Consent, Explanation, and Appeals in AI-Mediated Services: Designing for Understanding, Fairness, and Recourse

Akshaya Jayasankar
Email: akshayajayasankar01@gmail.com

***Abstract*—AI-mediated services—such as voice IVRs, chatbots, and automated decision assistants— increasingly handle tasks that affect access to benefits, credit, and support. Yet users often have limited insight into what they are consenting to, why a recommendation was made, or how to contest an outcome. This paper outlines a concept study on *consent, explanation, and appeals (CEA)* in AI-mediated services. We propose a concrete experimental scenario, a set of interface variants, and a synthetic-data pipeline to explore how different designs shape users' understanding, perceived fairness, and willingness to appeal.**

I. Introduction

AI-mediated services are becoming the front door for many routine and high-stakes processes, from customer support to eligibility checks and dispute resolution. Users increasingly interact with chatbots, voice assistants, or web-based decision flows instead of speaking directly to a human agent. While these systems can scale access and reduce waiting times, they also introduce new risks: people may click through consent flows they do not understand, receive opaque recommendations, and be unsure how to challenge an outcome they perceive as unfair.

Existing work on explainable AI and algorithmic decision-making has focused heavily on model-focused explanations or transparency reports [1]–[3]. In many real-world services, however, what the user encounters is a *conversation or flow*: a sequence of prompts, notices, explanations, and options to escalate or appeal. The design of that flow shapes whether consent is meaningful, whether explanations are useful, and whether users feel they have recourse.

This concept paper asks: *How do different designs for consent, explanation, and appeals in AI-mediated services affect users' understanding, perceived fairness, and willingness to contest decisions?* I focus on a stylized but realistic scenario: a benefits-eligibility chatbot that screens users for a support program and offers a recommendation (provisionally eligible, ineligible, or needs more documentation).

To make this question tractable, I propose an experimental design with synthetic data rather than real user records. The idea is to first build an *evidence-ready* prototype: a small-scale environment where interface variants, logging, and metrics are carefully specified, so that future pilots can be evaluated credibly and audited if needed.

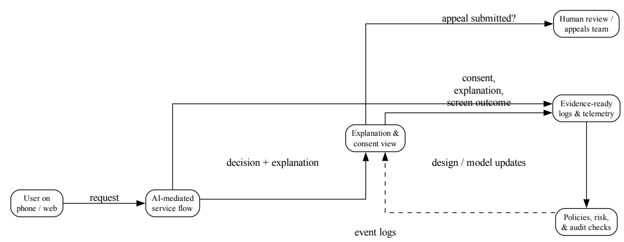

Fig. 1. Conceptual flow for consent, explanation, and appeals (CEA) in an AI-mediated service. The evaluation uses synthetic logs that mirror this pipeline without touching real user data.

The paper makes three contributions:

- It articulates a CEA design space for AI-mediated services, emphasizing consent flows, explanation styles, and appeal pathways.
- It outlines a concrete experimental scenario with interface conditions, measures, and synthetic logs that can be implemented and iterated on without touching real user data.
- It sketches an artefact bundle (code, synthetic data, and documentation) that can support future preregistration and field studies.

II. Background and Related Work

A. AI-Mediated Services and Algorithmic Decisions

Many organizations now deploy AI-mediated channels—chatbots, virtual agents, and automated decision assistants—as gatekeepers for information and services. These systems often make or support decisions about eligibility, prioritization, and next steps. Prior work on algorithmic decision-making highlights concerns about opacity, bias, and accountability, especially when decisions affect people's opportunities [2]–[5].

In practice, users experience these systems through interfaces, not models: they see consent checkboxes, short notices, explanation snippets, and occasionally links to appeal. The design of those interfaces shapes trust, perceived legitimacy, and whether people feel they can safely disagree with the system.

