automated Decision-Making (ADM) processes in public policies does not neutralize social inequalities but often reproduces or intensifies them. From an STS view, this is defended as a consequence of treating socio-technical problems as just technocratic ones.

This work follows this critical path and aims to transfer to the Spanish and Catalan context. Its objective is to analyze how certain public policies are automatised in that context, what effects do they produce over affected populations and how do the ones in charge of the management of the systems perceive it. For that purpose two methodologic strategies are combined:

- Literature review and semi structured interviews: interviewing relevant ecosystem actors in order to reconstruct the logics behind the construction, deployment and transparency of these systems, preceded by a consequent literature review in order to conceptualize the interviews addressing gaps and problems academically referred.
- Case study: centered around a deterministic scoring system used by the Catalan public administration.

The overarching aim is: how is AI and the automation of public policy taking shape in Catalan and Spanish public administration, and what does it imply for the social consequences of these systems? This breaks down into three research questions, each answered primarily by one of the three methodological strands (Figure 1):

• Contact E-mail: 1630353@uab.cat
 • Supervised by: Núria Vallès Peris (UAB, IIIA-CSIC)
 • Academic Year 2025/26

- What is understood by “AI” in public policy, and how is it audited?
- How is it being implemented and audited in Catalonia and Spain?
- Through a case study, how can such a system be audited and what does that reveal?

The three questions are interrelated rather than parallel: First research question produces, from the literature, the working definition of AI/ADM and the audit vocabulary (the code-driven/data-driven gradient, the auditability notions, the audit-method typology) that the second and third questions then apply; The second one contextualizes the case within actual administration practice; and the last one puts the audit vocabulary to work on a concrete system and feeds back into the literature by testing how far AI-derived methods travel. Prospective implications for future AI deployments are treated in the Discussion, not as a research question.

2 OBJECTIVES

The general objective of the project is to analyze how automation affects and conforms public administration’s decision making in Spain and Catalonia, what role does AI play in that environment and how can those systems be audited, addressing also the challenges this entails. This general objective is broken down into three specific objectives, that correspond to methodology strategies described in section 3:

2.1 Conceptual mapping of state of the art

Synthesizing academic literature, institutional documents and relevant reports from civil society organizations to situate the debate about automation, identifying documented tensions and building a conceptual and empirical base that justify the scripts for the interviews.

2.2 Reconstructing institutional logic and criticism through key actor interview

Performing semi-structured interviews to three informants that represent three different parts of the Catalan ecosystem (administration, open governance/research and civil society) to identify how supervision, governance and transparency challenges of deployed algorithms are perceived from inside.

2.3 Study Case

Applying the conceptual categories and perspectives applied to a specific case as an illustrative and contrasting element: the grant scale for rent payment for young people from the Housing Agency of Catalonia.

All three objectives feed back one into each other: the literature review guides the interviews, the interviews helps contextualizing the case, and the case findings aim to inform the final discussion on the applicability of the reviewed analytical frameworks.

3 METHODOLOGY

3.1 Overall design

The followed methodology combines bibliographic review, semi-structured interviews and the case study in such a way that all three mutually contribute to each other.

3.2 Literature review

The literature review was performed for the dual purpose of delimiting the conceptual framework and preparing the interview script, and its outcome is laid out in full in section 4 (Context and State of the Art). The search parameters have been organized around four thematic axes:

- Automated Decision-Making and algorithmic governance in public administration
- Methodologies for auditing AI systems
- Fairness and biases in deterministic and machine learning systems
- Spanish, Catalan, and European legal and institutional contexts.

The selected sources include articles indexed in Web of Science and Scopus, chapters from reference manuals (like [8]), reports from official bodies [9, 10, 11], and publications from civil society with documented impact (Civio, Eticas Foundation).

As said, the review objective is not accumulating bibliography for the theoretical framing but to build an argued map to formulate informed questions during the interviews. The resulting analytical categories such as the code-driven/data-driven distinction, Li and Goel’s [12] typology of audits, notions of auditability, fairness in scoring systems, and the AI Act’s regulatory framework; will be used both to structure the interview guide and to interpret the responses.

3.3 Key informant interviews

Sampling logic. Three semi-structured interviews were conducted with informants selected under a key informant, intentional maximum-variation logic. The aim is not statistical representativeness nor thematic saturation, but covering the three structurally distinct positions from which deployed public algorithms are observed in Catalonia, mirroring the government, observatory and civil society triad present in the adversarial audit literature:

- I1: a researcher and former senior public official in the fields of citizen participation and public administration training (former Director General de Participació Ciutadana and former Director of the Escola d’Administració Pública de Catalunya), bringing the institutional-academic insider perspective on digital transformation and policy design.
- I2: a jurist at a Catalan observatory on ethics in artificial intelligence and co-author of an institutional impact-assessment model, bringing the legal-regulatory observatory perspective.

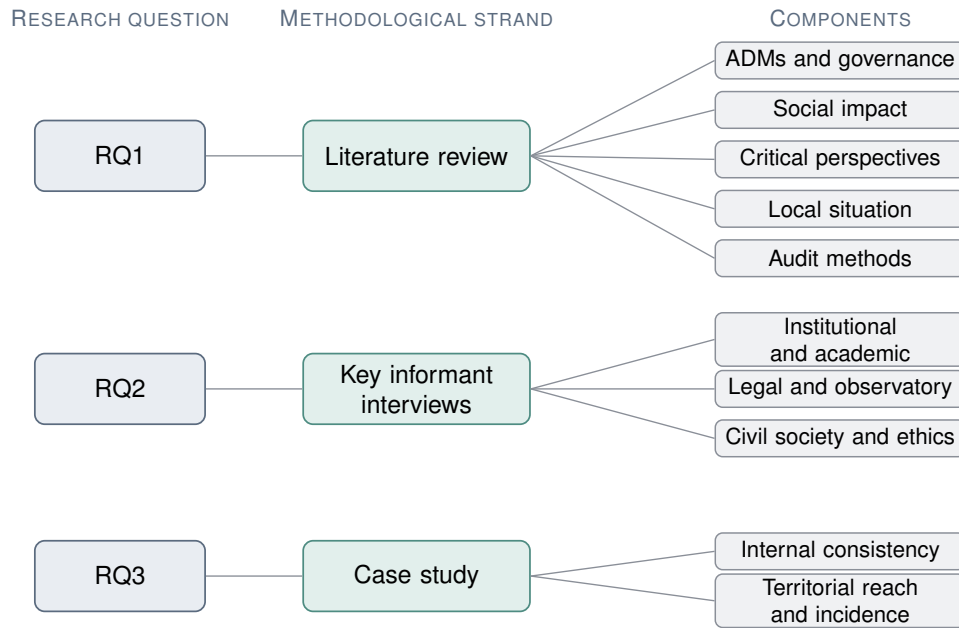


Fig. 1: Research questions, the methodological strand that answers each, and the components that each strand develops. The three strands are interrelated: the review supplies the categories the interviews and the case apply, the interviews contextualize the case, and the case tests how far the audit methods apply.

