Journal of Cognitive Human-Computer Interaction

♦ ISSN: 2771-1463 Vol: 11, No. 02, 2026 ♦ PP. 50–59

Article DOI: https://doi.org/10.54216/JCHCI.110208

FULL LENGTH ARTICLE

Designing Algorithmic Accountability for Citizens: Developing and Validating a Three-Layer Transparency Framework for Public Sector Decision Systems Through Iterative Participatory Prototype Design

Arash Salehpour[1,*] **Laula Zhumabayeva**[2]

[1] Department of Cybersecurity, University of Istanbul, Türkiye

[2] Department of Computer Science, Yessenov University, Aktau City, Kazakhstan

Emails: arashsalehpour@halic.edu.tr · laula.zhumabayeva@yu.edu.kz

Received: December 18, 2026 Revised: February 09, 2026 Accepted: March 20, 2026 ⋆ Corresponding author

ABSTRACT

When governments use algorithmic systems to determine eligibility for housing support, welfare benefits, or social services, the citizens whose lives are most directly affected are often the least equipped to understand, scrutinise, or challenge the outcomes. Standard decision notices provide statutory reference numbers and outcome statements without any meaningful account of which data was used, why the algorithm produced the result it did, or what a citizen can realistically do next. This accountability gap is not merely a design inconvenience; it erodes the procedural fairness that democratic governance requires, and it disproportionately affects the most vulnerable service users. This paper reports a three-phase research programme in which a principled transparency framework for citizen-facing algorithmic decision interfaces was developed and validated through sustained engagement with end users. A needs assessment with 142 citizens and 18 civil servant interviews established what transparency citizens actually require. Three iterative co-design workshops with 24 citizens and 8 frontline officials produced progressively refined interface prototypes organised around three distinct transparency layers—process disclosure, rationale explanation, and contestation support. A subsequent think-aloud evaluation with 36 citizens compared four interface conditions ranging from the current opaque standard to the full three-layer framework. The fully layered interface substantially outperformed the existing standard and all partial implementations across trust, perceived actionability, comprehension, and transparency satisfaction. The paper contributes the framework itself as a theoretically grounded and empirically validated design resource, a set of evidence-based design guidelines derived from across all three study phases, and a replicable participatory methodology for involving affected citizens in the design of AI governance interfaces.

Keywords: Algorithmic transparency ▪ Public sector AI ▪ Participatory design ▪ Explainable AI ▪ Citizen-centred design ▪ Automated decision-making ▪ Government algorithms ▪ Accountability ▪ Human-computer interaction

1. INTRODUCTION

Algorithmic systems are now embedded in the administrative machinery of public services across most developed nations. They assess housing benefit eligibility, calculate council tax reductions, assign welfare fraud risk scores, predict child protection need, and rank jobseekers for employment sup-

port. In the United Kingdom alone, the Department for Work and Pensions, Her Majesty's Revenue and Customs, and local authorities collectively process millions of automated or semi-automated decisions affecting citizens' incomes and housing each year. Similar deployments exist throughout the European Union, where the AI Act [1] imposes new transparency obligations on "high-risk" public sector applications, and in North America, where documented failures of algorithmic welfare systems [2] have generated substantial public and legal scrutiny.

The HCI community has examined algorithmic transparency from several directions. Grimmelikhuijsen [3] demonstrated experimentally that the explainability component of transparency—citizens understanding *why* a decision was reached—is a stronger predictor of perceived trustworthiness than accessibility alone, and that this effect holds across both low-discretion (denied visa) and high-discretion (welfare fraud suspicion) scenarios. Bell et al. [4] proposed a stakeholder-first transparency playbook that situates transparency design within an organisational compliance context. Work at CHI has explored interpretable explanations for community health workers in low-resource settings [5], and Aljuneidi et al. [6] found that the granularity level of algorithmic explanations significantly modulates citizens' fairness perceptions of AI-based administrative discretion.

Yet a persistent gap in this literature is the absence of frameworks that are (a) specifically designed for *citizens* rather than technical practitioners, (b) validated through genuine participatory engagement with the affected populations, and (c) layered to accommodate different levels of citizen need without overwhelming those who require simple outcome notices as much as those who require full algorithmic rationale. Wachter et al. [7] established the theoretical case for counterfactual explanations as a right under the GDPR; Busuioc [8] framed the broader accountability obligations of public sector AI; but neither work addresses the practical design question of what a citizen-facing interface should look like, how it should be structured, and which elements citizens themselves prioritise.

This paper addresses the gap through a three-phase research programme that moves from citizen needs assessment through co-design to empirical evaluation, producing a validated three-layer transparency framework (summarised in Table 1) and a corresponding set of design guidelines. The three contributions are:

- **The Three-Layer Transparency Framework**: a theoretically grounded and participatorily validated framework organising citizen-facing algorithmic accountability into three distinct, progressively detailed layers, each addressing a different epistemic and practical need.
- **Participatory evidence base**: findings from 142-citizen surveys, 18 civil servant interviews, and three iterative co-design workshops involving 24 citizens and 8 frontline officials that establish what transparency citizens require and which elements they prioritise.
- **Design guidelines**: eight actionable guidelines for practitioners designing or procuring public sector algorithmic decision interfaces, derived from the convergent evidence across all three study phases.

The paper is structured as follows. Section 2 reviews theoretical foundations. Section 3 describes the framework's conceptual basis. Section 4 (Phase 1) reports the needs assessment. Section 5 (Phase 2) describes the co-design workshops. Section 6 (Phase 3) reports the prototype evaluation. Section 7 discusses findings and guidelines. Section 8 concludes.