B. Consent and Data Use in AI Systems

Legal and policy debates discuss the "right to explanation" and informed consent in automated decision-making [2], [6]. Under regulations such as the GDPR, organizations must provide information about automated processing and, in some cases, offer meaningful safeguards and human review. However, user studies suggest that many consent flows function more as formalities than as meaningful opportunities for understanding and choice. Long, dense privacy notices and single "I agree" buttons often lead to superficial engagement.

In the context of AI-mediated services, there is an opportunity to design *layered consent*: brief, plain-language summaries with the option to expand for more detail; explicit statements of what data will be processed; and clear opt-out boundaries. This study treats consent as a design variable that can be systematically varied and evaluated.

C. Explanations and User Understanding

Explainable AI research proposes different forms of explanations for model outputs, such as feature importance lists, example-based explanations, or contrastive messages [1], [7], [8]. User-centered work emphasizes that explanations must match people's goals and mental models, not only describe internal model mechanics [9].

In service settings, explanations may need to answer questions like: "Why was I told I am likely ineligible?" or "What information changed the recommendation?" Explanations that are too technical or vague can reduce satisfaction and perceived fairness [3]. Conversely, well-designed explanations can support understanding and appropriate trust, especially when combined with clear information about uncertainty and limits.

D. Appeals, Procedural Justice, and Recourse

Procedural justice theory argues that people are more likely to accept decisions—even unfavorable ones—when they perceive the process as fair, understandable, and featuring a meaningful opportunity to voice concerns and seek review [10]. For algorithmic systems, appeals and contestation mechanisms are central to procedural justice: users should know *how* to challenge an outcome and *who* is responsible for reviewing it.

The recourse literature in machine learning similarly stresses that individuals should have actionable ways to change outcomes or request reconsideration [11]. Sociotechnical critiques of algorithmic governance highlight the need for contestability and human responsibility, not just accuracy [4], [12]. In many AI-mediated services, however, the appeal path is hidden behind generic "contact support" links or not surfaced at all. This study treats the visibility and structure of appeal options as an experimental factor.

E. Summary of Gaps

Across these literatures, three gaps motivate this concept study:

- Many discussions of consent and explanation are abstract or policy-focused; fewer provide concrete, testable designs in realistic service scenarios.
- Appeals and recourse are often mentioned as principles but rarely implemented and evaluated at the interface level.
- There are few end-to-end prototypes where consent, explanation, and appeals are integrated, instrumented, and evaluated together, even with synthetic data.

F. Design Lenses for CEA Interfaces

Building on this work, the proposed study uses three design lenses:

- **Transparency calibrated to user goals.** Transparency should support users' tasks and decisions rather than overwhelming them with technical detail [9], [13]. Layered consent and explanations are one way to offer more depth without forcing everyone to read everything.
- **Contestability and recourse as first-class features.** Recourse should be built into the interaction flow, not treated as an afterthought [4], [11]. This implies visible, understandable appeal options that connect to real human review.
- **Sociotechnical alignment.** Fairness and accountability depend on how systems, institutions, and people interact [5], [14]. Interface designs must therefore be evaluated not only for usability but also for how they shape responsibility, trust, and power.

These lenses guide the design space defined in Section III and the metrics used to evaluate different consent, explanation, and appeal configurations.

III. Proposed Study Design

A. Scenario and System Overview

The focal scenario is an online benefits-eligibility assistant. Users answer a small set of questions about their situation (e.g., income band, household composition, region). The AI-mediated service returns a recommendation:

- **Provisionally eligible** (can proceed to full application),
- **Probably ineligible** (based on current information), or
- **Needs additional documentation**.

In a real deployment, this assistant might support human caseworkers or act as a first filter. In the synthetic prototype, we simulate user attributes, outcomes, and interaction logs.

The interface flow has four key stages (Figure 1):

1) **Consent**: users are informed about data collection and use.
2) **Interaction**: users answer screening questions.
3) **Explanation**: the system provides a brief explanation of the recommendation.
4) **Appeal**: users may request a review or escalate to a human agent.

We vary the design of consent, explanation, and appeals across experimental conditions.

TABLE I
EXPERIMENTAL CONDITIONS IN THE CEA DESIGN SPACE.

