The Role of Initial Acceptance Attitudes Toward AI Decisions in Algorithmic Recourse

Tomu Tominaga
NTT Human Informatics Laboratories
NTT Corporation
Yokosuka, Kanagawa, Japan
tomu.tominaga@ntt.com

Naomi Yamashita*
NTT Communication Science Laboratories
NTT Corporation
Keihanna, Japan
naomiy@acm.org

Takeshi Kurashima
NTT Human Informatics Laboratories
NTT Corporation
Yokosuka, Kanagawa, Japan
takeshi.kurashima@ntt.com

Abstract

Algorithmic recourse provides counterfactual suggestions to individuals who receive unfavorable AI decisions; the aim is to help them understand the reasoning and guide future actions. While most research focuses on generating reasonable and actionable recourse, it often overlooks how individuals' initial reactions to AI decisions influence their perceptions of subsequent recourses and their ultimate acceptance of the decision. To explore this, we conducted a user experiment ($N = 534$) simulating an automobile loan application scenario. Statistical analysis revealed that participants who initially reacted negatively to the AI decision perceived the recourse as less reasonable and actionable, reinforcing their negative attitudes. However, when the recourse was perceived as explaining decision criteria or proposing realistic action plans, participants' attitudes shifted from negative to positive. These findings offer design implications for recourse systems that enhance the acceptance of individuals negatively affected by AI decisions.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing methodologies** → *Artificial intelligence.*

Keywords

Algorithmic Recourse, Counterfactual Explanation, XAI, Human-AI Decision Making

ACM Reference Format:
Tomu Tominaga, Naomi Yamashita, and Takeshi Kurashima. 2025. The Role of Initial Acceptance Attitudes Toward AI Decisions in Algorithmic Recourse. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713573

1 INTRODUCTION

With the increasing integration of artificial intelligence (AI) into high-stakes decision-making processes and the growing demand for explainable AI technologies, algorithmic recourse has emerged as a promising field of study. Algorithmic recourse provides individuals who have received unfavorable outcomes from AI systems with counterfactual action plans [85], thereby illustrating the rationale behind AI decisions [18, 34, 85] and the necessary actions for achieving favorable outcomes in the future [41, 42]. For instance, it might inform a rejected loan applicant, "Your application would be approved if your annual income were $10,000 higher". The ultimate goal is to provide individuals with the means to remedy and overcome unfavorable decisions [40].

Leveraging the contrastive nature of counterfactual explanations [56, 57], previous research has mainly focused on developing recourses that are reasonable and actionable, where a reasonable recourse explains the decision rationale [85] and an actionable recourse suggests feasible changes [77]. Many technical solutions have been proposed [40, 82], including methods for generating multiple consistent recourses [60] and determining optimal action sequences [38]. Additionally, research on individual reactions to recourse found that people often seek further explanations when recourses are counterintuitive or illogical [78] and prefer recourses with easily implementable suggestions [87]. These methods assumed that recourse should be reasonable as an explanation and actionable as a plan to foster AI decision acceptance.

However, initial attitudes of decision subjects (i.e., those who receive AI decisions) toward AI decisions, prior to receiving the recourse, may influence how they perceive the subsequent recourse and affect whether they finally accept that decision. Indeed, recent studies on AI-assisted decision-making [1, 65, 66, 68] suggest that a negative first impression of AI can reduce trust and reliance on future AI decisions [17, 58, 75], even when the AI's performance is accurate [16]. We hypothesize that the initial acceptance attitude shapes the perception of subsequent recourses and influences the final acceptance of the AI's decision. Given that reasonability and actionability are considered key dimensions of recourse perception [69, 80], we address the following research question: **(RQ1)** ***How do decision subjects' initial acceptance of AI decisions affect their perceptions of reasonability and actionability of the subsequent recourse provided by the AI?*** The analysis provides insights for accurately identifying recourses that reliably lead to final acceptance. If initial attitudes influence subsequent perceptions, it indicates that some recourses favored by decision subjects with positive initial attitudes are dismissed by those with negative ones. Understanding these effects allows for excluding such recourses and uncovering ones even acceptable to decision subjects with negative initial attitudes. Later in this paper, we elucidate the connection between recourse perceptions and final acceptance through the

*Currently working at Graduate School of Informatics, Kyoto University

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713573

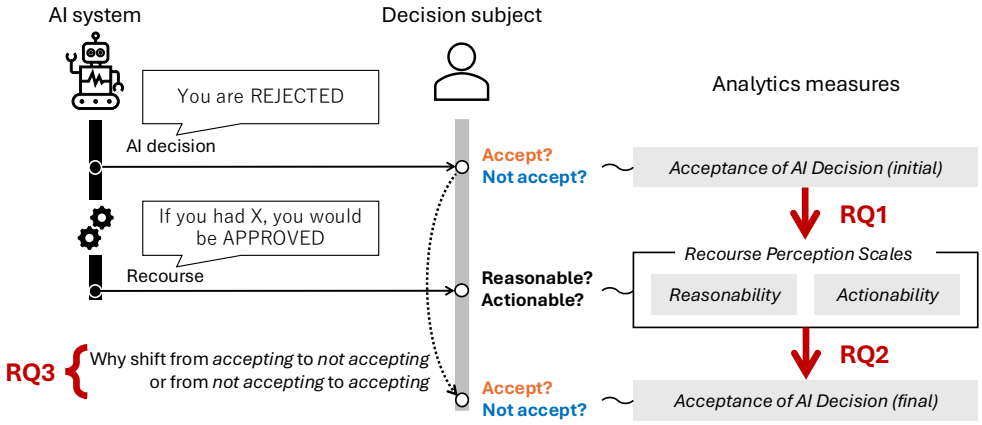


Figure 1: The scope of research questions in this study. In the experiment, five recourses are presented to each participant: participants provided one initial acceptance attitude rating, five sets of recourse evaluations, and five final acceptance attitude ratings corresponding to the recourses (see Section 3.3.3).

following RQ2, identify recourses that reverse acceptance attitudes based on findings of RQ1 and RQ2, and explore the characteristics of these recourses in the following RQ3. The relationship of the research questions in this study is illustrated in Figure 1.

We then examine how perceptions of recourse influence the final acceptance of AI decisions. Existing literature suggests that reasonable and actionable recourse provides a satisfactory explanation, allowing decision subjects to accept AI decisions [40]. However, the direct impact of these factors on acceptance attitudes has not been empirically investigated. To address this gap, we pose the following research question: **(RQ2)** ***How do decision subjects' perceptions of recourse reasonability and actionability affect their final acceptance of AI decisions?***

Finally, we investigate the recourse characteristics that contribute to changes in the subjects' attitudes toward AI decisions. Specifically, we analyze the types of recourse provided in cases where an initially negative attitude turned positive, or vice versa: **(RQ3)** ***What types of recourse lead decision subjects who initially did not accept AI decisions to finally accept them, or those who initially accepted AI decisions to finally not accept them?***

To investigate these research questions, we conducted a user study with 534 participants simulating an automobile loan scenario. While the applicability of this scenario to domains requiring distinct expertise or domain knowledge (e.g., healthcare, legal decisions) remains an open question for future research, we chose its scenario because financial credit assessment has been a focal area in numerous recourse studies [23, 39, 52, 77, 87]. To enhance the realism of the scenarios, we recruited participants genuinely interested in purchasing a car and asked them to submit their profile data for AI evaluation. After receiving the AI evaluation, participants assessed their acceptance of the AI's decision at two points: before and after the recourse was provided. They also evaluated the reasonability and actionability of the recourse and offered open-ended explanations of their assessments. Note that this study does not consider *feasibility*, the likelihood of a sequence of actions proposed by a recourse being observed in reality [67], which may cause ambiguity in how participants interpret recourses (Section 6.3.1) and may confound the relationship of recourse perceptions with acceptance attitudes (Section 3.2.5), but we took experimental measures to mitigate its impact and ensure the validity of our study (Section 3.2.4).

We found that initial attitudes toward an AI decision significantly influence perceptions of the subsequent recourse and, consequently, the final acceptance of that decision. Specifically, the more negative the initial attitude, the less reasonable and actionable the recourse was perceived, leading to a more negative final acceptance. However, acceptance attitudes shifted from initial to final based on how individuals interpreted the recourse. When the recourse was seen as clearly explaining the decision criteria, highlighting their shortcomings, and offering a plan that was achievable with some effort, the attitudes changed from negative to positive. Conversely, the attitudes shifted from positive to negative when the recourse was perceived as disregarding fairness or lacking relevance to real-world contexts.

The contributions of this study are three-fold:

(1) We provide empirical evidence that initial attitudes toward AI decisions significantly influence how decision subjects perceive the recourse offered by the identical AI. Specifically, negative attitudes lead to the recourse being viewed as less reasonable and actionable, resulting in a more negative final acceptance of the decision.
(2) We identified key recourse characteristics influencing acceptance attitudes, including clear explanations of decision criteria and personal shortcomings, proposals for plans feasible with a practical range, consideration of fairness, and alignment with individuals' daily lives, work, and societal contexts.
(3) Based on these findings, we presented design implications for systems that generate effective recourse to facilitate the acceptance of AI decisions and outlined future research directions to advance these systems.

2 RELATED WORK

2.1 Perceptions of AI Decision Making

As AI systems increasingly engage in complex tasks with high accuracy and participate in human decision-making, understanding how people perceive automated decisions has become a key focus of recent AI research. "Automated decision-making" here refers to the process where AI or machine learning models make evaluations or assessments based on predefined criteria, such as financial credit assessments [48, 59], medical diagnoses [5, 12], employment decisions [46, 71], and legal judgments [6]. Given that such high-stakes decisions significantly impact people's time, money, and lives, it is crucial that AI systems are transparent [15, 21], reliable [79, 93], unbiased and fair [20, 27, 55], and accountable [63].