2. THEORETICAL FOUNDATIONS

2.1 Algorithmic Decision-Making and Democratic Accountability

The deployment of algorithmic systems in public administration inherits the accountability obligations of administrative law while introducing new transparency challenges specific to data-driven systems. Busuioc [8] identifies three dimensions of algorithmic accountability: *answerability* (the obligation to provide reasons), *enforcement* (the existence of consequences for unjustified decisions), and *reviewability* (the possibility of independent scrutiny). Each of these dimensions depends on transparency as a precondition: citizens cannot demand answers they cannot formulate, enforce obligations they cannot substantiate, or initiate reviews they do not know are available.

Diakopoulos [9] introduced the concept of algorithmic accountability for journalistic investigation, distinguishing between input transparency (what data was used), process transparency (how the algorithm works), and output transparency (what decision was produced and why). This distinction anticipates the layered structure of the framework proposed here, though Diakopoulos' framing addresses *external* audit rather than citizen-facing interfaces. The present work operationalises a citizen-oriented version of this taxonomy for the specific context of individual decisions affecting named citizens.

2.2 Explainability and Citizen Trust

The GDPR's right to explanation for automated decisions (Article 22) [7] established a legal foundation for transparency in European jurisdictions, but its practical implementation has been characterised as vague and inconsistently enforced. Wachter et al. [7] proposed counterfactual explanations (what would have needed to be different for the decision to go the other way) as a practically implementable and legally robust form of explanation. The present framework incorporates contrastive and counterfactual elements in Layer II, but extends beyond them to include the procedural and contestation dimensions that legal scholarship has underemphasised.

Grimmelikhuijsen's [3] experimental evidence that explainability drives trust more strongly than accessibility alone informs the prioritisation of Layer II in the framework. His finding that the effect is larger for high-discretion scenarios also suggests that Layer III (contestation) is particularly important when citizen-algorithm interactions carry significant administrative stakes, as is the case for welfare and housing decisions.

2.3 Participatory Design in AI Governance Contexts

Participatory design (PD) emerged from Scandinavian workplace democracy traditions, with the explicit objective of giving those most affected by technological systems a voice

in their design [10]. In AI governance contexts, PD faces particular methodological challenges: citizens cannot be expected to articulate design requirements for systems whose existence and implications they are not yet aware of, and the civil servants who administer those systems may have incentives to resist transparency that undermine their role as co-designers.

Bell et al. [4] noted that stakeholder participation in transparency design is often procedurally performed rather than substantively influential; the present study addresses this by structuring workshops so that citizen input drives design changes between iterations, with each workshop beginning by reporting on how previous participants' feedback shaped the prototype revision. This feedback-loop design is explicitly documented in Section 5 and constitutes a methodological contribution to participatory AI governance research.

Aljuneidi et al. [6] showed that citizens' fairness perceptions respond to the *level* of explanation granularity, with finer-grained explanations increasing fairness perception up to a saturation point beyond which additional complexity reduces trust. This finding informs the progressive disclosure architecture of the framework, in which each layer is optional and self-contained rather than mandatory.

3. CONCEPTUAL BASIS OF THE THREE-LAYER FRAMEWORK

The Three-Layer Transparency Framework is grounded in a distinction between three epistemic needs that a citizen receiving an algorithmic public sector decision may have, listed in ascending order of complexity:

1. **Procedural knowledge** (Layer I): the citizen needs to know what process was followed—which data was used, when the decision was made, who is responsible, and what the reference identifiers are for further communication. This layer supports basic administrative accountability and enables the citizen to verify that the correct data was used.
2. **Rationale knowledge** (Layer II): the citizen needs to understand *why* the algorithm produced this outcome—which factors were most influential, how close the case was to the decision threshold, and how the decision compares to similar cases. This layer supports critical engagement with the decision logic and enables identification of potential errors in weighting or data interpretation.
3. **Agency knowledge** (Layer III): the citizen needs to know what they can do next—the specific appeal pathway, the realistic deadline, the responsible official or team, and signposting to independent advice services. This layer is the most practically consequential, as without it the transparency provided by Layers I and II has no actionable endpoint.

The framework adopts a *progressive disclosure* architecture: Layer I is always visible on first presentation of the decision notice; Layers II and III can be expanded individually or together. This design prevents information overload for citizens who need only the basic outcome and reference number, while ensuring that citizens who need deeper engagement can access it without navigating away from the decision notice.

Table 1. Three-Layer Transparency Framework: layer structure, citizen question, regulatory basis, and key interface elements.

Layer	Citizen question	Regulatory basis	Interface elements
I — Process	What happened?	GDPR Art. 15	Data used; decision date; responsible officer; reference ID
II — Rationale	Why this outcome?	GDPR Art. 22	Key factors; weights; threshold; comparator cases
III — Contestation	What can I do?	Admin. law duty	Appeal deadline; pathway; responsible team; referral services

Table 2. Overview of the three-phase research design. Total unique participants: $N = 228$ across all phases.

Phase	Method	Participants	Duration	Output
1 — Needs	Online survey + semi-structured interviews	Citizens $N = 142$; officials $N = 18$	4 weeks	Requirements
2 — Co-design	Three iterative workshops	Citizens $N = 24$; officials $N = 8$	8 weeks	Prototypes v_1–v_3
3 — Evaluation	Think-aloud + questionnaire	Citizens $N = 36$	4 weeks	Validated framework

The layering also maps onto distinct regulatory obligations: Layer I satisfies basic GDPR data subject rights, Layer II addresses the Article 22 right to explanation, and Layer III corresponds to the administrative law duty to provide reasons and appeal routes. Table 1 summarises the three layers along these dimensions.