Cond.	Consent	Explanation	Appeal Path
C1	Minimal	None	Hidden link
C2	Layered	Brief reason	Hidden link
C3	Layered	Layered (Why/How)	Simple button
C4	Layered	Layered	Structured appeal

B. Experimental Conditions

Table I summarizes the main conditions. For a first experiment, a between-subjects design with four arms is sufficient.

- **Consent (Minimal vs. Layered).** Minimal consent presents a short, generic notice and a single checkbox. Layered consent presents a one-sentence summary plus expandable sections for data use, retention, and sharing.
- **Explanation (None vs. Brief vs. Layered).** In the *none* condition, the system provides only the outcome label. In *brief* conditions, it adds a one-sentence reason. In *layered* conditions, users can expand to see factors that mattered and simple contrastive "what if" statements.
- **Appeal Path (Hidden vs. Simple vs. Structured).** A hidden path is a small "contact support" link in the footer. A simple path is a visible "Request Review" button that opens a free-text form. A structured path offers a guided form with reasons for disagreement and space to provide additional information.

C. Measures

We propose three primary outcome families, collected via post-interaction survey items and behavior logs.

1) Understanding: Self-reported understanding of the decision and data use (e.g., "I understand why I received this recommendation"). We can approximate this with Likert items and simple comprehension questions about data use and next steps.

2) Perceived Fairness and Procedural Justice: Perceived fairness of both the outcome and process (e.g., perceived respect, voice, and neutrality) drawing on procedural justice measures [10]. Items assess whether users felt they could share relevant information and whether they believe their perspective was considered.

3) Willingness to Appeal and Actual Appeal Behavior: Stated willingness to appeal (e.g., "If I disagreed, I would feel comfortable requesting a review") and actual clicks on appeal options or completion of appeal forms. Appeal rates under different conditions are a key behavioral outcome.

D. Synthetic Data and Logging Plan

To avoid handling real eligibility data at this stage, we rely on synthetic data generation. A simulation script will:

- Sample synthetic users with attributes (e.g., age band, income band, household size).
- Assign them to experimental conditions (C1–C4).
- Generate simulated recommendations and survey responses based on simple response models (e.g., layered explanations increase understanding scores on average).

TABLE II
ILLUSTRATIVE SYNTHETIC RESULTS BY CONDITION (MEANS ON A 1–5 SCALE). THESE VALUES ARE PLACEHOLDERS GENERATED FOR DESIGN AND ANALYSIS TESTING, NOT EMPIRICAL FINDINGS.

Cond.	Understanding	Fairness	Appeal rate
C1	2.8	2.9	0.05
C2	3.4	3.3	0.06
C3	3.9	3.7	0.10
C4	4.1	3.9	0.14

Interaction logs will be stored in structured tables, for example:

- **fact_interactions:** one row per user-session, including condition, recommendation, whether explanation was viewed, whether appeal was initiated, and synthetic outcome ratings.
- **dim_users:** synthetic user attributes.
- **dim_conditions:** mapping of condition codes to consent/explanation/appeal settings.

Even with synthetic data, we can compute and visualize condition differences, run simple regression or ANOVA models, and test analysis code and logging pipelines end-to-end.

E. Analysis Plan (Illustrative)

With N synthetic users per condition, we can estimate:

- Mean differences in understanding and fairness scores between conditions (e.g., C1 vs. C3/C4).
- Effects of visible appeal paths on willingness-to-appeal scores and actual appeal behavior.
- Interaction patterns, such as whether explanations matter more when the recommendation is unfavorable.

The goal at this stage is not to claim real-world effect sizes but to ensure that metrics, visualizations, and analysis scripts are well specified and reproducible.

F. Example Synthetic Outcome Patterns

To make the analysis plan more concrete, Table II shows a simple synthetic outcome pattern that could be generated by the simulation script. Here, layered explanations and structured appeals (C3–C4) increase understanding scores and appeal usage relative to a minimal, opaque baseline (C1).