- I3: a member of civil-society and digital-rights organisations and of the Generalitat's data ethics advisory structures, bringing the civil society and digital rights perspective.

The three informants are referred to by the codes I1, I2 and I3 in the analysis and triangulation that follow, and are identified only by position in order to preserve their confidentiality.

Procedure. Interviews were conducted in Catalan, recorded with informed consent (UAB CER023 template adapted with an explicit opt-in checkbox for attribution by name and position), transcribed locally with Whisper, and analysed in three complementary formats: per-interviewee executive summaries, thematic coding across cross-cutting axes, and a triangulation table crossing each empirical case finding against each informant's contributions. Generative AI tools were used, always as an assistance tool, to translate the Whisper outputs and personal interview notes from Catalan, as well as for formal language and style correction of the document.

Limitations. (1) The interview with the Director of AI of the Generalitat de Catalunya was cancelled by the interviewee on the scheduled day and could not be replaced within the timeline. This leaves a gap in the administrative-technical institutional voice, partially mitigated by I1's past institutional roles. (2) The design does not include beneficiaries or affected parties of the studied system. This is assumed as a scope limit and discussed in the Discussion. (3) The final panel may carry a selection bias toward voices critical of current deployment practices. The analysis treats convergences between informants as more robust than individual claims for this reason.

3.4 Embedded case study

Object. The case study is focused on the criteria that the Housing Agency of Catalonia applies to the Grant for the Payment of Rent or Assignment Price for Housing or Room for Young People. The selection of the case is justified for two reasons: first, its material impact on a group directly affected by the housing crisis, which, according to the latest barometers, is the main concern of the citizens [13, 14]; second, the accessibility of the decision logic and results access: unlike opaque systems such as EuroCop's Pred-Crime or VioGén, the regulatory bases of the scoring system are published in the DOGC, which allows for its formalization and reconstruction.

The empirical analysis of the case has already been performed, with the findings described in section 5. It is intended to be a supporting component for the conceptual framework and interviews and its function is to illustrate how the usual assumptions of an adversarial audit perform when contrasted with a real life scenario.

Audit technique and its adaptation. The case is read through the method-selection logic of the Eticas adversarial guide [15], whose decision tree routes an audit to a technique according to the access available to the auditor. The resolved outputs of the system, the scores and the final resolution of each case, were obtained as microdata. This is precisely the entry condition for the output branch of the tree, the one keyed on access to publicly disclosed outputs such as risk scores, which points to a comparative output audit. The guide defines that technique as the comparison of an algorithm's predicted outcomes against the actual outcomes, or of one system against another or a benchmark, as in the Eticas audit of VioGén, where risk scores are contrasted with realised outcomes.

The barem, however, does not let that comparison close, and the reason is itself one of the results. A comparative output audit needs an outcome against which the output can

be read, and the *barem* produces none: the score is not a prediction that some later event could confirm or refute, and, as the analysis shows (Finding 2), it is not the mechanism that allocates the aid either, so there is no realised outcome it is meant to track. The output exists but has no referent to compare it to. The technique was therefore adapted rather than applied as written. What remains feasible and informative on a fully disclosed deterministic rule is, first, a verification of the published rule against the released scores, an internal consistency audit that occupies the extreme code-driven end of Hildebrandt's gradient [16] and recovers what a code audit would, given that the rule is public; and second, an ecological analysis of the territorial reach and incidence of the aid, which moves the question from the inert score to where the distributional effect actually lands. This second layer was originally conceived as a comparative machine-learning model replicating the assignment, an idea abandoned once Finding 2 made clear that any model trained on the same data and target would only approximate an inert scoring rule, so that the instrument retained from that toolbox is the ecological-inference apparatus rather than a learned replica. The two layers below are thus an adaptation of the adversarial guide to a deterministic, reproducible system, keeping its process posture, namely restricted access, post-deployment focus and stakeholder context, while departing from the single named technique its access conditions would nominally select.

4 CONTEXT AND STATE OF THE ART

4.1 Automated decision-making and algorithmic governance

Automated Decision-Making (ADM) covers a wide spectrum of systems, from deterministic hard-coded rules to predictive models trained over great data volume. Some authors, like Hildebrandt [16], propose a distinction between code-driven regulation where rules are implemented as auto-executable code, building algorithms from legal or administrative criteria and data-driven regulation where the way the system behaves is directly inferred from historical data and patterns. Hildebrandt remarks this distinction because of the transparency and revision challenges each one faces: code-driven systems could be audited through the plain revision of the rules, while data-driven require additional techniques for interpretation. However there's a general understanding that those code-driven systems have not less powerful consequences [1] and audit guidelines treats both Hildebrandt definitions (what could be translated as deterministic vs. Machine Learning) inside the same auditing paradigm.

From STS it is argued that algorithmic systems cannot be understood as isolated tools but as socio-technical artifacts integrated in complex cycles, in each of which political decisions are embedded that cannot be reduced to technical problems. This work adopts two classic anchors of that tradition. Winner [17] argues that artifacts have politics: design choices such as which variables count, what weight they carry, where a threshold is drawn, and that they embody and stabilize social arrangements. Callon [18] describes how the definition of a problem (problematization) and the enrolment of actors around it shape what the result-

ing device can and cannot do; in the case at hand, a scoring rubric designed without an explicit allocation problem to solve becomes a device that scores but does not decide.