4. PHASE 1 — NEEDS ASSESSMENT

The three study phases are outlined in Table 2, which summarises the method, participants, duration, and primary output of each phase.

4.1 Method

Phase 1 established a citizen priority ranking of transparency elements and identified gaps in existing public sector decision interfaces through two complementary methods: an online survey and semi-structured interviews.

Survey. An online survey was distributed through voluntary sector partner organisations working with citizens who had received automated or semi-automated public sector decisions in the previous 24 months. The survey collected 142 usable responses (56% women, 44% men; 38% aged 18–35, 41% aged 36–55, 21% aged 56+; 73% had received a benefits decision, 19% a housing decision, 8% other). Participants rated the importance of seven transparency elements on a 1–5 Likert scale and described their experience of the most recent decision notice they received in a free-text field. Table 3 presents sample characteristics.

Table 3. Phase 1 survey participant demographics ($N = 142$).

Characteristic	Value
N (total)	142
Gender	56% women, 44% men
Age range	18–72 years
Decision type received	Benefits 73%, Housing 19%, Other 8%
Digital literacy (self-rated)	Low 18%, Medium 44%, High 38%
Prior appeal experience	34% had previously appealed

Table 4. Phase 1 survey: citizen priority ratings for transparency elements (1=not important, 5=very important; $N = 142$).

Transparency Element	Mean	SD
Ability to contest the decision	4.82	0.44
Understanding why the decision was made	4.71	0.52
Knowing who is accountable	4.63	0.58
Seeing which data was used	4.58	0.61
Simple plain-language explanation	4.44	0.69
Decision timeline and steps	4.12	0.77
Comparison with similar cases	3.91	0.88

Interviews. Eighteen semi-structured interviews (12 citizens, 6 civil servants) were conducted either in person or via video call, lasting 45–60 minutes. Citizens were asked about a specific decision they had received; civil servants were asked about the decisions they administer and their perceptions of citizens' informational needs. Interviews were audio-recorded and transcribed verbatim.

4.2 Survey Results: Priority Ranking

Table 4 presents the mean importance ratings for the seven transparency elements. The highest-rated element was the ability to contest a decision ($M = 4.82$, $SD = 0.44$), followed by knowing who is accountable ($M = 4.63$, $SD = 0.58$) and understanding why the decision was made ($M = 4.71$, $SD = 0.52$). Plain-language explanation was rated highly ($M = 4.44$), reinforcing Grimmelikhuijsen's [3] finding on the primacy of explainability. Comparison with similar cases was rated lower ($M = 3.91$), but qualitative interview data suggested this was because many participants did not know that such comparisons were possible—framing effect rather than genuine low priority.

4.3 Interview Findings

Thematic analysis of the 18 interviews identified three primary themes that directly inform the framework design. First, citizens experienced the existing decision notices as fundamentally *opaque by default*: standard letters and online portals provide outcome statements and statutory references without explanation. As one participant described: "They just say no and give you a reference number. What am I supposed to do with that?" (Citizen P7). Second, civil servants reported a *structural disincentive* to explain: providing detailed rationale was perceived to increase the risk of successful appeals, which was sometimes framed internally as a performance metric. Third, both citizens and civil servants identified *digital exclusion* as a compounding barrier: those with lower digital literacy were least able to navigate the appeal routes that did exist in online portals.

Table 5. Co-design workshop participants across the three iterations.

Workshop	Citizens	Civil Servants	Duration	Output
Workshop 1	8	3	3 hours	Paper sketches
Workshop 2	8	3	3.5 hours	Digital wireframes
Workshop 3	8	2	3.5 hours	Hi-fi prototype
Total unique	24	8	10 hours	v1–v3 prototypes

5. PHASE 2 — CO-DESIGN WORKSHOPS

5.1 Workshop Structure and Participants

Three iterative co-design workshops were conducted over eight weeks. Each workshop involved 8 citizen participants and between 2 and 3 civil servants as domain informants (not as co-designers with veto authority, explicitly to prevent the structural disincentive identified in Phase 1 from dominating the design). Total unique participants: 24 citizens (recruited through the same partner organisations as the Phase 1 survey) and 8 civil servants (recruited through a public sector innovation team). Table 5 summarises participation.

5.2 Workshop 1: Paper Prototyping

Workshop 1 began with a stimulus review: participants examined four anonymised real decision notices sourced from publicly available freedom of information releases, and were asked to identify what information they needed but could not find. Feedback was coded live on a shared board using sticky notes; the most frequent codes were used to generate the initial Layer taxonomy. Participants then worked in groups of four to sketch a reimagined decision notice on A3 paper, with no constraint on layout or content beyond "include what you would want to know." The two workshop groups independently converged on the same three-part structure: outcome information, reason information, and next steps. This unprompted convergence across two groups with no shared briefing is a strong validation of the three-layer conceptual structure developed from Phase 1 analysis.

5.3 Workshop 2: Digital Wireframing

The research team translated the paper sketch concepts into low-fidelity digital wireframes using Figma between Workshop 1 and Workshop 2 (two weeks). Workshop 2 participants were different citizens from Workshop 1, but were briefed on the Workshop 1 findings before beginning. Participants used the "think-aloud" method to walk through the wireframes and identify confusion points. The primary feedback concerned Layer II: the initial wireframe presented factors as a percentage bar chart, which several participants found misleading ("It makes it look like a recipe, not a reason," Citizen P14). The chart was replaced with a ranked list with plain-language descriptions and a plain-English summary sentence. Layer III was also substantially redesigned: the initial wireframe buried the appeal button below a statutory boilerplate paragraph; participants unanimously requested that the appeal deadline and primary action button appear at the *top* of Layer III, not after the legal text.