In the actual GitHub repository and any future OSF project, this table would be reproduced by analysis scripts that operate directly on logged synthetic data. This ensures that metrics, code, and documentation stay aligned and that the artefact bundle is ready to support preregistration and later field deployments.

IV. DISCUSSION AND NEXT STEPS

This concept paper proposes a small, focused study on consent, explanation, and appeals in AI-mediated services. By working initially with synthetic data and a stylized benefits-eligibility scenario, the project aims to develop:

- A concrete design space for CEA in service interfaces.
- A set of measures and logs that can support evaluation and audit.
- A reusable artefact bundle (code, data, and documentation) for future pilots.

A. Planned Artefacts and Reproducibility

To make the study evidence-ready, I plan to maintain a public repository containing:

- The LaTeX source and compiled PDF of this 3–4 page concept paper.
- Interface specifications (e.g., JSON or YAML describing consent, explanation, and appeal variants), and prompt templates for any AI components.
- Synthetic data generators and example synthetic datasets matching the logging schema.
- Analysis scripts that reproduce summary tables and figures from the synthetic experiment.

This mirrors best practices in reproducible AI research and creates a clear path to future preregistration and field studies.

B. Implementation Roadmap and Milestones

The work is structured as a sequence of milestones that progressively increase realism while keeping the project manageable and auditable:

Milestone 1: Design freeze and synthetic specification: Finalize the CEA design space (conditions C1–C4), survey items, and logging schema. At this stage, the focus is on aligning interface variants, measures, and synthetic data fields so that Table I and Table II can be fully reproduced from the planned logs.

Milestone 2: Synthetic prototype and analysis scripts: Implement the benefits-eligibility flow in a simple web prototype or scripted environment. Develop a synthetic data generator that produces interaction logs and survey outcomes under each condition, and write analysis scripts that recreate the illustrative patterns in Table II. Release this as a tagged version of the repository (e.g., v0.3–v0.5) with complete documentation.

Milestone 3: Evidence-ready artefact bundle: Consolidate the paper, code, synthetic data, and analysis scripts into an "evidence-ready" bundle. This includes a clear README, a data dictionary for log fields, and instructions for rerunning simulations and analyses. At this point, the project is prepared for preregistration and external review, even if only synthetic data have been used so far.

Milestone 4: Pilot and field collaboration (future work): In collaboration with an organizational partner, adapt the prototype to a real AI-mediated service (e.g., a customer-support or public-benefits context). Run a small-scale pilot with real users, using the same logging and analysis pipeline, and update the repository and documentation accordingly. A preregistered, larger-scale field study could then build on this pilot to estimate real-world effects of CEA design choices.

C. Limitations

Several limitations are important. First, synthetic data cannot capture the full complexity of real users' experiences, constraints, and emotional responses. Second, the stylized scenario focuses on one service type; other domains (e.g., credit scoring or hiring) may raise different concerns and require domain-specific design. Third, the initial study focuses on short-term outcomes; longer-term impacts on trust and behavior would require longitudinal designs.

Despite these limitations, a synthetic, evidence-ready prototype can be valuable. It provides a safe environment for stress-testing logging, metrics, and interfaces before involving real users, and it can help clarify which questions are worth investing in with richer studies.

V. Stakeholder Co-Design and Pilot Milestones

To keep the consent, explanation, and appeals (CEA) pipeline grounded in real service constraints, I plan a staged co-design process with practitioners and affected users.

First, I will conduct short, focused co-design workshops with product managers, operations leaders, and front-line agents. These sessions will map current consent and dispute-resolution flows, identify points where users commonly feel confused or powerless, and surface constraints around latency, call-handling time, and regulatory obligations. The synthetic experimental flows in this paper will be iteratively refined based on these workshops.

Second, I will run a small internal pilot using the synthetic dataset and prototype interfaces described above. A subset of employees will step through simulated scenarios in a lab-style environment, allowing me to (a) validate that log schemas are sufficient to reconstruct user journeys, (b) test the analysis scripts on realistic but non-sensitive data, and (c) refine the questionnaires for understanding, perceived fairness, and willingness to appeal.