4.2 AI-based versus non-AI automated decision-making

Not every automated decision system is an AI system, and the distinction is consequential for both analysis and regulation. Building on the code-driven/data-driven gradient [16], the operative line is whether the system infers how to produce its output from data or merely executes rules fixed in advance by humans. The AI Act adopts exactly this criterion: its definition of an AI system turns on inference, autonomy and adaptiveness, and its recitals explicitly exclude systems based on rules defined solely by natural persons to execute operations automatically [19]. It is important to make this distinction clear because even though everything falls inside algorithmic audits field, many projects worded as AI end up being simpler algorithms, what has important implications towards its openness possibilities. A purely deterministic rubric such as the *barem* studied here therefore falls outside the legal notion of AI, even though it performs the same function, deciding eligibility for an essential benefit, that the same regulation classifies as high-risk when carried out by a learned model. Legal scholars have however also criticised this demarcation, noting that complex rule-based systems can produce the very risks (opacity, lack of contestability, discriminatory effects) the Act is meant to govern [20]. This work takes that tension as a premise: the relevant class for auditing is automated decision-making and algorithmic audit as a whole, of which AI is one technical sub-case, and the deterministic remainder is extended in administrative practice.

4.3 Automation beyond decision-making

Automation in public administration is not confined to systems that decide individual cases. A growing body of work in information systems studies the automation of routine administrative processes through Robotic Process Automation (RPA) and related service-automation technologies, which reproduce clerical workflows without any learned model and are increasingly adopted in local and national administrations [21, 22]. In parallel, generative AI and large language model assistants are being introduced not as decision engines but as augmentation tools, for drafting, summarising and front-office support under a human-in-the-loop principle, alongside shared data infrastructures, data pools and federated systems that pool information across organisations [23]. These modes matter here for two reasons. First, they relocate AI away from the decision rule itself, so that the question of where automation actually bears on citizens cannot be answered by inspecting a single scoring artifact. Second, the line between aid and substitution, which the interviews repeatedly foreground, is precisely what separates the contested deployments (automating a decision) from the productive ones (supporting a professional who still decides). However, the use of these systems as professional aid also raises problems such as privacy concerns: during the interviews, examples arose of LLMs being used privately, without training, in ways that implied uploading

citizens' data to private companies such as OpenAI or Anthropic without consent.

4.4 Impact of automation on public decision systems

The adoption of algorithmic systems in public administration has risen in a sustained trend in fields like justice, education, frontier control and social benefits management [1]. Eubanks [24] documents how this implementation in the United States has affected already precarized population in bigger extent, producing what she describes as digital poor-house: a set of digital infrastructures that control access to public resources and tend to shift the cost of error onto the most vulnerable applicants. Zajko [2] extends this analysis to the case of automated benefits systems, identifying recurring patterns of surveillance, conditionality, and exclusion associated with their implementation.

In Europe, regulators have tried to respond this trend in latest years. The Regulation (EU) 2024/1689 (Artificial Intelligence Act) classify as high risk systems those that evaluate the eligibility of citizens for essential social benefits and services, making them comply with transparency, robustness, human supervision and conformity requisites [19]. It is the most important regulatory frame for this project.

4.5 Critical perspectives

The criticism to algorithmic systems deployment in public administration is articulated at least in three complementary registers:

4.5.1 Sociology and STS criticism

From Science and Technology Studies, authors maintain that ADM introduction tends to depoliticize decisions that early on were taken at deliberative forums, presenting them as technical problems solvable by experts. STS authors defend this depoliticization is problematic because the codified criteria such as which variables count, what weight they carry or where the threshold is drawn are ultimately normative choices, not neutral. Eubanks [24] adds a dimension of inequality: the promise of efficiency implementing new systems is often paid for with added friction for low resources applicants.

4.5.2 Criticism from civil society

Civil society organizations like Civio or Eticas Foundation have driven applied investigations that have made effects of opaque systems visible. Inside this field, the BOSCO case is paradigmatic: after years of legal process over the source code since finding inconsistencies, the Spanish Supreme Court ruled in September 2025 that the code must be accessible for auditing purposes, establishing a constitutional precedent on algorithmic transparency in Spain [3]. Along the same lines, audits by Eticas of VioGén [4] and RisCanvi [25] have documented reliability and transparency issues in systems with severe impacts on the affected populations.

4.5.3 Technical critique from AI research

The fair machine learning field has developed a theoretical body to identify and quantify the biases that algorithmic systems can induce or amplify. Mehrabi et al. [26] provide a comprehensive map of the sources of bias and the most established definitions of fairness in the literature, while Chouldechova [27] formally demonstrates that various desirable notions of fairness are mutually incompatible, a fact that requires any design decision to have a normative component. For scoring and ranking systems like the one at hand, Zehlike et al. [28] provide the most comprehensive reference on fairness in ranking. Barocas et al. [8] synthesize the state of the art in a reference handbook that connects these technical contributions with their social and regulatory implications.

4.6 Spanish and Catalan situation

Spanish and Catalan administration has articulated its orientation regarding AI through its National Strategy on Artificial Intelligence (ENIA), which explicitly establishes the objective of transparent decision-making by Administration and the development on data-driven, evidence-based policies [10]. In Catalonia the Catalonia AI Strategy 2030 deploys a similar framework, with guiding principles such as democratic ethics, transparent, continuous innovation and equity [11]. These documents provide the internal regulatory framework according to which the actual deployed systems must be measured.

At the state level, Law 19/2014 on Transparency, Access to Public Information, and Good Government of Catalonia [29], constitutes the legal basis for requests for access to information about algorithmic systems. At the European level, the aforementioned AI Act [19] adds specific obligations for high-risk systems, a category to which the case studied belongs.

4.7 Audit methodologies

According to the systematic review by Li and Goel [12] AI auditing is structured around four primary categories: (1) ethics-based audits, (2) compliance audits, (3) process and governance audits, and (4) quality assurance. To facilitate these evaluations, the authors propose an AI Auditability Framework that identifies five essential components for examination: training data, the underlying model, process, governance, and people. Key auditability measures within this framework include ensuring data accessibility, providing detailed documentation of model intended use, and implementing traceability through sufficient logs and explainability features. Furthermore, successful auditing requires a multidisciplinary team with technical competencies in data science and AI engineering, complemented by knowledge of compliance, business processes, and the social implications of the systems.