Among factors affecting perceptions of AI systems such as the expertise required for the decision [28, 31], the nature of the task [13, 73], and AI performance [35], the roles of the person in the decision-making process are particularly significant [32, 50]. Key stakeholders include decision makers, who utilize or collaborate with AI in making judgments, and decision subjects, who receive judgments from AI [81]. Research focusing on decision makers' perspectives is notably extensive and of significant interest [1, 65, 66, 68]. In the context of AI-assisted decision making and Human-AI collaboration, researchers have investigated the process of building trust with decision makers [36, 53, 93] and the impact of algorithmic biases and errors [58]. They found that, if users perceive an AI system negatively due to errors or biases, their trust in the system diminishes significantly [75], leading them to avoid reliance on AI for decision support [16, 58]. Conversely, when users initially view AI as highly accurate, they are prone to accept its suggestions [61, 62, 91]. These findings are useful for developing AI decision-making systems that foster trust and confidence of decision makers.

While decision subjects interact with AI systems differently from decision-makers, they often want to understand how AI decisions affect them [23]. For instance, when given explanations, decision subjects may be concerned not only with whether the decisions are fair and reliable [47, 86], but also with whether the decisions are mechanical [47] and whether they have the opportunity to contest the suggestions [33, 85]. Although research on decision subjects is less extensive compared to that on decision-makers, some studies have explored design principles to address these concerns. Research on algorithmic review in AI decision-making shows that decision subjects prefer having explanations and the ability to contest decisions [92] and generally favor human reviewers over AI reviewers [51, 52]. Moreover, while decision subjects desire fairness in AI decisions, they are more likely to continue using the system if they expect it to deliver favorable outcomes [23]. These effects are particularly pronounced when there is a high level of distrust in existing systems [8] or when the decisions will significantly impact the subjects [51].

Inspired by these research findings, we hypothesize that, if decision subjects initially have a negative (positive) perception of AI decisions, their assessment of subsequent recourses and final acceptance of the decisions may worsen (improve). We argue that our research is novel in that it tests this hypothesis within the field of algorithmic recourse specifically focusing on decision subjects. In this investigation, it is important to target decision subjects because they are directly impacted by negative outcomes from AI systems and actively seek to overcome them. According to Girotto et al. [24], when individuals engage in counterfactual thinking following a negative outcome (e.g., being denied a car loan by a bank), they explore alternative solutions within the context (e.g., working longer hours for higher income) if they are actors directly involved, whereas they avoid facing the outcome by choosing not to tackle a given problem directly (e.g., opting not to apply for a loan) if they are external observers. In other words, actors attempt to modify the problem's available elements, unlike observers, who seek to alter the problem's underlying assumptions. Since algorithmic recourse is designed to help actors address such challenges, attitudes toward AI decisions and perceptions of recourse should be evaluated by actors, that is, decision subjects rather than observers. The process of selecting appropriate participants for our experiment (i.e., decision subjects) is described in Section 3.

2.2 Algorithmic Recourse and Its Fundamental Assumption

The field of algorithmic recourse began with Wachter's formulation of the problem in 2018 [85]. As it is an emerging field, its definition varies among researchers. For example, Joshi et al. [34] defines it as "an actionable set of changes a person can undertake to improve their outcome". Ustun et al. [77] describes it as "the ability of a person to achieve a desired outcome from a fixed model". More broadly, Venkatasubramanian and Alfano [80] defines it as "the systematic process of reversing unfavorable decisions made by algorithms and bureaucracies across various counterfactual scenarios". Although no single definition is universally accepted, the core philosophy of algorithmic recourse research is to develop methods that allow decision subjects to understand and overturn unfavorable decisions made by AI systems [40].

To this end, algorithmic recourse provides counterfactual suggestions to decision subjects, enabling them to understand the rationale behind AI decisions [18, 34, 85] and to act on these suggestions [41, 42]. Recent research has, therefore, concentrated on developing counterfactual suggestions that are both reasonable and actionable. From a technical perspective, this involves defining objective functions to assess the reasonability and actionability of recourses and optimizing these functions (see extensive reviews in [40, 82]). Researchers have proposed methods to generate recourses that ensure fairness [27, 84], provide diverse and consistent options [60], capture the distribution or causal relationships among suggestion items [37, 42, 67], and outline the sequence for implementing the suggestions [37]. Given the high computational cost of implementing recourse systems, research has also developed efficient and scalable techniques for recourse calculation [54, 83].

Despite significant technological advances, few studies have demonstrated the effectiveness of these developments [43]. This gap has prompted several user experiments that attempt to address the issue, and preliminary results have been reported [45, 76, 78, 88]. Typical experimental designs in previous studies involve training AI models with existing datasets [87], having participants role-play as data subjects extracted from these datasets or as fictional characters in specific scenarios [23, 51, 52, 87], and then asking them to assess the provided counterfactual suggestions. According to

a study simulating a loan application scenario, decision subjects requested additional explanations when presented with counterintuitive or illogical recourse options [78], because people generally expect AI systems to closely replicate human reasoning [78] and offer perfect and consistent responses [19]. Wang et al. [87] developed an interactive interface, GAM Coach, that allows decision subjects to select recourse options that fit their preferences. Their exploratory study of loan-related recourse, analyzing user logs and survey responses, revealed that users preferred recourses that they could easily implement at their discretion [87]. These results suggest that decision subjects tend to favor recourses that are both reasonable and actionable.

However, the underlying assumption—that decision subjects accept AI decisions whenever the recourse is reasonable and actionable—has not yet been tested. In other words, the connection between perceptions of recourse reasonability and actionability and the acceptance of AI decisions remains unclear. Therefore, we aim to validate this assumption. Through this validation, we seek to link final acceptance of AI decisions with perceptions of recourses and explore how initial acceptance attitudes influence final acceptance attitudes through recourse perceptions.

3 EXPERIMENT

We conducted an online experiment to investigate the relationship between decision subjects' acceptance of AI decisions and their perceptions of recourses. To replicate a realistic situation where decision subjects receive unfavorable outcomes, we used an auto loan application scenario and carefully selected participants appropriate for this context. Participants submitted their profile data, received a rejection notice, and then indicated whether they could accept the decision at that stage. Each subsequently assessed the recourse provided in terms of reasonability and actionability, wrote their evaluation reasons, and finally indicated whether they could accept the decision after considering the recourse.

This experiment involving human subjects was reviewed and approved by the external review board of the Public Health Research Center[1] (approval number: PHRF-IRB 24A0002). All procedures were conducted in accordance with the provided guidelines.

3.1 Scenario Design

This study employs a hypothetical scenario in which participants apply for an auto loan. Participants are instructed to imagine themselves visiting a financial institution and undergoing the loan application process. The auto loan scenario was selected due to its relevance in recourse research within the finance domain [39, 77]. It is also a critical area of study in high-stakes decision-making in both the ML and HCI communities [23, 52, 87]. Among services and goods financed through loans—such as education, housing, and business—cars are the most common, relatively expensive, and primarily owned and used by participants themselves. This makes the auto loan scenario familiar and relatable to participants, enabling them to easily imagine the situation, recognize the AI decision as high stakes, and consider their personal context.

The scenario is described as follows:

[1]https://www.phrf.jp/rinri/ (only in Japanese)

	Current profile		Ideal profile
...	...		...
Employment	Private company		Private company
Job Title	Employee	⇒	Manager
Years of Service	1-3 years	⇒	5-10 years
...	...		...

Figure 2: An example of recourse provided to a decision subject

> *Participants seek an auto loan amounting to one-third of their annual income. They visit a financial institution, submit their profile data for evaluation, and an AI system assesses their application. The AI determines their eligibility based on criteria derived from a vast internal database and ultimately rejects their loan request. To understand the rejection and identify actions necessary for future approval, they will review several recourse options generated by the AI.*

Here, we assumed that the financial institution has a policy of rejecting applicants seeking loans exceeding one-quarter of their annual income. Consequently, all participants' applications were denied. This policy was not disclosed to participants during the experiment.

In this scenario, the loan amount is set relative to each applicant's annual income to adjust the repayment burden among participants. For instance, if the loan amount were fixed at $50K, approval would require an annual income exceeding $200K for all applicants. This would result in significant variation: a participant earning $60K would face a much larger gap to the approval threshold compared to someone earning $180K, potentially leading to more burdensome recourse suggestions for the lower-income participant than the higher-income participant. To mitigate such disparities in perceived recourse burden and gather within-subjects insights into how participants evaluate recourse options (e.g., a participant who perceives a recourse as less burdensome is more likely to accept the decision outcome), we designed the scenario with loan amounts proportional to each participant's income.

3.2 Recourse Computation

Given a profile that was rejected by the AI, recourse is generally computed by identifying its contrastive profile from the AI's training database. Here, we refer to the given profile as the input sample and the contrastive profile from the AI's database as the counterfactual sample. The difference between these samples is the basis for constructing the recourse, which is then presented to the decision subjects. An example of this is illustrated in Figure 2 (current and ideal profiles correspond to input and counterfactual samples, respectively). The following sections will explain the general mechanism of this calculation and the steps for generating the recourses presented to participants in this experiment.

3.2.1 Basic Scheme. As mentioned in the related studies [40, 82], solving the following optimization problem is a fundamental and

straightforward approach for generating recourse: Given N dimensional features $X = X_1 \times \cdots \times X_N$, a subset $P(X) \subseteq X$ defined by specific conditions, and a pre-trained binary classification AI model (e.g., for loan assessment) $F : X \rightarrow \{\text{Approval}, \text{Rejection}\}$, the objective is to find a counterfactual sample x^* for an input sample x that receives an unfavorable decision (i.e., $F(x) = \text{Rejection}$). This problem can be formulated as follows:

$$x^* = \underset{x, x^* \in P(X)}{\operatorname{argmin}} \; d(x, x^*) \text{ subject to } F(x^*) \neq F(x) \tag{1}$$

Here, the objective function $d : X \times X \rightarrow \mathbb{R}_{>0}$ represents a distance metric that quantifies the degree of divergence between the input sample and the counterfactual sample, such as L_0 and L_1 norms.