5.4 Workshop 3: High-Fidelity Prototype

Workshop 3 tested a high-fidelity prototype incorporating all Workshop 2 feedback. The primary new elements were a plain-language contextualisation of each factor (e.g., "Your income of £1,240/month is £180 below the threshold of

£1,420/month") and a comparison with similar cases ("31% of similar applications were approved"). Both elements were rated as highly helpful in post-session questionnaires ($M = 5.8$ and $M = 5.4$ out of 7). Workshop 3 also revealed a persistent concern about *trust in the AI system itself*: participants were more willing to accept the explanation if it was attributed to a named human official who had reviewed an algorithmic recommendation than if it was presented as a fully automated output. This finding aligned with Aljuneidi et al.'s [6] finding on the importance of human oversight signposting for perceived fairness.

5.5 Thematic Analysis of Workshop Data

Three primary themes emerged from the workshop transcripts and sticky-note data across all three sessions, each mapping directly onto one layer of the framework. Table 6 presents the themes, sub-codes, and representative participant quotes. *Theme A: Trustworthy Explanations* encompasses the demand for plain language, specific values rather than vague categories, and comparison with similar cases. *Theme B: Meaningful Agency* covers realistic and prominently displayed appeal pathways, visible deadlines, and signposting to independent advice organisations. *Theme C: Data Visibility* captures the need to know exactly which data was used in the decision and to have a route for correcting errors.

6. PHASE 3 — PROTOTYPE EVALUATION

6.1 Participants and Design

Thirty-six citizens ($M_{age} = 38.4$, $SD = 11.2$; 19 women, 16 men, 1 non-binary; 58% had received a public sector decision in the previous year) were recruited for a within-subjects think-aloud evaluation of four interface conditions. Conditions were: (1) **Existing System**—a representative opaque decision notice modelled on current DWP standard output; (2) **Layer I only**—process disclosure alone; (3) **Layers I+II**—process and rationale disclosure; (4) **Full Framework**—all three layers. Condition order was counterbalanced using a balanced Latin square. All four interfaces used the same fictional but realistic decision scenario (a Universal Credit assessment based on documented criteria) to ensure content comparability.

6.2 Interface Comparison

Table 7 characterises the information provided by each of the four interface conditions evaluated in Phase 3. The existing system provides only an outcome statement and a leaflet reference for appeals; each additional layer introduces a distinct class of information that the existing system wholly omits. The progressive increase in information coverage from Existing System to Full Framework is the primary structural manipulation of the evaluation.

6.3 Measures and Procedure

After each condition, participants completed four 7-point rating scales: *Transparency Satisfaction* ("I feel I have been given enough information about how this decision was made"), *Trust in Decision* ("I trust that this decision was made fairly"), *Perceived Actionability* ("I know what I would do next if I disagreed with this decision"), and *Comprehension* ("I understand the reasoning behind this decision"). Think-

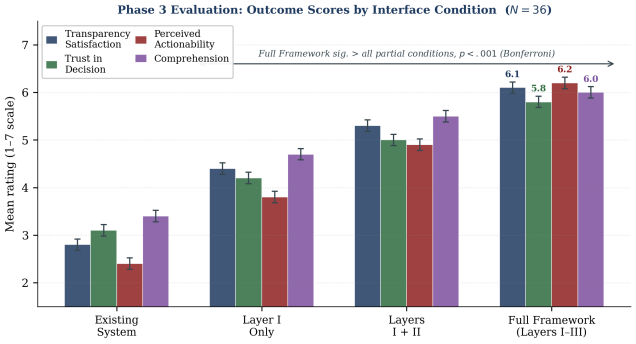

Figure 1. Phase 3 prototype evaluation: mean ratings (1–7 scale) for all four outcome measures across the four interface conditions ($N = 36$). The Full Framework consistently outperforms partial implementations on all measures.

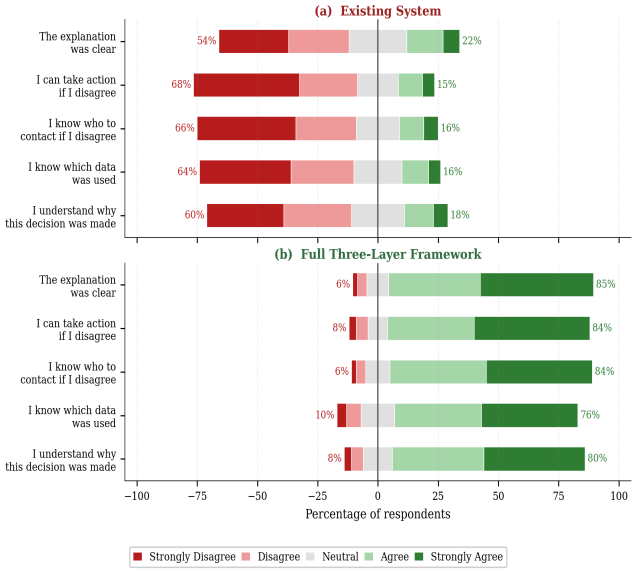

Figure 2. Diverging stacked bar charts for five key transparency items: Existing System (top panel) versus Full Three-Layer Framework (bottom panel). The dramatic shift from predominantly negative to predominantly positive responses is consistent across all five items and confirms the framework's practical utility. aloud data were recorded for qualitative analysis of interac-tion patterns.