On the other hand, within this mapping, the Adversarial Algorithmic Auditing Guide published by Eticas Foundation [15] is the reference that fits best the context of this work. It defines adversarial auditing as the process by which an independent third party or an affected community examines the impact and, to the extent possible, the functioning of an algorithmic system, in order to detect anomalies.

lies or unfair practices towards protected groups or society as a whole. The distinguishing feature they add, according to the authors, is restricted access to the algorithm and its design and validation databases, which limits the audit to the post-deployment phase. In parallel, Eticas Foundation [30] has developed the methodology Community-Led Audits (CLA), which places the affected communities at the center of the process and seeks to correct the elitist bias of other forms of auditing.

5 RESULTS

5.1 Interviews

The thematic coding consolidated the three interviews into five cross-cutting themes, used here to organise the informants' contributions and to connect them to the case. Positions are paraphrased in English, and the original Catalan wording is preserved in the transcripts.

(a) Data stewardship and governance. I1 reads the data layer as the part of public-sector automation that is actually advancing well: interoperability protocols, data-protection practice and the new data offices are being built solidly, in several cases pioneering at the state level, with universities and the administration sharing stewardship for both technical and ethical reasons. He places this, however, at a pre-governance stage: the administration is still handling non-interoperable datasets without metadata, some not even collectively owned but held by sub-agencies, so ethical and collective governance, meaning citizen monitoring and co-design, remains aspirational while the data office concentrates on data-protection compliance. He is explicit that this stewardship should not sit with the infrastructure provider (the AOC): infrastructure can be supplied centrally, but the decision over its use belongs next to each sector's policy, and governance should be networked and distributed rather than concentrated in a single body. I2 frames stewardship normatively rather than technically: the public administration is the only actor structurally obliged to protect fundamental rights, since a private provider's objective is profit, and therefore it cannot delegate that guardianship or audit itself through the market. I3 sharpens this into a critique of the present arrangement: the administration cannot be the auditor of its own systems, which is the role of civil society and independent research, and the existing venue, the Generalitat's data-ethics committee, is in practice the ceiling of civil participation while consisting essentially of a single civil-society representative alongside the departmental data-protection officers. The clearest divergence among the three sits here: where I1 treats data governance as the relative success of the whole process, I3 reads the same terrain as defined by its gaps. One of them is a jurisdictional hole they raised over police and security systems, which the executive placed outside the committee's remit on the grounds that it only covers AI use within the Generalitat's own administration, a reading they considers plainly wrong since the force is the Generalitat's.

(b) Planning, purpose and the infrastructure framing. The most consequential thread, carried mainly by I1 but present in all three, is that the binding constraint on public automation is not technical but the near-absence of strategic planning and purpose: half a century into modern adminis-

tration, strategic planning remains anecdotal, and AI is being asked to fix what are really problems of organisation and planning. From this follows his central reframing, namely that AI should be treated as infrastructure, a shared investment governed accordingly, and not as a commodity bought and consumed like any procurement item. The procurement step itself is reasonably solved; what is missing are the stages around it, defining purpose *ex ante* and evaluating impact *ex post*, and defining the purpose of an infrastructure is a political rather than technical decision that requires the concurrence of many actors. His positive illustration is the shared inter-university data pool, where the institutions stepped back to ask which functions they actually perform (certification, student mobility, library loans) and built common, interoperable data infrastructure underneath, instead of layering AI onto an unexamined process. Related to the analyzed need of a previous thinking the policy stage, asked where participation matters most, the three informants converge on the same answer: the design stage, and before it the definition of the problem. I1 argues that after design it is too late, and that diagnosis is itself where non-expert citizens can contribute, provided they are asked concrete questions rather than abstract ones. I2 likewise places participation at design and even at problem definition: not patching holes at speed but stopping to ask whether, and for what, the tool is needed. I3 states it most directly: the best moment is the start, when the problem is defined, since framing a need as "a chatbot that replaces the case interview" versus "a tool that drafts notes for the worker to revise" already decides whether the system aids or displaces the professional.

(c) Institutional capacity, talent and sovereignty. The three converge that the administration lacks the internal capacity to lead and over-relies on outsourcing. I1 puts it bluntly: the administration has essentially no capacity to develop technology and subcontracts almost everything, not from bad faith but from a real capability gap. His remedy is strategic workforce planning rather than reactive course-taking: inventorying the new functions and profiles needed years ahead and, above all, the process and organisational re-engineering that nobody is doing, given that the leap from form to data has not even been made. I2 makes capacity teaching and formation of every ecosystem agent her first priority, beginning with the administration that must lead and then society at large, which needs to understand what AI is and is not. I3 adds that the administration in fact holds substantial, underused internal talent and should stop working in patches, building shared knowledge and collective intelligence and, where the money is public, keeping the code public so it can be accountable. On the institutional form this should take, the three diverge. I1 is uncertain but generative: nobody yet knows which institutions are needed, so he welcomes even an under-defined new agency as a place to start thinking, and imagines its role as articulating the ecosystem, that is diagnosing, proposing methodologies and incentives, rather than executing. I2 is openly sceptical of a single state-level agency (AE-SIA): in an asymmetric state of very different autonomous communities it will act late and unevenly, and she would rather build from territorial scales that already make sense (comarques, vegueries, rural and urban realities) than from the state level. I3 locates the answer outside the administration altogether, in public-community and civil-society

collaboration as an alternative to the Silicon Valley offer, and folds in a notion of situated technological sovereignty: sovereignty is not only owning data centres but building them under the ecological constraints of a drought-prone territory and asking what the compute is actually for, reasoning from scarcity rather than assumed abundance.

(d) Openness and the limits of the registry. On the transparency of deployed systems the informants split over feasibility. I1 doubts that comprehensive external oversight is even tractable: the proliferation and crossing of systems is so fast that monitoring legal and ethical compliance everywhere is already impossible and will only worsen, so he favours distributed governance devices per purpose and, ultimately, empowering the affected individual to recognise when a system bears on them, backed by incentives that reward audited, free-software, open-library implementations over closed ones. I2 and I3 instead treat openness as necessary and currently failing. I2 defends open code because it makes systems auditable and their data provenance and errors knowable, while noting that the right degree of openness depends on the case. Both bring the same concrete evidence, which is the finding carried into the interviews: the Generalitat's algorithm registry lists on the order of thirty-three cases, yet for example the health-AI observatory alone reports a figure an order of magnitude higher, and of those thirty-three only three concern health. I3 details why it under-covers: it is not legally mandatory, the AI-regulation duty reaches only high-risk systems with broad exceptions, and it covers only the Generalitat's own public-function systems, excluding municipalities and, pointedly, security and policing, which is where small and rural municipalities are left most unprotected. Their standard is that the registry must not become performative, that is a static spreadsheet that lists systems without letting anyone audit the code, contest a decision or be genuinely empowered as a citizen.