The input sample x and the counterfactual sample x^* are N-dimensional vectors representing their profile data. In this experiment, the use of the distance function d is detailed in Section 3.2.2, while the roles of the constraints $F(x^*) \neq F(x)$ and the subspace $P(X)$ are mentioned in Section 3.2.3.

3.2.2 Distance Metrics. We use L_0 norm and L_1 norm as the distance function d to measure the divergence between the input sample x and the counterfactual sample x^*. These are most commonly used in prior research [40, 82]. These metrics are referred to as *Sparsity* and *Proximity*, respectively. *Sparsity* counts the number of features changed, while *Proximity* represents the Manhattan distance between the samples.

Specifically, we first compute the perturbation δ needed to transform x into x^* as follows:

$$\delta_i = \begin{cases} \mathbb{I}[x_i^* \neq x_i] & \text{if } x_i \in \mathrm{C}, \\ |x_i^* - x_i| / M_i & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbb{I}$ is the indicator function, C is a set of nominal variables (i.e., $i = 1, 2, 4, 15, 16$ in Table 1), and M_i is the maximum range for the i-th feature, normalizing each feature scale from 0 to 1.

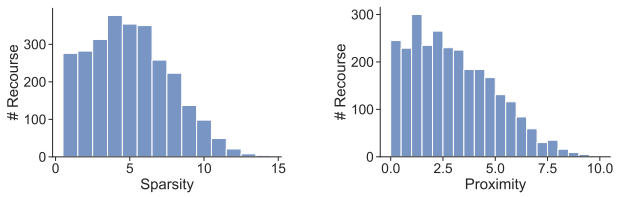
Using the perturbation δ, we compute *Sparsity* and *Proximity* as follows:

$$\begin{aligned} \text{Sparsity} &= \sum_i \mathbb{I}[\delta_i \neq 0], \\ \text{Proximity} &= \sum_i \delta_i. \end{aligned} \tag{3}$$

Smaller values indicate a recourse that is more sparse (i.e., fewer feature changes) and more proximate (i.e., closer).

3.2.3 Conditions of Counterfactual Samples. In this experiment, a counterfactual sample x^* for an input sample x is extracted from an existing pool of profile data [76]. The extraction process is governed by two key conditions. First, the annual income of the counterfactual sample x^* must be at least 4/3 times that of the input sample x. In this scenario, given that the financial institution follows the internal policy of rejecting applicants whose loan requests exceed one-quarter of their annual income, participants applying for a loan amounting to one-third of their annual income are rejected. Consequently, for a participant's application to be approved, the counterfactual sample must have an annual income at least 4/3 times greater than the participant's income. This requirement ensures that $F(x^*) = \text{Approval}$, while $F(x) = \text{Rejection}$ (i.e., $F(x^*) \neq F(x)$).

Second, both the input sample and counterfactual sample must adhere to the constraints related to features listed in Table 1. These constraints define the subspace $P(X)$ and ensure that the recourse does not involve changes to features that are immutable or realistically impossible. For example, features such as degree (#3), job title (#5), years of service (#6), or management experience (#7) can be increased but not decreased, as reducing these features is considered highly unrealistic. Similarly, removing personal experiences or skills—such as job change experience (#11), overseas work experience (#12), study abroad experience (#13), or best scores in ability tests (#14)—is also not feasible. Other features are not subject to these constraints. These conditions allow us to construct recourse within the theoretically possible range.



(a) Sparsity (b) Proximity

Figure 3: Frequency distribution of distance metrics for the selected recourses used for evaluation

3.2.4 Process for Selecting Counterfactual Samples to Construct Recourse. When an input sample is given, counterfactual samples meeting the conditions in Section 3.2.3 are selected as recourse candidates. Based on the prior finding that smaller recourse distances are associated with a greater propensity to act the recommended changes [76], we selected five counterfactual samples from the recourse candidates for each participant using the following procedure: one of the most sparse samples; one of the most proximate samples, other than the one selected in the first step; and three samples randomly chosen, other than those selected in the first two steps. This procedure assigned both short- and long-distance recourses to each participant, thereby balancing the distribution of recourse distances across participants and controlling for the influence of the distance metrics on participants' recourse evaluation. The distribution of distance metrics of the recourses selected using this procedure is shown in Figure 3.

3.2.5 Methodological Limitations. In computing recourses, we did not account for the feasibility of recourse, indicating the likelihood that the sequence of actions suggested by the recourse aligns with the distribution of the observed data [67]. Considering that feasibility may influence participants' recourse perceptions, if it is unevenly distributed across participants, it could introduce confounding biases that distort our results. For example, even if RQ1 reveals that participants with a negative initial acceptance attitude rate the recourse lower, presenting such participants with recourse of low feasibility in a biased manner makes it difficult to determine whether the low ratings are due to their negative initial attitude or the low feasibility of the recourse.

While we acknowledge the limitation of not directly observing feasibility, we emphasize that we minimized this potential impact

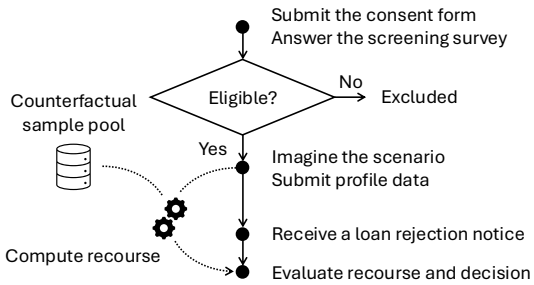


Figure 4: Overall experimental procedure

by randomly selecting a part of recourses for each participant (Section 3.2.4), promoting even distribution of recourse feasibility across participants. Although there remains room for further validation to achieve more refined experimental results, the experimental data is, nevertheless, sufficient to support the validity of our study.

3.3 Experimental Procedure

The experiment was conducted from February 7 to March 21, 2024 in Japan. The overall flow of the experiment is illustrated in Figure 4. Participants first read the experiment's instructions and submitted the consent form, followed by a screening survey to confirm eligibility. Eligible participants were then asked to imagine the experiment scenario and submit their profile data. Using this data, we constructed five recourses for each participant. Finally, participants received a loan rejection notice and evaluated their acceptance of the decision and their perception of the recourses presented to them.

3.3.1 Recruitment and Screening of Participants. Participants were recruited through the online research company ASMARQ[2]. A total of 550 Japanese individuals were gathered. The sample comprised 402 males (73.1%) and 148 females (26.9%), with an average age of 48.9±10.3 years. Participants who completed all surveys were compensated with 600 JPY[3]. The median completion times for the surveys were approximately 1 minute for the screening, 2 minutes for the profile data, and 15 minutes for evaluating AI decisions and recourses.

A screening survey was administered to ensure participants met the following four criteria: (1) currently employed by a private company or public institution, (2) interested in purchasing a car, (3) not holding a loan at the time of participation, and (4) having an annual income of less than 10 million JPY. The first criterion ensured the exclusion of students and focused on working individuals. The second criterion made the experiment more realistic by involving participants who were actually interested in buying a car. The third criterion avoided situations where participants with existing loans might consistently rate actionability as too low for any recourse. The fourth criterion ensured we could select five counterfactual samples per participant from the data pool[4].

[2]https://www.asmarq.co.jp/global/
[3]600 JPY is equivalent to 4.24 USD as of September 2024.
[4]If an applicant's income is 10 million JPY, the ratio of counterfactual samples that meet the conditions is only 1.47% in the data pool [76].

3.3.2 Submission of Profile Data. After confirming eligibility, participants submitted profile data necessary for evaluation based on the experimental scenario. As shown in Table 1, we selected 16 items for the profile data.

We used basic demographic information (#1-#3), current employment details (#4-#10), professional experience and skills (#11-#14), and personal network information (#15, #16). These items correspond to attributes found in loan-scenario user studies [76, 92], machine learning datasets for financial credit assessments [4, 29, 90], or typical resumes [9, 14], assuming that they relate to individuals' creditworthiness and employment stability.

We excluded immutable features such as gender, place of birth, and nationality, which cannot be changed by individual's choice or abilities [44]. Although these features might sometimes offer reasonable explanations, they always make recourses unactionable [40]. Additionally, due to the experimental design, annual income was omitted from the profile data. Participants were placed in a scenario where they were rejected due to insufficient income relative to the requested loan amount, which was not disclosed to them. Inspired by the idea that recourse is more useful when it covers a diverse range of counterfactuals [70, 85], we aimed to replicate scenarios where various suggested changes are included in the recourse. However, if annual income was included in the profile, the recourse would invariably suggest an increase in income, which significantly restricts its diversity. For example, for recourses with $Sparsity = 1$, the variation in the suggested items would be completely eliminated. To prevent this issue, annual income was not included as a profile data item for the recourse.

3.3.3 Evaluation of AI Decisions and Recourse. Participants were notified of their loan application rejection and subsequently evaluated both the decision and the recourse provided. The procedure followed these steps:

(1) Participants receive a brief overview of the experimental scenario, followed by the rejection notice.
(2) They indicate their **initial acceptance attitudes** toward the decision: *"Do you accept this decision on your loan application?"* (rated on a 7-point scale from strongly no to strongly yes).
(3) They are briefed on the purpose and details of the action plans (recourses) suggested by the AI.
(4) They receive five recourses and perform the following evaluations for each one. Since the recourses are presented in succession, they repeat these evaluations a total of five times.
   (a) They evaluate **recourse reasonability** and provide their evaluation reasons in an open-ended response: *"Do you consider the AI system's proposed plan to be a reasonable explanation for the rejection of your loan application?"* (rated on a 7-point scale: strongly no to strongly yes), and *"Why did you rate it this way?"*.
   (b) They evaluate **recourse actionability** and provide their evaluation reasons in an open-ended response: *"How difficult or easy do you find it to implement the AI system's proposed plan?"* (rated on a 7-point scale: very difficult to very easy), and *"Why did you rate it this way?"*.
   (c) They indicate their **final acceptance attitudes** toward the decision: *"Having seen the AI system's proposed plan,*

Table 1: Profile data items, options, and constraints. x_i and x_i^* are i-th features of input and counterfactual sample vectors (e.g., $x_3^* = 1$ means that the counterfactual sample's "Education" is a "High school"). The constrains column describes feature conditions of a counterfactual sample x^* to be selected for a suggested recourse given an input sample x.