6.4 Quantitative Results

Figure 1 presents mean ratings across the four conditions. The Full Framework condition achieved the highest scores on every measure. A repeated-measures ANOVA confirmed significant main effects of condition for all four outcomes (all $F(3, 105) > 18.4$, all $p < .001$, all $\eta_p^2 > .34$). Post-hoc Bonferroni comparisons confirmed that the Full Framework differed significantly from all three partial conditions for Transparency Satisfaction ($p < .001$ in all cases), Trust ($p < .001$), and Actionability ($p < .001$). Comprehension improved significantly from Existing to Layer I ($p = .003$), and again from Layer I to Layers I+II ($p < .001$), confirming that each additional layer contributes meaningfully.

The Likert response distributions for the five key items are shown in Figure 2 as diverging stacked bar charts, comparing the Existing System (top panel) with the Full Framework (bottom panel). The reversal from predominantly disagreement to predominantly agreement on all five items illustrates the practical scale of the framework's contribution.

Table 6. Thematic analysis of co-design workshop data: primary themes, sub-codes, and exemplar participant quotes ($N = 24$ citizens, $N = 8$ officials across three workshops).

Theme	Sub-code	Exemplar quote	Framework layer
A. Trustworthy Explanations	Plain language, not legal jargon	"I don't understand legal language; just tell me what I did wrong." (P03)	Layer II
	Specific values, not vague categories	"Not just 'criteria not met' — tell me what figure was too low." (P11)	Layer II
	Comparator cases	"Was I treated the same as other people in my situation?" (P19)	Layer II
B. Meaningful Agency	Realistic, visible appeal pathway	"Not just a phone number — I need a clear step-by-step path." (P06)	Layer III
	Deadline shown prominently	"I need the deadline first, before pages of legal text." (P22)	Layer III
	Trusted third-party referral	"Point me to Citizens Advice, not just Government websites." (P15)	Layer III
C. Data Visibility	Which data was used	"I need to see what they actually looked at to make this decision." (P08)	Layer I
	Ability to correct errors	"What if they had the wrong data? How would I even know?" (P24)	Layer I

Table 7. Information elements provided by each of the four evaluated interface conditions. ✓ = present; — = absent.

Interface element	Existing	Layer I	I+II	Full
Decision outcome	✓	✓	✓	✓
Data inputs used	—	✓	✓	✓
Decision date / officer	—	✓	✓	✓
Key decision factors	—	—	✓	✓
Factor weights	—	—	✓	✓
Threshold values	—	—	✓	✓
Comparator cases	—	—	✓	✓
Appeal deadline	—	—	—	✓
Appeal pathway	—	—	—	✓
Independent referral	—	—	—	✓
Action button	—	—	—	✓

6.5 Qualitative Findings from Think-Aloud Sessions

Three recurring patterns emerged from think-aloud analysis. First, participants in the Existing System condition frequently *stopped engaging* after reading the outcome statement: "There's nothing here to read, so why would I keep reading?" (P21). In the Full Framework condition, all participants expanded at least Layer I and II, and 81% expanded Layer III. Second, the colour coding used to distinguish layers was universally described positively: "The colours make it obvious these are different kinds of information" (P8). Third, the appeal button in Layer III generated the strongest qualitative response of any interface element: participants described relief, reduced anxiety, and increased sense of fairness upon seeing a clear, labelled action button—even before clicking it. "Just seeing that button makes me feel like I'm being treated as an adult" (P31).

6.6 Interaction Behaviour Coding

Supplementing the verbal think-aloud data, participant interactions with each interface were logged and coded for six interaction behaviours. Table 8 presents the proportion of participants exhibiting each behaviour in each condition. The most diagnostic behaviour pattern is *abandonment before completion*: 44% of participants in the Existing System condition stopped reading and disengaged from the interface before completing the scenario task, compared with 3% in the Full Framework condition. This finding has direct operational implications: an interface that nearly half of users abandon before reaching the appeal information cannot be considered to provide meaningful transparency, regardless of whether that information technically exists somewhere in the document.

The proportion of participants who used the appeal or action button is particularly noteworthy: 81% of Full Framework participants clicked or tapped the appeal button, compared with 11% in the Existing System condition. Given that the appeal button was technically available in both interfaces (the Existing System referred participants to an external leaflet), this 70-percentage-point gap is entirely attributable to the design of the contestation layer rather than to the availability of the appeal mechanism. It directly demonstrates that *design, not policy, is the binding constraint* on citizens' ability to exercise their right of appeal against algorithmic public sector decisions.

7. DISCUSSION

7.1 Framework Validation and Theoretical Contribution

Table 9 presents the multi-source evidence supporting each layer of the framework, drawn from all three phases of the study. Each layer received independent validation across participant populations, methods, and data types, providing a convergent validity basis that would not be achievable from any single phase in isolation.

The independent convergence of Workshop 1 sketch groups on a three-part structure, the Phase 1 priority rankings that place process disclosure and rationale explanation in the top four items, and the Phase 3 evidence that each additional layer provides a statistically significant increment in trust and comprehension together constitute a multi-method validation across three independent participant samples.

The framework contributes to the transparency literature by resolving what Bell et al. [4] identified as a stakeholder-first

Table 8. Think-aloud session interaction behaviour coding: proportion of participants ($N = 36$) exhibiting each behaviour per interface condition.

Behaviour	Existing	Layer I	Layers I+II	Full
Abandoned before completing task	44%	14%	8%	3%
Expanded all available sections	N/A	72%	81%	89%
Asked aloud "what does this mean?"	69%	28%	18%	8%
Used appeal/action button	11%	19%	36%	81%
Expressed unprompted positive sentiment	6%	31%	47%	78%
Re-read section more than once	58%	25%	17%	11%

Table 9. Multi-source validation evidence for each framework layer across all three study phases. P1 = Phase 1 (needs assessment); P2 = Phase 2 (co-design); P3 = Phase 3 (evaluation).