(e) Positive counter-cases. Across the three, the positive examples share a pattern: automation that aids a decision or builds shared infrastructure without displacing a political choice or depositing public data on Big Tech. I1 cites the agent-based procedure simplification piloted in Sant Boi de Llobregat and, as his preferred case, the inter-university data pool noted above. I2 offers two cases of small, proximity engineering: a Les Borges Blanques municipal web assistant built from scratch in conversation with the town hall and trained only on its own data rather than resting on Big Tech search engines, and a federated system among Delta de l'Ebre rice growers that pools pest and crop data so fields can warn one another. She also points to the OEIAC's own public self-assessment model as a transparency mechanism users can apply to the systems they use. I3 offer instead of examples some possible ideas, as a Whisper-based tool that drafts notes for social workers to revise, that is AI as aid rather than substitute, defensible only because of how the problem was framed. They stress that the field is still very incipient and that, tellingly, they know of no civil-society participation that has yet become binding public policy.

5.2 Embedded case study

5.2.1 Data petition and acquisition

Anonymized data with the variables of the scoring rubric, zip code, and final resolution of the cases for the 2025 call was formally requested from the Housing Agency of Catalonia, citing the 19/2014 Transparency Law [29]. The request was granted and provided the empirical data used for the case analyses.

5.2.2 Internal consistency audit

The steps followed were (a) the verification of the published scoring rule against the actual scores applying the public weights to the per-applicant flags and checking the match, which also serves to pin down implementation details absent from the bases; (b) four red-flag checks: applicants violating the IRSC ceiling with favourable resolutions, high-scoring rejected applicants, low-scoring funded applicants, and recomputation mismatches; and (c) the analysis of the implicit budget cutoff and of the section 11.1 of the system's priority mechanism.

It has found two principal findings:

- The scoring system does not decide: for the set of cases that reach the scoring phase, the analysis shows that the score does not act as a selection mechanism: eligible cases without points receive the aid. What acts as the operational filter is the set of eligibility criteria applied beforehand (income limits, age, contract characteristics, documentation status, etc.). This finding has significant methodological implications: the substantive question is not whether the scoring discriminates, but whether and how eligibility does.
- Inconsistencies detected in the scoring formula: the process of reconstructing the formula from the regulatory bases and validating it against the actual data has allowed for the identification of a set of minor inconsistencies between what the call says and how points are applied in practice. These inconsistencies do not alter the aggregate results, but are relevant as indicators of the system's level of internal control.

5.2.3 Territorial reach and incidence audit

The four metrics were computed at two geographic scales: municipality, joined exactly to the INE Atlas income by municipal code [31], and postal code, obtained by population-weighted areal interpolation of census-section income onto postal-code polygons. Following the discipline fixed ex ante, every result is a claim about zones and not individuals [32], and a finding is reported as solid only if it holds at both scales. Table 2 summarises the four metrics at both scales, and Figure 2 shows Metrics A and B graphically.

Metric A, incidence by income decile. Zones were ordered by equalised net income and grouped into application-weighted deciles. The concession rate, computed as favourable resolutions over applications, stays in a narrow band across the lower and middle deciles, around 0.72 to 0.82, and falls in the richest deciles, down to 0.69 at municipality scale and 0.67 at postal-code scale. The decline at the top appears at both scales. This is coherent with Finding 1: the gradient is not produced by the score but by the

TABLE 1: Five cross-cutting themes across the three informants.

Theme	I1	I2	I3
(a) Data stewardship and governance	Data layer advancing well; pre-governance phase; distributed, sector-close stewardship, not the AOC	Only the administration can guard rights; cannot audit itself via the market	Administration cannot audit itself; committee is the ceiling of civil participation; security and police governance falls in a gap
(b) Planning, purpose and infrastructure	Binding constraint is planning, not tech; AI as infrastructure, not commodity; data pool as exemplar	Do not patch and run; build from scratch; participate in design and problem definition	Best participation is at problem definition; framing decides aid vs. substitute
(c) Capacity, talent and sovereignty	Low internal build capacity; needs workforce and process re-engineering; welcomes even an undefined new agency	Training as first priority; sceptical of a state-level agency; build from territorial scales	Underused internal talent; public-community collaboration; situated sovereignty and scarcity
(d) Openness and the registry	Comprehensive oversight is intractable; distributed devices, individual empowerment and open-source incentives	Pro open code (auditability, provenance); the 33-case registry is far too partial	Registry not mandatory and Generalitat-only; excludes municipalities and security; must not be performative
(e) Positive counter-cases	Sant Boi agent-based procedures; inter-university data pool	Les Borges Blanques assistant; Delta de l'Ebre federated system; OEIAC self-assessment model	Whisper-for-social-workers (aid, not substitute); civil-society tools; none yet binding policy

income ceiling, which leaves out the higher-income applicants who are more present in the wealthier zones.

Metric B, concentration. The signed concentration index is -0.021 at municipality scale and -0.017 at postal-code scale. With the convention that a positive index means concessions skew towards richer zones, both values are slightly negative, so the distribution is essentially proportional to where the applications are, with a small pro-poor tilt, and it is not regressive. The concentration curve stays close to the diagonal at both scales.

Metric C, three-level decomposition. Decomposing the funded rate as $\text{gate-pass} \times \text{scored-given-gate} \times \text{funded-given-scored}$ isolates where the difference between low and high-income zones is produced. The last factor, funded given scored, is approximately 1.000 at both scales: once a case is scored it is funded, with no budget rejection. The difference in the funded rate between low and high-income zones is small and favours the poorer zones, about $+3.3$ points at municipality scale and $+2.3$ at postal-code scale. That difference is located entirely before the score: partly at the IRSC gate, with gate-pass close to 0.95 in low-income zones against 0.93 in high-income ones, and partly at the remaining eligibility requisites, with scored-given-gate around 0.82 against 0.80. The score and the budget add nothing to the territorial gradient. As stated before, gate-pass here is a lower-bound proxy, because the IRSC ceiling is the only gate criterion observable in the microdata.