#	Profile item (feature)	Option	Constraints
1	Residence	1. Tokyo / 2. Other than Tokyo	–
2	Type of residence	1. Owned house / 2. Rental housing	–
3	Education	1. High school / 2. Junior college / 3. University (bachelor) / 4. Graduate school (master) / 5. Graduate school (doctor)	$x_3^* \geq x_3$
4	Employment	1. Private company / 2. Public institution	–
5	Job title	1. Employee / 2. Supervisor / 3. Section head / 4. Section chief / 5. Assistant general manager / 6. Manager / 7. General manager / 8. Executive director / 9. Senior executive director / 10. President	$x_5^* \geq x_5$
6	Years of service	1. 0-1 year / 2. 1-3 years / 3. 3-5 years / 4. 5-10 years / 5. 10-20 years / 6. 20- years	$x_6^* \geq x_6$
7	Management career	1. No / 2. 0-1 year / 3. 1-3 years / 4. 3-5 years / 5. 5-10 years / 6. 10-20 years / 7. 20- years	$x_7^* \geq x_7$
8	Daily working hours	1. 0-2 hours / 2. 2-4 hours / 3. 4-6 hours / 4. 6-8 hours / 5. 8-10 hours / 6. 10-12 hours / 7. 12- hours	–
9	Daily remote working hours	1. 0-2 hours / 2. 2-4 hours / 3. 4-6 hours / 4. 6-8 hours / 5. 8-10 hours / 6. 10-12 hours / 7. 12- hours	–
10	Number of side jobs	1. No / 2. 1 job / 3. 2 jobs / 4. 3 jobs / 5. 4 jobs / 6. 5- jobs	–
11	Job change experience	1. No / 2. Yes	$x_{11}^* \geq x_{11}$
12	Overseas work experience	1. No / 2. Yes	$x_{12}^* \geq x_{12}$
13	Study abroad experience	1. No / 2. Yes	$x_{13}^* \geq x_{13}$
14	Best TOEIC† score	1. No / 2. 10-400 / 3. 400-495 / 4. 500-595 / 5. 600-695 / 6. 700-795 / 7. 800-895 / 8. 900-990	$x_{14}^* \geq x_{14}$
15	Facebook use	1. No / 2. Yes	–
16	LinkedIn use	1. No / 2. Yes	–

†This stands for the **T**est **O**f **E**nglish for **I**nternational **C**ommunication (TOEIC®) Listening & Reading Test (https://www.iibc-global.org/english/toeic/test/lr.html), one of the most widely recognized English proficiency tests in Japan, commonly used for admissions and job interviews. Scores range from 10 to 990, in increments of 5 points.

do you now accept this decision on your loan application?" (rated on a 7-point scale from strongly no to strongly yes).

In step 4b, we also asked whether the suggested changes were immutable for participants. As noted in Section 3.3.2, recourses involving immutable changes are not actionable. However, this ultimately depends on the individual. Therefore, we directly asked participants to assess immutability, and any recourse deemed immutable was excluded from the analysis.

Through these steps, participants provided one initial acceptance attitude rating, five sets of recourse evaluations, and five final acceptance attitude ratings corresponding to the recourses. To minimize the effect of recourse presentation order on the final acceptance ratings, five recourses were presented in a randomized order. This within-subject design allows for examining participants' reactions to different recourses and the resulting acceptance attitudes toward the AI decision. For more details of the survey flow and questions, please refer to the supplementary materials.

4 ANALYSIS

A total of 2750 data points were collected from 550 participants regarding their acceptance of the AI decision and their perceptions of the recourse reasonability and actionability. Of these, 231 data points were excluded because participants indicated that the suggested profile changes were immutable, leaving 2519 data points from 534 participants for further analysis. For the data availability, please refer to the supplementary materials.

4.1 Impacts of Initial Acceptance on Perceived Reasonability and Actionability (RQ1)

To examine the impact of initial acceptance of AI decisions on the perceived reasonability and actionability of recourse, we utilize Generalized Additive Models (GAMs). This non-parametric regression method with nonlinear basis functions captures complex relationships between explanatory and objective variables through smoothing splines. For the mathematical structure of the model, please refer to Appendix A.1.1.

For RQ1, the explanatory variable is initial acceptance, while the objective variable is either perceived reasonability or perceived actionability. Consequently, two distinct smoothing spline functions were constructed using the GAM. An F-test was performed within the model to assess whether initial acceptance significantly influences perceived reasonability or actionability.

We also conducted a post hoc power analysis. To assess the power of analytical models involving nonlinear functions like GAM, we used two approaches. The first approximated power using a calculation method of linear multiple regression models (via G*Power), and the second estimated the probability of the GAM's predictors achieving the significance level through repeated simulations (via RStudio). In the first approach, the effect size was based on the

adjusted R-squared values of the GAM. In the second, we repeated the process of generating simulated data, fitting the GAM to the data, and testing the statistical significance of its predictors, and measured the probability of achieving the significance level. Hereafter, let the power calculated from the first approach be denoted as P_{lmr} and that from the second as P_{sim}. In both cases, the significance level was $\alpha = 0.01$.

4.2 Impacts of Perceived Reasonability and Actionability on Final Acceptance (RQ2)

This analysis uncovers the impact of perceptions of recourse reasonability and actionability on the final acceptance of AI decisions. Given that perceived reasonability, perceived actionability, and final acceptance are paired data repeatedly measured from a single participant, we utilized Generalized Additive Mixed Models (GAMMs) to derive within-subject insights (e.g., if a decision subject perceives recourse as more reasonable or actionable, she/he is more likely to accept the AI decision). GAMM extends the generalized additive model (GAM) by incorporating mixed effects, accounting for the individual subject-specific effects. For the mathematical structure of the model, please refer to Appendix A.1.2.

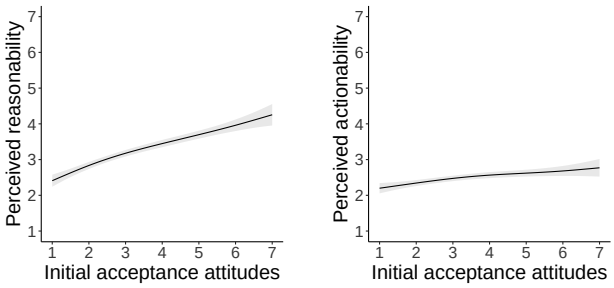
In this GAMM, the objective variable is final acceptance, and the explanatory variables are perceptions of reasonability and actionability. As with RQ1, we performed an F-test to determine whether explanatory variables significantly affect the final acceptance and confirmed its statistical power P_{lmr} and P_{sim}.

4.3 Properties of Recourses Affecting Shift in Acceptance Attitudes (RQ3)

RQ3 aims to identify the properties of recourses from the perspective of decision subjects when their acceptance attitude changes. We analyzed the open-ended responses explaining the reasons for evaluating the recourse that was presented when acceptance attitudes changed from negative to positive or from positive to negative. These patterns of changes in acceptance attitudes are called NP group (negative to positive) and PN group (positive to negative) hereafter.

Since initial and final acceptance attitudes are assessed on a 7-point scale, we categorized them as negative if the rating is 3 or below and as positive if it is 5 or above. Here, ratings of 4 were considered neutral and excluded. This yielded 307 recourse evaluation data from 97 participants for the PN group and 321 recourse evaluation data from 145 participants for the NP group. Among all participants (534), 267 finally accepted the decision, while 448 did not. It should be noted that each participant evaluated their final acceptance attitudes for five recourses, potentially exhibiting both positive and negative attitudes.

For the open-ended responses to *"Why did you rate it this way?"* on these recourses, we performed a qualitative analysis using reflexive thematic analysis [7]. One author generated initial codes from participants' responses, which were then reviewed and refined through discussions with another author to group them into themes. Afterwards, we asked two external researchers to code the reasons for the evaluation of reasonability and actionability, respectively (four coders in total), and confirmed their substantial agreement for each (Cohen's kappa $\kappa = 0.64, 0.61$). This qualitative analysis provides a detailed understanding of the recourse characteristics from the decision subjects' perspective that lead to changes in acceptance attitudes.



(a) The smoothing spline function of initial acceptance attitudes against perceived reasonability obtained by the GAM fitting.

(b) The smoothing spline function of initial acceptance attitudes against perceived actionability obtained by the GAM fitting.

Figure 5: Smoothing spline functions derived from the generalized additive models (GAMs) explaining perceptions of reasonability (a) and actionability (b) based on initial acceptance of AI decisions. The black curve represents the mean, and the gray shading shows the 95% confidence interval.

At the same time, we acknowledge that readers of this study may be concerned with the consistency of our analysis results (RQ3). In response to this, we asked two external researchers (for a total of four coders) to independently code the free-text responses concerning the reasons for evaluating reasonability and actionability, using the codebook agreed upon by the authors. The results revealed Cohen's kappa values of 0.64 and 0.61, indicating substantial agreement between the coders.