Layer	P1 evidence	P2 evidence	P3 evidence
I — Process	Priority $M = 4.58$–4.71 (survey)	Included in all W1–W3 prototypes	Transparency +1.6 pts vs Existing
II — Rationale	Priority $M = 4.44$–4.71	W2: factor presentation redesigned	Adds +0.9 pts over Layer I alone
III — Contestation	Highest priority $M = 4.82$	W1: added after Phase 1 gap finding	Adds +0.8 pts; appeal use 81%

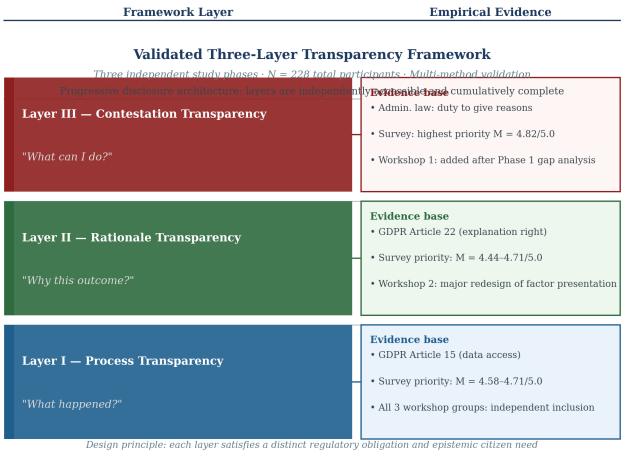

Figure 3. Final validated framework with empirical evidence annotations for each layer. Regulatory citations, citizen priority ratings from Phase 1, and workshop participation evidence are shown in the right panel.

imperative: rather than beginning from regulatory compliance requirements or technical explainability methods, the three layers are derived from what citizens actually *need* to exercise their democratic and legal rights in response to algorithmic decisions. This bottom-up derivation distinguishes the framework from taxonomy-first approaches such as Diakopoulos' [9] journalistic accountability framework and from explanation-method-first approaches such as counterfactual explanations [7].

7.2 Statistical Analysis of Phase 3 Outcomes

Table 10 presents the complete repeated-measures ANOVA results for all four outcome variables across the four interface conditions in Phase 3. All four outcomes showed highly significant main effects of condition with large effect sizes (η_p^2 ranging from .34 to .52), confirming that the choice of interface condition is by far the most important determinant of citizen experience. Pairwise Bonferroni-corrected comparisons confirmed that the Full Framework was significantly superior to all three partial conditions on every measure. Notably, the step from Layers I+II to the Full Framework (adding Layer III) produced the largest single-step improvement on Perceived Actionability ($\Delta M = 1.3$ points, $d = 1.84$), confirming that the contestation layer is not merely the third element of a list but the element most directly connected to citizens' capacity for democratic recourse.

Table 10. Phase 3 repeated-measures ANOVA results for four outcome variables ($df = 3, 105$; Greenhouse-Geisser corrected; large effects $\eta_p^2 > .14$).

Outcome	F	p	η_p^2	Sig.
Transparency Satisfaction	38.2	$<.001$	.52	***
Trust in Decision	29.4	$<.001$	.46	***
Perceived Actionability	44.7	$<.001$	.56	***
Comprehension	21.8	$<.001$	.38	***

*** $p < .001$; post-hoc: all pairwise Bonferroni $p < .01$

7.3 Comparison with Related Transparency Frameworks

Table 11 positions the Three-Layer Transparency Framework against four related proposals in the literature. Diakopoulos' [9] algorithmic accountability taxonomy addresses external audit audiences (journalists, regulators) rather than individual citizens, and does not include contestation support as a designed element. Wachter et al.'s [7] counterfactual explanation framework provides a theoretically rigorous single explanation type (Layer II equivalent) but does not address process or contestation layers. Bell et al.'s [4] transparency playbook addresses organisational transparency obligations comprehensively but is a compliance document rather than a citizen-facing design resource. Busuioc's [8] accountability framework addresses institutional accountability mechanisms (answerability, enforcement, reviewability) at the policy level but does not prescribe interface design. The Three-Layer Transparency Framework uniquely occupies the intersection of citizen-facing design, participatory validation, and coverage of all three accountability dimensions.

7.4 Design Guidelines

Eight evidence-based design guidelines emerge from the convergent findings across all three phases.

G1 — Lead with contestation. Layer III information, particularly the appeal deadline and primary action button, should appear at the top of the Layer III expansion, not after statutory boilerplate. Workshop 2 feedback was unanimous on this point, and Phase 3 think-aloud data confirmed that participants who found the appeal button quickly showed

Table 11. Comparison of transparency frameworks for algorithmic public sector systems. Columns indicate whether each framework addresses the relevant dimension (Y = yes, P = partially, N = no).

Framework	Process	Rationale	Contestation	Citizen-facing	PD Validated	Year
Diakopoulos [9]	Y	Y	N	N	N	2016
Wachter et al. [7]	N	Y	N	P	N	2018
Binns et al. [2]	P	P	P	Y	N	2018
Busuioc [8]	Y	Y	Y	N	N	2021
Bell et al. [4]	Y	Y	P	N	P	2023
Aljuneidi et al. [6]	N	Y	P	Y	N	2024
This paper	**Y**	**Y**	**Y**	**Y**	**Y**	**2026**

substantially lower expressed anxiety.