Metric D, origin incidence. The soft-weighted estimate [33] of the native-minus-foreign concession gap is about $+0.004$ at both scales, that is, at parity. The Manski partial-identification interval is $[-0.29, +0.79]$ at municipality scale and $[-0.30, +0.74]$ at postal-code scale, and both intervals contain zero. The conclusion is that the ecological data do not identify an origin gap: the point estimate sits at parity, but the bounds are too wide to sustain any directional claim, which is the expected situation when the pro-

tected attribute and the outcome are observed in separate datasets [34].

The four metrics agree across the two scales, so the results are reported as solid. The postal-code scale is more than a robustness check here: Barcelona is a single municipality that concentrates around 18% of all applications and collapses into one income value at the municipal scale, while the areal interpolation splits it into around forty postal codes ranging from roughly 15,000 to 53,000 euros of equivalised income, where most of the within-city variation actually is. Read together, the territorial layer reproduces Finding 1 from another direction: the eligibility gate, and not the score, governs who receives the aid, and the only territorial gradient that exists is mild, located at the gate, and slightly favourable to the lower-income zones.

6 CONCLUSIONS

RQ1. What is understood by AI in public policy, and how is it audited. The review shows that AI in public administration is best understood not as a single technology but as a gradient of automated decision-making, from code-driven deterministic rules to data-driven learned models [16], of which the deterministic class is largely deployed yet does not claim so much attention currently. The review further shows that the distinction between AI and non-AI systems is however consistently recognised in both the literature and the AI Act. Auditing a deterministic system also draws on AI-derived frameworks, in particular Li and Goel's [12] four-category typology and the ETICAS adversarial guide [15], which examine training data, model, process, governance and people, presuppose restricted post-deployment access, and rest on the auditability notions of traceability, explainability and contestability [9]. The conceptual answer the work defends is that the risk these audit obligations target follows the function, that is deciding ac-

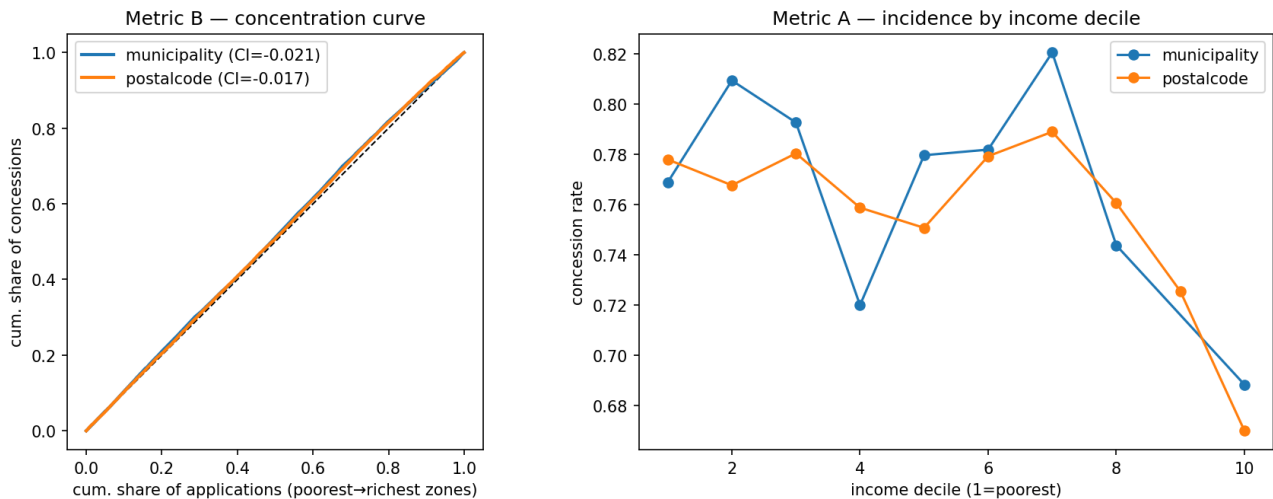


Fig. 2: Territorial reach and incidence, Metrics A and B at both geographic scales.

TABLE 2: Territorial reach and incidence: the four metrics at both geographic scales. A result is reported as solid only when it holds at the municipality and the postal-code scale.

Metric	Municipality	Postal code	Reading
A. Concession rate, low–middle deciles → richest decile	0.72–0.82 → 0.69	0.72–0.82 → 0.67	Flat across low and middle income, declines at the top
B. Signed concentration index	−0.021	−0.017	Near-proportional, slight pro-poor tilt; not regressive
C. Funded-rate gap (low – high income)	+3.3 pp	+2.3 pp	Small, favours poorer zones; located at the gate, not the score
D. Native – foreign gap (point estimate)	+0.004	+0.004	At parity
D. Native – foreign gap (Manski bounds)	[−0.29, +0.79]	[−0.30, +0.74]	Interval contains 0: origin gap not identified

cess to an essential benefit, and not the technique, so deterministic rules belong in the same auditing paradigm as ML even though each case needs appropriate and precise audit methodology.

RQ2. How it is being implemented and audited in Catalonia and Spain. From the three informants, adoption is real but uneven and largely unexamined: generative tools spread as shadow use without guidelines or training, perimeter control over them is impossible, and a precaution asymmetry holds whereby AI systems receive opaque pilots while being treated as open and audited. Auditing in practice is closer to nominal than substantive: transparency is performative and the binding constraint is political rather than technical. Positive counter-cases exist, where AI aids rather than substitutes a decision, but they are local and exceptional. Implementation seems to outpace governance, and systematic auditing is largely absent, as well as deployment objectives or characterization.

RQ3. How such a system can be audited, and what that reveals. The case demonstrates a concrete, low-cost audit of a deterministic ADM in two layers: a consistency audit that verifies the public rule against the released microdata and an ecological territorial reach audit reported with partial-identification bounds. Its central revelation is that the score does not decide: the binary eligibility gate does, since 2,661 zero-point applicants were funded and no

scored applicant was rejected for budget, showing that auditing is fully feasible once the data is released. Methodologically, AI-derived audit methods transfer in their process dimension, but their built-in assumption that the scoring artifact is the decision mechanism must be checked empirically before any fairness metric is applied. Taken together, the three answers support the overarching claim: the automation already shaping Catalan public policy is predominantly under-audited.