5 RESULTS

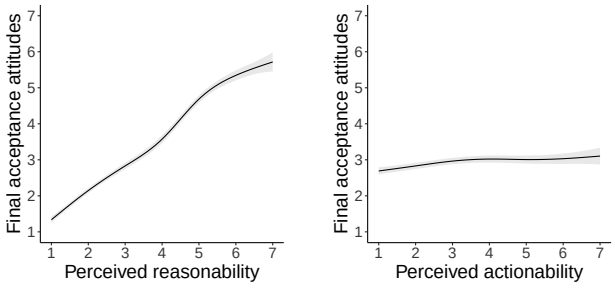
5.1 Impacts of Initial Acceptance on Perceived Reasonability and Actionability (RQ1)

Table 2 shows the statistical values drawn from the generalized additive models (GAMs). We found that initial acceptance has a statistically significant impact on both perceived reasonability ($F = 61.08$, $p < 0.001$, $R^2 = 0.069$, $P_{\text{lmr}} > 0.80$, $P_{\text{sim}} > 0.80$) and actionability ($F = 9.78$, $p < 0.001$, $R^2 = 0.009$, $P_{\text{lmr}} > 0.80$, $P_{\text{sim}} > 0.80$). Figure 5 illustrates the smoothing spline functions of each GAM. Perceived reasonability increases with more positive initial acceptance and decreases with more negative initial acceptance. While perceived actionability also shows a positive correlation with initial acceptance, its impact is less pronounced compared to perceived reasonability. Specifically, when comparing the output values of the smoothing spline functions for initial acceptance scores of 1 and 7, the function of perceived reasonability shows a difference of approximately 1.85, while that of perceived actionability displays a difference of about 0.57.

Overall, initial acceptance of AI decisions significantly influences both perceptions of reasonability and actionability, with a stronger effect on the former. This suggests that those who find it hard

Table 2: Statistical values obtained from the generalized additive models (GAMs) explaining perceived reasonability or actionability from the initial acceptance attitudes toward AI decisions (*Smth*: Whether or not the variable is a smoothing term. *Coef, S.E., t value*: Coefficient values, standard errors, and t values of the intercept. *EDF*: Effective degrees of freedom of the smoothing term. Higher EDF implies more complex, wiggly splines. When the EDF value is close to 1, it is close to being a linear term. *F value*: The statistical value from the F-test to verify whether the smoothing term is equal to zero. If it is significant, the smoothing term is an influential variable. *AIC*: Akaike information criteria.)

Objective variable	Explanatory variable	Smth	Coef	S.E.	t value	EDF	F value	p value	AIC
Perceived reasonability	Intercept		3.23	0.03	96.53			<0.001	9762.12
	Initial acceptance attitudes	✓				2.43	61.08	<0.001	
Perceived actionability	Intercept		2.48	0.03	81.41			<0.001	9283.54
	Initial acceptance attitudes	✓				1.96	9.78	<0.001	



(a) The smoothing spline function of perceived reasonability against final acceptance attitudes obtained by the GAMM fitting.

(b) The smoothing spline function of perceived actionability against final acceptance attitudes obtained by the GAMM fitting.

Figure 6: Smoothing spline functions derived from the generalized additive mixed models (GAMMs) explaining final acceptance attitudes toward AI decisions based on perceptions of reasonability (a) and actionability (b), accounting for individual subject-specific effects. The black curve represents the mean, and the gray shading shows the 95% confidence interval.

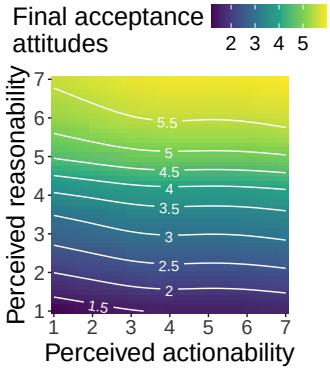


Figure 7: A contour plot of final acceptance attitudes toward AI decisions estimated by the generalized additive mixed model (GAMM), with perceived actionability on the x-axis and perceived reasonability on the y-axis.

to accept an AI decision are more likely to view the subsequent recourse as unreasonable or difficult.

5.2 Impacts of Perceived Reasonability and Actionability on Final Acceptance (RQ2)

As shown in Table 3, we confirmed that both perceptions of reasonability ($F = 714.62$, $p < 0.001$) and actionability ($F = 12.71$, $p < 0.001$) significantly impact final acceptance of AI decisions (R^2=0.816, $P_{\text{lmr}} > 0.80$, $P_{\text{sim}} > 0.80$). Figure 6 presents the smoothing spline functions derived from the generalized additive mixed model (GAMM). Perceived reasonability shows a strong correlation with final acceptance, while perceived actionability shows a positive correlation only in its lower regions.

To intuitively understand the simultaneous effects of non-linear functions of perceived reasonability and actionability on final acceptance, Figure 7 illustrates contour lines of final acceptance values estimated by the GAMM on a plane with perceived reasonability on the vertical axis and perceived actionability on the horizontal axis. The contour lines show that, while perceived actionability has some influence, perceived reasonability is the dominant factor in final acceptance. For example, when perceived actionability is at 7, manipulating perceived reasonability from 1 to 7 shifts final acceptance from values below 2.0 to values above 5.5, representing a change of more than 3.5 units. Conversely, shifting perceived actionability from 1 to 7 results in an approximately 0.5-unit change in final acceptance. This suggests that even if the recourse is perceived as very easy, the perception of recourse reasonability significantly determines the final acceptance attitude.

In summary, our analysis reveals that both perceptions of reasonability and actionability have statistically significant impacts on final acceptance, but perceived reasonability has a much stronger effect compared to perceived actionability.

5.3 Properties of Recourse Affecting Shift in Acceptance Attitudes (RQ3)

Since we found that perceived reasonability correlates more strongly with both initial and final acceptance of AI decisions compared to perceived actionability, we focused on participants' open responses for reasonability evaluations to investigate recourse properties affecting changes in acceptance attitudes. Using thematic analysis [7],

Table 3: Statistical values obtained from the generalized additive mixed model (GAMM) explaining final acceptance attitudes toward AI decisions from perceived reasonability and actionability. This model controls for individual subject-specific effects by introducing random intercepts (R.I.) and random slopes (R.S.) of the smoothing term coefficients corresponding to each explanatory variable.

Objective variable	Explanatory variable	Smth	Coef	S.E.	t value	EDF	F value	p value	AIC
Final acceptance attitudes	Intercept		3.11	0.02	136.80			<0.001	5969.83
	R.I.	✓				0.01	0.00	1.000	
	Perceived reasonability	✓				5.33	714.62	<0.001	
	Perceived actionability	✓				3.40	12.71	<0.001	
	R.S. (perceived reasonability)	✓				250.34	1.37	<0.001	
	R.S. (perceived actionability)	✓				35.82	0.11	0.011	

we identified four key themes: clarity of the cause-and-effect relationship, transparency of the decision criteria, feasibility and desirability of implementation, and concerns about assessment and AI. The relationship between these themes and the properties of recourse is outlined in Table 4. For more detailed information on the distribution and examples of participants' comments regarding these recourse properties, see Tables A1 and A2 in Appendix A.2.

The following sections report the details of the four themes (Section 5.3.1 to 5.3.4) and the results of an exploratory analysis of participants' open-ended responses regarding their reasons for actionability evaluations to understand why the association between perceived actionability and changes in acceptance attitudes was weak (Section 5.3.5).

5.3.1 Theme #1. Clarity of the Cause-and-Effect Relationship. Many participants sought the cause of their rejection through the recourse, regardless of whether participants' acceptance attitudes shifted from negative to positive (NP group) or from positive to negative (PN group). Participants in the NP group recognized the recourse were explaining key evaluation items (NP1, 84/321) or their own shortcomings (NP2, 21/321), such as *"I think that the length of service is important"* (P663) and *"No home ownership, a high school graduate, and not a section manager"* (P696). Conversely, when participants felt the recourse failed to justify the rejection decision by the suggested changes (PN1, 44/307), their acceptance attitudes deteriorated. They expressed frustration with "why" or "not understand" in their articulated concerns or questions, such as *"I don't understand why having a degree means the application will be accepted"* (P436) or *"I cannot understand why I must be in Tokyo"* (P345).

5.3.2 Theme #2. Transparency of the Decision Criteria. Whether the decision criteria are clearly communicated to participants through the recourse impacts participants' acceptance attitudes. Participants were more likely to accept AI decisions when they could infer and agree with the decision criteria and processes likely used by the financial institution based on the recourses (NP3, 74/321). Despite not being informed of the financial institution's decision rules, they frequently mentioned factors related to decision criteria such as income, reliability, or risks. For instance, *"The idea that increasing income by working longer hours will improve repayment ability is convincing"* (P498) and *"Because the reliability has increased"* (P694) were common responses. Acceptance attitudes also improved when participants perceived that the recourse adhered to pre-established rules beyond their control (NP4, 15/321). However, this acceptance seemed to stem from resignation to the likelihood of being approved, e.g., *"If it's said to be a rule, there's no room for disagreement"* (P742).

On the contrary, some participants whose acceptance attitudes worsened noted irrelevance and ambiguities in the decision criteria (PN2, 80/307; PN3, 9/307) when reviewing the changes suggested by the recourse. For example, they remarked, *"Education level and working hours are unrelated to the loan"* (P087) and *"Looking at the specific details, sometimes stability is valued and sometimes it isn't, which made it even less clear why it was rejected"* (P531).

5.3.3 Theme #3. Feasibility and Desirability of Implementation. We also found the recourse properties as action plans. Specifically, when participants viewed the changes suggested by the recourse as feasible within a practical range, their acceptance attitudes improved (NP5, 41/321). Many of them expressed willingness to follow the plan using terms like "effort" and "a little more", e.g., *"[...] it feels like I am close to making it"* (P126), *"It's within reach if I put in the effort"* (P646). Conversely, when participants perceived the suggested changes as either minimal or overly simplistic, their acceptance attitudes deteriorated (PN5, 20/307). Some believed that *"the difference is not significant"* (P077) and that *"[the current profile] should be enough"* (P015), making it hard for them to accept the rejection.

When participants found numerous and unrealistic changes, their reactions fell into two patterns: some accepted the decision out of resignation (NP6, 66/321), while others resisted it (PN4, 68/307). Those who accepted out of resignation felt helpless due to the high hurdle, e.g., *"Because it's not a level that I can achieve"* (P037). Those who resisted expressed dislike for plans because they viewed the plans as *"[...] unrealistic and impossible [...]"* (P212) and thought that *"[...] only a limited number of people can meet the criteria [...]"* (P911).