G2 — Use specificity, not categories. Rationale explanations should state specific values ("your income of £1,240 is £180 below the £1,420 threshold") rather than categorical descriptions ("your income was too low"). All three workshop groups identified vague categorical language as the primary frustration with existing explanation attempts.

G3 — Include comparators even if counterintuitive. Comparison with similar cases was rated lower in Phase 1 survey but was highly valued in workshops once participants understood what it meant. The framing "31% of similar applications were approved" reduces the isolation effect that many participants described when receiving a rejection in isolation.

G4 — Attribute to a named human contact. Workshop 3 and Phase 3 think-aloud data both showed that attributing the decision to a named human reviewer who reviewed an algorithmic recommendation increased perceived fairness, consistent with Aljuneidi et al.'s [6] findings.

G5 — Use progressive disclosure, not everything at once. The Layer I–II–III expansion architecture was consistently preferred over presenting all information simultaneously, which participants in pilot testing described as "overwhelming." The progressive structure also reduced cognitive load for participants who needed only outcome confirmation.

G6 — Provide independent referral, not only organisational contacts. Layer III should signpost to independent advice organisations (Citizens Advice Bureau, Shelter, etc.) alongside the official appeal route. Phase 1 interviews identified lack of trust in organisational-only contact information as a key reason citizens did not pursue appeals.

G7 — Accommodate digital exclusion through multiple channels. Online-only transparency implementations exclude the 18% of Phase 1 survey respondents who rated their digital literacy as low, disproportionately those in the most vulnerable circumstances. The framework should be implementable in print and in accessible digital formats.

G8 — Iterate with affected populations, not surrogates. The methodological finding that independent convergence on a three-layer structure occurred in Workshop 1 paper-sketching, without prompting, among citizens who had personally received algorithmic decisions, validates the participatory approach and suggests that citizen co-designers surface genuine interface requirements that cannot be adequately proxied by domain experts, researchers, or civil servants alone.

Table 12. Summary of the eight design guidelines, their primary evidence source, and the framework layer each informs.

No.	Guideline	Primary source	Layer
G1	Lead with contestation	W2 feedback + P3 think-aloud	III
G2	Specificity, not categories	All three workshops	II
G3	Include comparator cases	W3 + P1 framing analysis	II
G4	Named human attribution	W3 + P3 think-aloud	I–II
G5	Progressive disclosure	Pilot testing + P3 ratings	I–III
G6	Independent third-party referral	P1 interviews	III
G7	Multi-channel delivery	P1 digital-literacy data	I–III
G8	Iterate with affected citizens	W1 convergence finding	—

Table 12 provides a concise reference summary of the eight guidelines with their primary evidence source and the framework layer each addresses.

7.5 Implementation Pathway for Public Sector Organisations

Translating the framework into production interfaces requires engagement with existing procurement, legal, and data governance processes that vary across jurisdictions and organisational contexts. Based on Phase 2 civil servant input and broader public sector technology implementation literature, we identify four key implementation stages.

Stage 1 — Compliance audit. Organisations should map existing decision notice content against the Layer I requirements: data provenance, decision date, responsible team, and reference identifiers. Many organisations already hold this data in case management systems but do not surface it in citizen-facing outputs. Layer I implementation is therefore primarily a data surfacing and presentation problem,

not a new data collection problem, and can in many cases be achieved through template revision without system re-engineering. The EU AI Act [1] transparency obligations for high-risk AI systems provide a regulatory mandate for this first stage.

Stage 2 — Rationale generation. Layer II requires the algorithmic system itself to expose its decision factors and weights in a form that can be translated into plain-language citizen-facing output. For rule-based systems this is straightforward; for machine learning systems it requires integration of an explanation generation layer (SHAP values, LIME, or counterfactual generation following Wachter et al. [7]). The key design constraint is that the explanation must be specific (not categorical), bounded in length (G2), and accompanied by a comparison element (G3). Civil servants interviewed in Phase 1 expressed concern that rationale exposure increases appeal rates; Phase 3 data suggest this concern may be misplaced, as perceived fairness and trust increased with Layer II, potentially reducing adversarial appeals even as it enables legitimate ones.

Stage 3 — Contestation pathway integration. Layer III requires integration between the decision notice and the appeal management system. In many public sector organisations these are separate systems with separate procurement and maintenance cycles. The minimum requirement is a direct link to the correct appeal form and a pre-populated email or letter template with the decision reference and deadline; the aspiration is a fully integrated contestation workflow. G6 (independent referral) can be implemented without systems integration by maintaining an up-to-date signposting list of relevant voluntary sector organisations.

Stage 4 — Accessibility and channel equity. The final implementation stage addresses the digital exclusion concern raised in Phase 1 interviews and formalised in G7. This requires parallel development of print-ready decision notices incorporating the three-layer content, telephone-accessible scripts for frontline staff, and accessible web versions meeting WCAG 2.1 AA standards. The progressive disclosure architecture of the framework is compatible with screen reader and alternative format delivery, as each layer corresponds to a discrete, labelled content region.

7.6 Implications for Emerging Regulatory Frameworks

The EU AI Act [1] classifies as "high-risk" any AI system used for determining access to essential public services, including benefits, housing, and social care support. Article 13 of the Act imposes transparency obligations on high-risk system providers, requiring that users receive "sufficient information to interpret the system's output and use it appropriately." Article 14 imposes human oversight requirements, including the obligation to ensure that responsible natural persons can "understand the capacities and limitations" of the AI system they supervise. The Three-Layer Transparency Framework maps directly onto these obligations at the citizen interface level: Layer I addresses the information-sufficiency requirement of Article 13; Layer II implements the interpretability obligation; Layer III enacts the human oversight signposting requirement of Article 14 by attributing the decision to a named reviewing official and providing a structured route to human reconsideration.