7 DISCUSSION AND FINAL CONSIDERATIONS

What the case says about auditability. The barem occupies the extreme code-driven end of Hildebrandt’s gradient, which is the category supplied by the review, and the audit confirms the theoretical expectation. A code-driven system is indeed auditable by rule revision, since the public rule reproduces the official scores exactly. But the exercise also shows that the published rules were not fully the implemented rules, given the four undocumented departures, and that the component the audit frameworks target, the scoring model, was not where allocation power resided, which was the eligibility gate. The transferability lesson is therefore double: the ETICAS adversarial framework transfers well

in its process dimension (stakeholder mapping, restricted-access posture, post-deployment focus), but its implicit assumption that the scoring or ranking artifact is the decision mechanism must be checked empirically before any fairness metric is computed. Applying ranking-fairness metrics [28] to a ranking that decides nothing would have produced formally correct but substantively empty results.

Reproducibility as advantage. It is worth stating plainly what the inability to audit an AI system actually cost and gained. The deterministic character of the barem is usually treated as the easy case, of lesser interest than learned models. For the question that turned out to matter, where the decision is made and whether it is political or technical, the opposite holds: precisely because the system is fully reproducible, it can be established with certainty that the scoring is inert and the eligibility gate is the lever, which is exactly the kind of structural finding an opaque model would have buried. Two qualifications show this: first, no ML system was built here, so the comparison across the rule-based and ML gradient is conceptual and literature-grounded [16, 8], not an empirical measurement. Second, the advantage is question-specific: transparency makes structural questions about where the power sits clean to answer, but tells us nothing about the class of harms that arise only inside learned systems, such as distributional bias in a trained model or feedback loops, for which AI-specific audit tooling remains necessary. The defensible claim is therefore not that deterministic audits are better, but that the implementation paradigm determines which questions an audit can answer cleanly.

Where the politics is (Winner, Callon). Following Winner [17], the politics of this artifact is not in the weights of the rubric but in the gate: income ceilings, contract requisites and documentation burdens are the design choices that distribute access. In Callon's [18] terms, the problematization failed: the rubric was enrolled to solve a scarcity-allocation problem that the budget of this call for applications dissolved, leaving a device that scores without deciding. The interviews independently supply both readings, the purpose-less tool and the genealogy of these systems as mere tools.

Competing narratives: solutionism versus purpose-first design. The clearest debate to emerge across the strands is between two stories about what automation is for. In the first, technology is laid on top of an existing process to make it faster, that is automation as a layer added to whatever the administration already does. In the second narrative, one asks first whether and for what AI is needed, and designs the tool and the process together. The informants' positive cases belong here, and the policy and polity diagnosis raised in the interviews names the failure mode of the first: the bottleneck is organizational and political, and automation layered on an unexamined process inherits, and hides, that process's problems rather than solving them. The position this work takes, while remaining a normative argument rather than an empirical result, is that automating a process before its purpose and design are settled tends to entrench friction and opacity rather than remove them, and that the payoff comes from co-designing the policy and the tool, not from adding intelligence on top of a process that was never interrogated.

Where automation helps and where it deepens the de-

bate. Drawing the three strands together, the case and the interviews suggest a tentative line between productive and contested uses of automation in public administration. On the productive side, the informants point to functions where added efficiency does not displace a political decision: accessibility-oriented service transformation, automated front-office attention, and genuinely data-driven, rather than data-dressed, policy design. The interviews' positive examples belong here, from procedure simplification and a shared inter-university data pool, to local-engineering cases such as a municipal assistant trained only on local data and a federated system among rice growers, to a Whisper-based drafting tool for social workers, AI as aid rather than substitute. On the contested side lie the functions that automation makes less inspectable rather than more: participation, transparency and contestability. The barem case shows the mechanism concretely, since the move from a deliberative allocation decision to a codified device shifts the normative choice, the eligibility thresholds, out of public view, where it draws no scrutiny precisely because it looks technical. The literature anticipates that this opacity cost grows as systems move from rule-based to learned: a published rule can in principle be read, whereas a data-driven model's internal logic resists the same inspection [16], and the resulting loss of contestability falls hardest on those least able to challenge it [24].

Limits. The main limitation is that the empirical case belongs to the deterministic end of the ADM spectrum. The original objective was to audit an AI-based system in order to contrast directly the literature on AI auditing with an operational public-sector deployment. However, no openly documented and sufficiently auditable AI system could be identified among the publicly disclosed Generalitat systems during the research period. This is itself indicative of a broader transparency problem: the systems that would have been most relevant for an AI audit were precisely those for which external scrutiny was least feasible. Interviews nevertheless pointed to the existence of additional deployments, including policing-related systems, whose future examination could provide a more direct test of the audit frameworks reviewed in this work. Also, the interview panel lacks the administrative-technical voice, the cancelled fourth interview, and the affected parties themselves. The case covers a single call for applications; in a budget-constrained year the barem would decide, and the documented discrepancies would become materially consequential.

Final considerations. Several threads are left open as natural continuations rather than loose ends. Repeating the audit on a budget-constrained call for applications, where the ranking actually operates, would test whether the discrepancies translate into differential outcomes. A follow-up information request on requisite-level rejection grounds would let the eligibility gate, the decisive but least-documented layer, be characterized directly instead of inferred. Extending the informant panel with the administrative-technical voice and, above all, with affected parties, ideally under a Community-Led Audit design, would close the participation gap this work could only name. And completing further robustness checks on the joined AHC and ADRH data would consolidate the second empirical layer.