Participants also evaluated whether the recourse plans were realistically implementable in their personal contexts, such as their life, work, and social environment (PN6, 20/307). For example, P751 noted, *"The current lifestyle is supported by various interconnected factors. It's not possible to easily change just one aspect [...]"*. Even if the recourse content seems simple, implementing it could impact various aspects of their lives. Additionally, constraints like legal limitations or professional obligations could affect the implementation of recourse plans, e.g., *"The proposed daily working hours are*

Table 4: The identified themes and their associated recourse properties (refer to Table A1 for the distribution of participants' open-ended responses on the recourse properties, and Table A2 for participants' perceptions and examples of their responses).

Recourse properties changing acceptance attitudes toward AI decisions	
From negative to positive (NP group)	From positive to negative (PN group)
Theme #1. Clarity of the cause-and-effect relationship	
(1) Explaining rejection reasons through feature changes	(1) Failing to explain rejection reasons through feature changes
(2) Highlighting one's shortcomings through feature changes	
Theme #2. Transparency of the decision criteria	
(3) Explaining decision criteria through feature changes	(2) Failing to explain decision criteria through feature changes
(4) Explaining out-of-control rules through feature changes	(3) Making rejection reasons or decision criteria unclear through feature changes
Theme #3. Feasibility and desirability of implementation	
(5) Proposing realistic plans via feature changes	(4) Proposing unrealistic plans via feature changes (unacceptable)
(6) Proposing unrealistic plans via feature changes (acceptable)	(5) Proposing unnecessary plans via feature changes
	(6) Proposing plans inconsiderate of personal contexts
Theme #4. Concerns about assessment and AI	
	(7) Involving one's undesired items in feature changes
	(8) Triggering AI-averse attitudes

illegal" (P542). Consequently, when the recourse did not consider participants' real-world context, acceptance attitudes deteriorated.

5.3.4 Theme #4. Concerns about Assessment and AI. When the changes suggested by the recourse were unfavorable or disadvantageous to participants, their acceptance attitudes worsened (PN7, 36/307). Specifically, they expressed concerns about the AI decisions when the recourse appeared unfair or discriminatory, e.g., *"Assessments based on residence or job changes feel unfair to me"* (P765), *"I believe this seems like discrimination based on place of origin"* (P958). They articulated concerns about various items, including job position (P468), educational background (P468), residential prefecture (P958, P765), and job change experience (P765).

Additionally, some participants showed aversion to the AI itself rather than the content of the recourse (PN8, 6/307). Considering their initial positive acceptance attitudes, it is likely that they developed a dislike of AI after reviewing recourses rather than an inherent aversion to AI. However, the participants' self-reported responses did not clarify the reasons for this. This issue will be assessed in our future work, as mentioned in Section 6.4.

5.3.5 Exploratory Analysis: Perspectives for Actionability Evaluations. To understand the reason why the association between perceived actionability and changes in acceptance attitudes was weak, we examined the perspectives participants used to evaluate the actionability of recourse. We analyzed participants' open-ended responses regarding their reasons for evaluating actionability using thematic analysis [7] and found six themes: feasibility of change, motivation, rationality, time or financial costs, external constraints, and unpleasantness. For details on these specific frequencies and participants' comments, please refer to Tables A3 and A4 in the Appendix A.3.

Comments indicating the infeasibility of the proposed recourse changes were observed in 48.6% of the NP group and 44.6% of the PN group. Evaluations related to motivation and rationality were generally more positive in the NP group compared to the PN group. The frequency of references to time or financial constraints exhibited little variation between the groups; however, mentions of external constraints were more prevalent in the PN group. While no comments from the NP group referred to unpleasantness, 3.3% of the PN group's comments aligned with this theme.

These results suggest that participants were often taken aback by infeasible recourses and rated actionability low, regardless of whether their acceptance attitudes changed positively or negatively. Consequently, the connection between perceived actionability and final acceptance was weak. We also observed some cases in which recourses that motivated or persuaded participants to act improved their acceptance attitude, while conversely, recourses that demotivated or dissatisfied them worsened their acceptance attitude. This result is consistent with the findings shown in Section 5.3.3.

6 DISCUSSION

6.1 Interpretation of Results

Our quantitative results suggest that decision subjects' initial acceptance attitudes toward AI decisions are pivotal in shaping how they later evaluate the reasonability and actionability of the recourse and whether they finally accept the decision. The finding that a negative initial acceptance leads to a poorer evaluation of the recourse (Section 5.1) is consistent with prior research showing that users who develop a negative impression of AI tend to avoid its advice [17, 58, 75]. This is novel in empirically validating the prior knowledge within the domain of algorithmic recourse. While the result itself may seem intuitive, it indicates that decision subjects with positive and negative initial attitudes respond differently to the same recourse. Without this insight, there is a risk of offering recourses that work for those with positive initial attitudes but fail for those with negative ones. However, if a recourse effectively addresses negative initial attitudes, it will reliably lead to acceptance.

Thus, the findings of RQ1 are essential for accurately identifying recourses that ensure final acceptance.

In addition, the finding that a decline in the perception of recourse leads to a more negative final acceptance (Section 5.2) aligns with the assumption in previous recourse generation studies, which posit a positive correlation between reasonability or actionability and acceptance attitudes [40, 85]. We contributed to this field by providing empirical evidence to support this. More importantly, we found that while both perceptions of reasonability and actionability are statistically linked to final acceptance, the influence of perceived reasonability is predominant, whereas that of perceived actionability is limited. Our results indicate that, in fostering positive acceptance attitudes in decision subjects, ensuring that the recourse is reasonable is more important than making it actionable.

Our qualitative results indicate what properties make a recourse reasonable for decision subjects and how these properties lead to positive acceptance attitudes. We identified several key properties: clearly explaining the decision criteria and the subjects' shortcomings, offering achievable plans that require some effort, and considering fairness perceptions and real-world contexts (Section 5.3). Among these properties, it is noteworthy that the second property—offering achievable plans that require some effort—provides unique insights into the nature of algorithmic recourse as action plans. As described in Section 5.3.3, participants are likely to abandon overly unrealistic and ambitious plans. They also perceive trivial and overly simplistic recourse plans as neither necessary nor meaningful. Conversely, recourse that offers a realistic challenge—neither too easy nor too difficult—motivates participants and enhances their acceptance attitude. This suggests that decision subjects evaluate recourse based on their willingness to follow through with the action plan, which in turn influences whether they accept the decision. While recent studies have focused on actionability [44, 77, 78, 87], Karimi et al. [40] pointed out the potential of recourse as an intervention for improving future outcomes. To make the intervention successful, it is important to consider willingness to act and explore its impacts on acceptance attitudes.

Furthermore, we uncovered recourse properties related to personal perspectives and contexts, including perceptions of fairness (Section 5.3.4) and incompatibility with life, work, and societal norms (Section 5.3.3). This is due to our experimental design. Unlike previous studies where subjects play the role of fictional characters in scenarios when evaluating explanations [23, 51, 52, 87], our participants were genuinely interested in purchasing a car. They submitted their own profile data, received a rejection, and then reviewed the recourse. By simulating the real conditions, we illuminated how recourse affects acceptance attitudes based on personal factors such as perceived fairness and its relevance to daily life. Our study empirically confirms the need for personal context-aware recourse, as noted in previous research [2, 82, 87], and offers insights into the specific personal contexts involved.

Consistent with existing research, our study also shows that both rational recourse (i.e., explaining the rejection reasons or identifying one's shortcomings) and transparent recourse (i.e., clearly explaining the decision criteria) contribute to improved acceptance attitudes (Section 5.3.1 and 5.3.2). Participant feedback indicated that the recourse enhances their acceptance attitudes when it aligns with their prior knowledge and helps them recognize their shortcomings or key evaluation criteria. This result supports previous findings that decision models aligning with human expectations [78] and presented transparently [66] are preferred by users.

This study sheds light on how negative initial acceptance attitudes can be manipulated, but does not answer where their origins are. Initial acceptance attitudes might be shaped by complex interactions with various factors, including personal experiences and expectations. For example, dissatisfaction with AI-driven loan rejections is often linked to heightened sensitivity from repeated similar experiences in the past [25]. Additionally, prior knowledge of AI has been shown to enhance trust and acceptance of such systems [22, 26]. Thus, interventions that adjust individuals' knowledge or experiences could potentially reduce negative initial acceptance attitudes. However, since such interventions usually occur before AI decisions are made, they fall outside the focus of recourse. This study offers insights into post-decision strategies for addressing negative initial acceptance attitudes through recourse.

6.2 Design Implications

We found that decision subjects who initially did not accept AI decisions viewed subsequent recourse negatively, resulting in poor final acceptance of the decisions. Such individuals, adversely affected by AI systems, are indeed the intended beneficiaries of algorithmic recourse. Based on our findings, we discuss here the type of recourse needed and the mechanisms for creating it to help them finally accept the decision.

To secure the acceptance of individuals negatively impacted by AI, recourse must focus on justifying the decision's rationale, ensuring transparency of evaluation criteria, and proposing achievable plans with some effort, all while accounting for fairness perceptions and real-world contexts. However, achieving this is particularly challenging, as these factors are highly dependent on the subjective perceptions of the decision subjects. While it is evident that real-world contexts vary among individuals, it is also obvious that what one considers reasonable, transparent, difficult, or fair is significantly shaped by personal experience, abilities, and knowledge.

For example, understanding of the decision results (i.e., reasonability) and the decision process (i.e., transparency) may be influenced by users' expertise with the AI models or the decision-making domain, as this expertise affects their level of reliance on AI-generated explanations [3, 74]. Moreover, perceptions of the difficulty of proposed actions can be deeply influenced by psychological factors such as self-efficacy [72] and motivation [49], making them highly subject to individual differences. Additionally, previous research reported individual differences (e.g., gender) in fairness perceptions of algorithmic decisions [86]. We also observed that one participant viewed a recourse suggesting only a change in the educational background as *"[...] it's discrimination"* (P929), whereas another viewed it as *"Understand it"* (P365). As such, perceptions and preferences regarding recourse are highly individual. Therefore, the standalone approach employed by current recourse generation studies, where the model computes recourse entirely from input to output without interacting with human factors [40, 83], poses challenges for producing personalized recourse.