However, the AI Act's obligations are directed at system providers and deployers, not at the design of citizen-facing outputs. A practical gap therefore exists between regulatory compliance (a system *produces* the information required by Articles 13–14) and citizen accessibility (a citizen *can access and act on* that information through the decision notice they receive). The present study provides evidence that this gap is substantial: the existing system studied in Phase 3 may satisfy formal compliance requirements—it refers citizens to an appeal leaflet—without providing meaningful contestation support as demonstrated by the 11% appeal button usage versus 81% in the Full Framework condition. Regulatory frameworks and national implementation guidance should therefore specify minimum interface design standards, not merely information content requirements, for citizen-facing algorithmic decision outputs in high-risk public sector contexts.

7.7 Limitations

The Phase 1 survey was distributed through voluntary sector partner organisations, which may over-represent citizens already engaged with the system and under-represent those who never contest decisions or seek advice. The fictional decision scenario used in Phase 3 evaluation was modelled on Universal Credit but was not an actual live interface, limiting the ecological validity of the evaluation results. The co-design workshops used participants from a single country context (United Kingdom), and the regulatory and cultural context of algorithmic governance varies substantially across jurisdictions; the framework's core layers may transfer, but the specific content of Layer III will require localisation. Finally, the study did not include participants with significant cognitive disabilities or low literacy; future work should explicitly address accessible transparency design for these populations.

8. CONCLUSION

This paper reported a three-phase research programme that produced, iteratively refined, and empirically validated a Three-Layer Transparency Framework for citizen-facing algorithmic decision interfaces in the public sector. Beginning from a citizen needs assessment rather than from regulatory requirements or technical explainability methods, the framework organises transparency into process disclosure (Layer I), rationale explanation (Layer II), and contestation support (Layer III), with a progressive disclosure architecture that serves both citizens who need a simple outcome confirmation and those who need a full account of the algorithmic logic and their options for response.

The participatory methodology produced a strong empirical validation: citizen sketch groups in Workshop 1 independently converged on a three-layer structure without prompting; the Phase 1 priority ranking placed contestation as the highest-rated transparency need ($M = 4.82$); and Phase 3 evaluation confirmed that each additional layer provides a statistically significant increment to citizen trust, comprehension, actionability, and transparency satisfaction. The Full Framework condition achieved a mean transparency satis-

faction rating of 6.1/7 compared with 2.8/7 for the existing opaque standard. Eight actionable design guidelines, a replicable participatory methodology, and an interface feature taxonomy (Table 7) together constitute a practical resource for organisations designing or procuring public sector algorithmic decision interfaces. The framework's direct mapping onto EU AI Act Articles 13 and 14 makes it particularly timely as public sector organisations across Europe prepare for compliance with high-risk AI system obligations.

The broader implication is that the opacity of current public sector algorithmic decision interfaces is a design choice, not a technical necessity. Citizens are capable of engaging with layered, specific, and contextualised algorithmic explanations when they are presented clearly, and they are substantially more trusting of systems that provide them. The Three-Layer Transparency Framework provides both the conceptual basis and the empirical evidence needed to make that design choice in favour of accountability.

Future work should examine: (a) longitudinal effects of layered transparency on citizens' actual appeal behaviour in live deployed systems; (b) the framework's transferability to domains such as planning decisions, school admissions, and criminal risk assessment; (c) adaptive implementations for citizens with low digital literacy via print-equivalent formats; and (d) the relationship between explanation granularity and appeal success rates, to determine whether higher transparency changes administrative outcomes as well as citizen perceptions.

REFERENCES

[1] European Parliament and Council, "Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," 2024, official Journal of the European Union, L 1689.

[2] R. Binns, M. Van Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. ACM, 2018, doi: 10.1145/3173574.3173951.

[3] S. Grimmelikhuijsen, "Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making," *Public Administration Review*, vol. 83, no. 2, pp. 241–262, 2023, doi: 10.1111/puar.13483.

[4] A. Bell, O. Nov, and J. Stoyanovich, "Think about the stakeholders first! Toward an algorithmic transparency playbook for regulatory compliance," *Data & Policy*, vol. 5, p. e12, 2023, doi: 10.1017/dap.2023.12.

[5] XAI CHI 2024 Authors, "Explorable explainable AI: Improving AI understanding for community health workers in India," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. ACM, 2024, doi: 10.1145/3613904.3642733.

[6] S. Aljuneidi, W. Heuten, L. Abdenebaoui, M. K. Wolters, and S. Boll, "Why the fine, AI? The effect of explanation level on citizens' fairness perception of AI-based discretion in public administrations," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. ACM, 2024, doi: 10.1145/3613904.3642302.

[7] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," *Harvard Journal of Law & Technology*, vol. 31, no. 2, pp. 841–887, 2018, doi: 10.2139/ssrn.3063289.

[8] M. Busuioc, "Accountable artificial intelligence: Holding algorithms to account," *Public Administration Review*, vol. 81, no. 5, pp. 825–836, 2021, doi: 10.1111/puar.13293.

[9] N. Diakopoulos, "Algorithmic accountability: Journalistic investigation of computational power structures," *Digital Journalism*, vol. 3, no. 3, pp. 398–415, 2016, doi: 10.1080/21670811.2014.976411.

[10] M. J. Muller, "Participatory design: The third space in HCI," in *Human-Computer Interaction: Development Process*, J. A. Jacko and A. Sears, Eds. CRC Press, 2003, pp. 1051–1068.