REFERENCES

- [1] K. Levy, K. E. Chasalow, and S. Riley, "Algorithms and decision-making in the public sector," *Annual Review of Law and Social Science*, vol. 17, pp. 309–334, 2021. [Online]. Available: <https://doi.org/10.1146/annurev-lawsocsci-041221-023808>
- [2] M. Zajko, "Automated government benefits and welfare surveillance," *Surveillance & Society*, vol. 21, no. 3, pp. 246–258, 2023. [Online]. Available: <https://doi.org/10.24908/ss.v21i3.16107>
- [3] Civio, "Civio abre camino en la transparencia algorítmica: el supremo condena al gobierno a entregar el código fuente de bosco," Fundación Civio, Sept. 2025. [Online]. Available: <https://civio.es>
- [4] Eticas Foundation, "The external audit of the VioGén system," Eticas Foundation, 2022. [Online]. Available: <https://eticasfoundation.org>
- [5] A. Andrés-Pueyo, K. Arbach-Lucioni, and S. Redondo, "The RisCanvi: A new tool for assessing risk for violence in prison and recidivism," in *Handbook of Recidivism Risk/Needs Assessment Tools*, J. P. Singh et al., Eds. Wiley, 2018, pp. 255–268.
- [6] B. Lopes Buarque and N. Lewis, "Algorithms at your service: Understanding how X's systems of recommendation likely fuelled the far-right riots in the United Kingdom by amplifying visual representations of racist conspiracy theories," *The British Journal of Politics and International Relations*, vol. 28, no. 2, 2025.
- [7] Los Angeles County Superior Court, "K.G.M. v. Meta Platforms, Inc. et al., jury verdict," Judicial Council Coordinated Proceeding (JCCP) No. 5255, Los Angeles County Superior Court, 2026, bellwether trial on social media addiction; jury found Meta and Google (YouTube) liable, awarding \$6 million in damages, 25 March 2026.
- [8] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023.
- [9] High-Level Expert Group on AI, "Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment," European Commission, Tech. Rep., 2020.
- [10] Ministerio de Asuntos Económicos y Transformación Digital, "ENIA: Estrategia nacional de inteligencia artificial," Gobierno de España, Tech. Rep., 2020.
- [11] Generalitat de Catalunya, "Estratègia catalunya IA 2030," Departament de Polítiques Digitals, Tech. Rep., 2024.
- [12] Y. Li and S. Goel, "Making it possible for the auditing of AI: A systematic review of AI audits and AI auditability," *Information Systems Frontiers*, vol. 27, pp. 1121–1151, 2024. [Online]. Available: <https://doi.org/10.1007/s10796-024-10508-8>
- [13] Centro de Investigaciones Sociológicas, "Barómetros mensuales," CIS, 2026. [Online]. Available: <https://www.cis.es>
- [14] Observatori Metropolità de l'Habitatge de Barcelona, "L'assequibilitat de l'habitatge a catalunya en el 2024: Informe anual," O-HB, Tech. Rep., 2026.
- [15] G. Galdon-Clavell, I. Nalbantova, A. Akisheva, and L. R. González, "Adversarial algorithmic auditing guide," Eticas Research and Innovation, Tech. Rep., 2023.
- [16] M. Hildebrandt, "Algorithmic regulation and the rule of law," *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2128, p. 20170355, 2018. [Online]. Available: <https://doi.org/10.1098/rsta.2017.0355>
- [17] L. Winner, "Do artifacts have politics?" *Daedalus*, vol. 109, no. 1, pp. 121–136, 1980.
- [18] M. Callon, "Some elements of a sociology of translation: Domestication of the scallops and the fishermen of st brieuc bay," in *Power, Action and Belief: A New Sociology of Knowledge?*, J. Law, Ed. Routledge, 1986, pp. 196–233.
- [19] European Parliament and Council, "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," Official Journal of the European Union, 2024. [Online]. Available: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- [20] H. Hermanns, A. Lauber-Rönsberg, P. Meinel, S. Sterz, and H. Zhang, "AI act for the working programmer," in *Int. Symp. on Leveraging Applications of Formal Methods (AISO LA 2024)*, 2024, preprint arXiv:2408.01449. [Online]. Available: <https://arxiv.org/abs/2408.01449>
- [21] G. Juell-Skielse, I. Lindgren, and M. Åkesson, Eds., *Service Automation in the Public Sector: Concepts, Empirical Examples and Challenges*, ser. Progress in IS. Springer, 2022. [Online]. Available: <https://doi.org/10.1007/978-3-030-92644-1>
- [22] A. Ranerup and H. Z. Henriksen, "Value positions viewed through the lens of automated decision-making: The case of social services," *Government Information Quarterly*, vol. 36, no. 4, p. 101377, 2019. [Online]. Available: <https://doi.org/10.1016/j.giq.2019.05.004>
- [23] OECD, "Governing with artificial intelligence: The state of play and way forward in core government functions," OECD Publishing, Paris, Tech. Rep., 2025. [Online]. Available: https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287.html
- [24] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

- [25] G. Galdon-Clavell *et al.*, “Automating (in)justice? an adversarial audit of RisCanvi,” Eticas Foundation, Tech. Rep., 2024.
- [26] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A survey on bias and fairness in machine learning,” *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2021. [Online]. Available: <https://doi.org/10.1145/3457607>
- [27] A. Chouldechova, “Fair prediction with disparate impact: A study of bias in recidivism prediction instruments,” *Big Data*, vol. 5, no. 2, pp. 153–163, 2017. [Online]. Available: <https://doi.org/10.1089/big.2016.0047>
- [28] M. Zehlike, K. Yang, and J. Stoyanovich, “Fairness in ranking, part I: Score-based ranking,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, 2022. [Online]. Available: <https://doi.org/10.1145/3533379>
- [29] Parlament de Catalunya, “Llei 19/2014, del 29 de desembre, de transparència, accés a la informació pública i bon govern,” DOGC núm. 6780, 2014.
- [30] Eticas Foundation, “Community-led AI audits: Methodology for placing communities at the center of AI accountability,” Extended abstract, Participatory AI Research Symposium, 2024.
- [31] Instituto Nacional de Estadística, “Atlas de distribución de renta de los hogares (ADRH),” INE, 2023. [Online]. Available: <https://www.ine.es>
- [32] W. S. Robinson, “Ecological correlations and the behavior of individuals,” *American Sociological Review*, vol. 15, no. 3, pp. 351–357, 1950.
- [33] J. Chen, N. Kallus, X. Mao, G. Svacha, and M. Udell, “Fairness under unawareness: Assessing disparity when protected class is unobserved,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ser. FAT* ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 339–348. [Online]. Available: <https://doi.org/10.1145/3287560.3287594>
- [34] N. Kallus, X. Mao, and A. Zhou, “Assessing algorithmic fairness with unobserved protected class using data combination,” *Management Science*, vol. 68, no. 3, pp. 1959–1981, 2022. [Online]. Available: <https://doi.org/10.1287/mnsc.2020.3850>