Overall, we argue that it is essential to consider these individual differences for recourse design to improve their acceptance

attitudes toward AI decisions. The simplest and most powerful approach to handling such significant individual differences is to engage directly with the decision subjects [89]. Specifically, we recommend explicitly gathering decision subjects' preferences and integrating them into the recourse generation process. A leading example of this approach is the interactive recourse generation interface developed by Wang et al. [87]. This interface enables users to specify which features they wish (or do not wish) to change and experiment with hypothetical adjustments. It allows them to explore various recourse options and identify one that best aligns with their preferences. As a result, we can expect to generate recourses that decision subjects will accept as they satisfy the key attributes we have identified.

Future research should focus on further developing these advanced interfaces to enable users to efficiently reach their preferred plans. "Efficiently" here means reducing the number of plans users need to explore. In general, counterfactual thinking imposes a significant cognitive load and is often associated with negative emotions such as guilt, self-blame, and regret [10]. In our experiment, P431 remarked, *"It's certainly a plan, but I feel like I'm being blamed for not having done enough in the past"*. As such, reviewing recourse options is often stressful. This burden can be exacerbated, especially for decision subjects who have a negative impression of AI systems. Consequently, it is crucial to design systems that streamline the process of generating appropriate recourse and reduce the cognitive burden for decision subjects with initial negative acceptance attitudes.

One possible approach for such system designs is to have decision subjects rate several test recourse plans and use their feedback to determine their preferences for tightening the recourse options. To make this approach successful, we recommend that future research address the following key issues. First, it is necessary to optimize the number of test plans so as to capture user preference accurately while minimizing cognitive load. Second, it is crucial to determine which recourses to include among a limited number of test plans to detect user preferences effectively. Including diverse and consistent recourses may be more beneficial than only similar ones [60]. Finally, it is needed to confirm how preference-aware recourses affect decision subjects' acceptance attitudes toward AI decisions. By implementing these challenges, researchers can develop intelligent systems that generate highly personalized resources without imposing a cognitive burden, helping decision subjects adversely affected by AI systems finally accept AI decisions.

6.3 Limitations

6.3.1 Potential Effects of Recourse Feasibility. Since this study does not consider the feasibility of recourse, it leaves some uncertainty about how participants interpret actionability. As shown in RQ2, perceived actionability was less strongly linked to final acceptance attitudes than perceived reasonability (Section 5.2). This was likely due to the fact that most participants, regardless of whether their acceptance attitude was changed positively or negatively, were confused by the difficulty of implementing the recourse (Section 5.3.5). If feasibility had been measured and recourse with a certain level of feasibility had been presented, it is expected that perceived actionability would have been higher. This would allow for a clearer understanding of how perceptions of actionability differ and how acceptance attitudes vary with perceived actionability. We will address this issue, along with the potential confounding by feasibility (Section 3.2.5), in future research that incorporates the measurement and control of feasibility, leading to more reliable results.

6.3.2 Generalizability to Other Contexts. As this study uses a car loan scenario, it is unclear whether similar results hold in other contexts. We chose this scenario because financial credit screening is a common focus in previous studies on AI-generated explanations [44, 52, 76, 78, 87] and purchasing a car is a familiar, high-stakes, and real-world decision. Our results may not be replicated when the scenario deviates greatly from these contexts. For example, Kuhl et al. [45] observed that, within the self-developed abstract game setting, users tend to accept counterfactual advice that suggests minimal changes from their current states. This is because users do not have prior knowledge or mental models of the setting [45]. As such, if the scenario is unfamiliar, lacks real-world contexts, and is low-stakes, our findings cannot be applied.

Since we selected the recourse items based on loan assessments such as auto [76], home, or travel [92] and credit evaluations such as defaults [4, 29, 90], our findings could be applicable to various financial contexts beyond auto loans, where recourse helps individuals improve their financial standing. These contexts align with recourse's primary application. However, the applicability of our findings to medical or legal decisions is likely limited because more complex considerations might be needed in those areas. For future research attempting to extend our approach to these areas, it would be essential not only to integrate domain-specific factors, such as medication records and treatment data in the medical domain or case logs and criminal history in the legal domain, into the recourse process but also to establish clear decision-making criteria of AI systems based on these factors. Consequently, it will be necessary to design experiments that account for the required domain expertise and specialized knowledge.

The interpretation of our results should consider the participants' demographics. While 53.7% monitors on ASMARQ are female[5], our study had 26.9% female. This is likely due to both the use of the auto loan scenario in this study and the relatively low number of women who have a driver's license [11] and regularly drive in Japan[6] [64]. This gender imbalance in this study could impact our results. For example, men are less sensitive to fairness in algorithmic decisions than women when faced with unfavorable decisions [86], potentially leading to an underestimation of fairness concerns due to the low number of women in the study. Cultural background also could influence the perception of recourse. For example, individuals from collectivist cultures, where harmony and consensus are highly valued [30], might be less likely to prefer recourses that inconvenience their surrounding people, compared to those from individualist cultures, where self-actualization is prioritized [30].

As such, gender and cultural factors could affect our results. However, some findings are consistent with prior research on diverse samples. For example, the link between negative initial acceptance

[5]https://www.asmarq.co.jp/monitor_info/ (only in Japanese)

[6]According to a survey conducted in March 2023 involving 847 men and 945 women aged 18 and older in Tokyo, 466 men (54.4%) and 212 women (22.4%) held a driver's license and drove at least once or twice a week [64]. These statistics reveal that the proportion of women among regular drivers is 31.3% (= 212/(466 + 212)).

and poorer recourse evaluations is in line with a study with participants from 30 countries (67% female), which found that poor first impressions of AI reduce trust in it [75]. Moreover, the result that reasonable recourse fosters final acceptance extends U.S.-based research with a gender-balanced sample, which identified users' preference for logical AI explanations [78]. Overall, the role of gender or culture in our findings remains uncertain and warrants further investigation in the future.

6.3.3 Recourse Features and Their Computation. As noted in Section 3.3.2, this study excluded annual income from recourse items to preserve recourse diversity. Since our primary focus is on perceptions of recourse, this exclusion minimally impacts the study's conclusions. However, as participants were evaluated based on annual income, the recourses excluding annual income do not necessarily provide accurate information. Applying such recourse directly to real-world services has potential risks of misleading the users. Due to experimental considerations aimed at ensuring recourse diversity, annual income was excluded from recourses in this study; however, real-world systems must incorporate all evaluation factors, including annual income, into the recourses.

We calculated recourse distance metrics using the most general method [40, 82], which does not account for the relative difficulties or dependencies between features in individuals' contexts. For example, it does not consider whether improving job position is harder than enhancing English skills or how job status improvement affects years of service. These factors could ultimately depend on individuals' abilities and circumstances, and no established methods currently account for these in recourse generation research. Given these limitations, we adopted the most commonly used approach. This study focuses on the relationship between recourse perception and acceptance of AI decisions, not on delivering individually optimized recourse, limiting the impact of this approach on our conclusions.

6.4 Future Directions

A key direction for future research is to develop efficient methods for generating personalized recourse, as outlined in Section 6.2. This is crucial because once decision subjects lose trust in AI systems after receiving recourses, it is very difficult for the AI systems to restore that trust. High-stakes decisions addressed by algorithmic recourse are uncommon in daily life, unlike routine choices such as movies or restaurants. Consequently, if decision subjects develop AI-aversion attitudes (as seen in Section 5.3.3), it poses a significant problem. According to Tolmeijer et al. [75], regaining lost trust is a slow process that requires numerous interactions between users and AI. Therefore, if trust is lost due to the recourse provided, recourse-generation AI systems are at a significant disadvantage because they have few chances to rectify the situation. To mitigate this risk, it is essential to understand and incorporate decision subjects' preferences in advance, allowing for recourse optimization based on these preferences. We believe that the approach outlined in Section 6.2 is a promising way of achieving this. We will also investigate why participants who initially accepted the decision positively developed AI-averse attitudes after viewing recourse.

Another future direction for this research is to investigate willingness to act. Here, willingness to act refers to an individual's desire or motivation to implement the recourse, while actionability concerns the feasibility or difficulty of executing the recourse. Although these concepts are closely related, they do not always align. As seen in Section 5.3.3, we revealed that if a recourse plan is overly simplistic, participants perceived it as trivial or meaningless and lost their motivation. Conversely, a recourse that requires some effort motivated them to implement. Therefore, we argue that, alongside reasonability and actionability, willingness to act can be a critical factor in enhancing the acceptance attitudes of decision subjects. We will explore which types of recourse can motivate decision subjects, especially those who initially hold a negative view, and develop methods to implement these insights into computational systems.

Addressing the limitations identified in this study is also essential. For example, testing the findings with different experimental tasks beyond loan approval and including participants from diverse cultural backgrounds is necessary to evaluate the generalizability of the results. Given the rapid advancements in Human-AI decision-making research, understanding the applicability of these findings and their limitations is crucial for the future research community.

7 CONCLUSION

Since the core purpose of algorithmic recourse is to assist individuals negatively impacted by AI decisions, unraveling the influence of a negative initial acceptance attitude is crucial. In this study, we conducted a user experiment with 534 Japanese participants and demonstrated that an initially negative acceptance attitude toward AI decisions deteriorated perceptions of the reasonability and actionability of the subsequent recourse, ultimately leading to a decline in final acceptance. However, contrary to this cognitive pattern, participants' acceptance attitudes shifted from negative to positive when the recourse justified the rejection, explained the decision criteria, or proposed realistic and actionable plans with consideration for fairness and personal contexts. Based on these findings, we outlined key principles and challenges for designing computational systems that can efficiently generate optimal recourses without imposing a cognitive burden. This approach is useful for guiding initially dissatisfied decision subjects toward eventual acceptance.