Finally, contingent on partner interest and ethics approval, I will plan a limited field deployment with *opt-in* users in a non-critical service context (e.g., low-stakes recommendations or informational routing). The goal is not to optimize a live product immediately, but to stress-test the full CEA pipeline — from consent copy and explanation cards through to appeals handling and telemetry capture — under realistic operational conditions.

Table III summarises the major milestones and artefacts.

VI. Stakeholder Co-Design and Pilot Milestones

To keep the consent, explanation, and appeals (CEA) pipeline grounded in real service constraints, I plan a staged co-design process with practitioners and affected users.

First, I will conduct short, focused co-design workshops with product managers, operations leaders, and front-line agents. These sessions will map current consent and dispute-resolution flows, identify points where users commonly feel confused or powerless, and surface constraints around latency, call-handling time, and regulatory obligations. The synthetic

TABLE III
PLANNED MILESTONES AND ARTEFACTS FOR THE CEA STUDY.

Phase	Key Outputs
Scoping & Co-design	Process maps of current consent/appeals flows; refined interface variants; updated risk register.
Synthetic Pilot	Synthetic logs and analysis scripts; validated survey instruments; preliminary effect-size estimates.
Field Preparation	IRB/ethics approvals; data-protection impact assessment; finalized logging and redress policies.
Limited Field Trial	Anonymized evaluation dataset; mixed-methods report on understanding, fairness, and appeals behaviour.

experimental flows in this paper will be iteratively refined based on these workshops.

Second, I will run a small internal pilot using the synthetic dataset and prototype interfaces described above. A subset of employees will step through simulated scenarios in a lab-style environment, allowing me to (a) validate that log schemas are sufficient to reconstruct user journeys, (b) test the analysis scripts on realistic but non-sensitive data, and (c) refine the questionnaires for understanding, perceived fairness, and willingness to appeal.

Finally, contingent on partner interest and ethics approval, I will plan a limited field deployment with *opt-in* users in a non-critical service context (e.g., low-stakes recommendations or informational routing). The goal is not to optimize a live product immediately, but to stress-test the full CEA pipeline — from consent copy and explanation cards through to appeals handling and telemetry capture — under realistic operational conditions.

Table III summarises the major milestones and artefacts.

A. Conclusion

As AI-mediated services become common in public and private decision-making, users' experiences of consent, explanation, and appeals will shape not only satisfaction but also perceptions of legitimacy and fairness. By treating CEA as a designable, measurable space—and by building a small but rigorous prototype with synthetic data—this project aims to provide a grounded starting point for future empirical work and collaboration with organizations deploying AI-mediated services.

REFERENCES

[1] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[2] S. Wachter, B. Mittelstadt, and L. Floridi, "Why a Right to Explanation of automated decision-making does not exist in EU law," *International Data Privacy Law*, vol. 7, no. 2, pp. 76–99, 2017.

[3] 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 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–14.

[4] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 59–68, 2019.

[5] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. fairmlbook.org, 2019, online draft.

[6] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a "right to explanation"," *AI Magazine*, vol. 38, no. 3, pp. 50–57, 2017.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[8] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 93:1–93:42, 2018.

[9] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, and J. D. Weisz, "Expanding explainability: Towards social transparency in AI systems," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 397–407.

[10] T. R. Tyler, *Why People Obey the Law*, 2nd ed. Princeton University Press, 2006.

[11] B. Ustun, A. Spangher, and Y. Liu, "Actionable recourse in linear classification," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 10–19.

[12] A. Alkhatib and M. S. Bernstein, "Street-level algorithms: A theory at the gaps between policy and decisions," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[13] R. F. Kizilcec, "How much information? effects of transparency on trust in an algorithmic interface," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2390–2395.

[14] A. D. Selbst *et al.*, "Fairness and abstraction in sociotechnical systems," *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019.