This study elucidates the role of initial acceptance attitudes in the process of providing recourse to individuals negatively impacted by AI and offers valuable insights into recourse generation systems to enhance the acceptance of AI decisions. We hope that our findings contribute to advancing the field of Human-AI decision making and pave the way for developing supportive and satisfying interactions, particularly for affected individuals.

References

[1] Daehwan Ahn, Abdullah Almaatouq, Monisha Gulabani, and Kartik Hosanagar. 2024. Impact of Model Interpretability and Outcome Feedback on Trust in AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–25. https://doi.org/10.1145/3613904.3642780

[2] Solon Barocas, Andrew D. Selbst, and Manish Raghavan. 2020. The hidden assumptions behind counterfactual explanations and principal reasons. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 80–89. https://doi.org/10.1145/3351095.3372830 arXiv:1912.04930

[3] Anna Bashkirova and Dario Krpan. 2024. Confirmation bias in AI-assisted decision-making: AI triage recommendations congruent with expert judgments

increase psychologist trust and recommendation acceptance. *Computers in Human Behavior: Artificial Humans* 2, 1 (jan 2024), 100066. https://doi.org/10.1016/j.chbah.2024.100066

[4] Barry Becker and Ronny Kohavi. 1996. Adult. https://doi.org/10.24432/C5XW20

[5] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. 2019. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence* 1, 1 (jan 2019), 20–23. https://doi.org/10.1038/s42256-018-0004-1

[6] Richard A. Berk, Susan B. Sorenson, and Geoffrey Barnes. 2016. Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. *Journal of Empirical Legal Studies* 13, 1 (mar 2016), 94–115. https://doi.org/10.1111/jels.12098

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300271

[9] Barbara K. Brown and Michael A. Campion. 1994. Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology* 79, 6 (dec 1994), 897–908. https://doi.org/10.1037/0021-9010.79.6.897

[10] Ruth M.J. Byrne. 2016. Counterfactual Thought. *Annual Review of Psychology* 67, 1 (jan 2016), 135–157. https://doi.org/10.1146/annurev-psych-122414-033249

[11] Cabinet Office. 2023. WHITE PAPER ON TRAFFIC SAFETY IN JAPAN. https://www8.cao.go.jp/koutu/taisaku/r05kou_haku/pdf/zenbun/1-1-2-3.pdf Accessed: 2024-Sep-06 (only in Japanese).

[12] Ramon Casanova, Fang Chi Hsu, Kaycee M. Sink, Stephen R. Rapp, Jeff D. Williamson, Susan M. Resnick, and Mark A. Espeland. 2013. Alzheimer's Disease Risk Assessment Using Large-Scale Machine Learning Methods. *PLOS ONE* 8, 11 (nov 2013), e77949. https://doi.org/10.1371/JOURNAL.PONE.0077949

[13] Noah Castelo, Maarten W. Bos, and Donald R. Lehmann. 2019. Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56, 5 (oct 2019), 809–825. https://doi.org/10.1177/0022243719851788

[14] Michael S. Cole, Robert S. Rubin, Hubert S. Feild, and William F. Giles. 2007. Recruiters' Perceptions and Use of Applicant Résumé Information: Screening the Recent Graduate. *Applied Psychology* 56, 2 (apr 2007), 319–343. https://doi.org/10.1111/j.1464-0597.2007.00288.x

[15] Karl de Fine Licht and Jenny de Fine Licht. 2020. Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legtimacy. *AI & SOCIETY* 35, 4 (dec 2020), 917–926. https://doi.org/10.1007/s00146-020-00960-w

[16] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114–126. https://doi.org/10.1037/xge0000033

[17] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2018. Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64, 3 (mar 2018), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

[18] Michael Downs, Jonathan L Chu, Yaniv Yacoby, Finale Doshi-Velez, and Weiwei Pan. 2020. CRUDS: Counterfactual Recourse Using Disentangled Subspaces. In *Proceedings of the 2020 ICML Workshop on Human Interpretablility in Machine Learning (WHI'20)*. 1–2.

[19] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (jun 2003), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

[20] Jessica Maria Echterhoff, Matin Yarmand, and Julian McAuley. 2022. AI-Moderated Decision-Making: Capturing and Balancing Anchoring Bias in Sequential Decision Tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3491102.3517443

[21] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Vol. 19. ACM, New York, NY, USA, 1–19. https://doi.org/10.1145/3411764.3445188

[22] Upol Ehsan, Samir Passi, Q. Vera Liao, Larry Chan, I-Hsiang Lee, Michael Muller, and Mark O Riedl. 2024. The Who in XAI: How AI Background Shapes Perceptions of AI Explanations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–32. https://doi.org/10.1145/3613904.3642474

[23] Meric Altug Gemalmaz and Ming Yin. 2022. Understanding Decision Subjects' Fairness Perceptions and Retention in Repeated Interactions with AI-Based Decision Systems. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 295–306. https://doi.org/10.1145/3514094.3534201

[24] Vittorio Girotto, Donatella Ferrante, Stefania Pighin, and Michel Gonzalez. 2007. Postdecisional Counterfactual Thinking by Actors and Readers. *Psychological Science* 18, 6 (jun 2007), 510–515. https://doi.org/10.1111/j.1467-9280.2007.01931.x

[25] Ana Rita Gonçalves, Amanda Breda Meira, Saleh Shuqair, and Diego Costa Pinto. 2023. Artificial intelligence (AI) in FinTech decisions: the role of congruity and rejection sensitivity. *International Journal of Bank Marketing* 41, 6 (aug 2023), 1282–1307. https://doi.org/10.1108/IJBM-07-2022-0295

[26] Manuel F. Gonzalez, Weiwei Liu, Lei Shirase, David L. Tomczak, Carmen E. Lobbe, Richard Justenhoven, and Nicholas R. Martin. 2022. Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior* 130 (may 2022), 107179. https://doi.org/10.1016/j.chb.2022.107179

[27] Vivek Gupta, Pegah Nokhiz, Chitradeep Dutta Roy, and Suresh Venkatasubramanian. 2019. Equalizing Recourse across Groups. arXiv:1909.03166 http://arxiv.org/abs/1909.03166

[28] Robert R. Hoffman, Matthew Johnson, Jeffrey M. Bradshaw, and Al Underbrink. 2013. Trust in Automation. *IEEE Intelligent Systems* 28, 1 (jan 2013), 84–88. https://doi.org/10.1109/MIS.2013.24

[29] Hans Hofmann. 1994. Statlog (German Credit Data). https://doi.org/10.24432/C5NC77

[30] Geert Jan Hofstede and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind, Third Edition*. https://doi.org/10.1057/jibs.1992.23

[31] Yoyo Tsung-Yu Hou and Malte F. Jung. 2021. Who is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (oct 2021), 1–25. https://doi.org/10.1145/3479864

[32] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 310–323. https://doi.org/10.1145/3531146.3533097

[33] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (sep 2019), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[34] Shalmali Joshi, Oluwasanmi Koyejo, Warut Vijitbenjaronk, Been Kim, and Joydeep Ghosh. 2019. Towards Realistic Individual Recourse and Actionable Explanations in Black-Box Decision Making Systems. , 19 pages. arXiv:1907.09615 http://arxiv.org/abs/1907.09615

[35] Ekaterina Jussupow, Izak Benbasat, and Armin Heinzl. 2020. Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In *Twenty-Eigth European Conference on Information Systems (ECIS2020)*. 1–16. https://aisel.aisnet.org/ecis2020_rp/168

[36] Patricia K. Kahr, Gerrit Rooks, Martijn C. Willemsen, and Chris C. P. Snijders. 2024. Understanding Trust and Reliance Development in AI Advice: Assessing Model Accuracy, Model Explanations, and Experiences from Previous Interactions. *ACM Transactions on Interactive Intelligent Systems* (aug 2024). https://doi.org/10.1145/3686164

[37] Kentaro Kanamori, Takuya Takagi, Ken Kobayashi, and Hiroki Arimura. 2020. DACE: Distribution-Aware Counterfactual Explanation by Mixed-Integer Linear Optimization. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, California, 2855–2862. https://doi.org/10.24963/ijcai.2020/395 arXiv:2002.05522

[38] Kentaro Kanamori, Takuya Takagi, Ken Kobayashi, Yuichi Ike, Kento Uemura, and Hiroki Arimura. 2021. Ordered Counterfactual Explanation by Mixed-Integer Linear Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11564–11574. https://doi.org/10.1609/aaai.v35i13.17376

[39] Amir-Hossein Karimi, Gilles Barthe, Borja Balle, and Isabel Valera. 2020. Model-Agnostic Counterfactual Explanations for Consequential Decisions. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*. 895–905. http://arxiv.org/abs/1905.11190

[40] Amir Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. 2022. A Survey of Algorithmic Recourse: Contrastive Explanations and Consequential Recommendations. *Comput. Surveys* 55, 5 (2022). https://doi.org/10.1145/3527848

[41] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic Recourse: from Counterfactual Explanations to Interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 353–362. https://doi.org/10.1145/3442188.3445899 arXiv:2002.06278

[42] Amir Hossein Karimi, Julius von Kügelgen, Bernhard Schölkopf, and Isabel Valera. 2020. Algorithmic recourse under imperfect causal knowledge: A probabilistic approach. In *Advances in Neural Information Processing Systems*. arXiv:2006.06831

[43] Mark T Keane, Eoin M Kenny, Eoin Delaney, and Barry Smyth. 2021. If Only We Had Better Counterfactual Explanations: Five Key Deficits to Rectify in the Evaluation of Counterfactual XAI Techniques. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, California, 4466–4474. https://doi.org/10.24963/ijcai.2021/609 arXiv:2103.01035

[44] Lara Kirfel and Alice Liefgreen. 2021. What If (and How ...)? - Actionability Shapes People's Perceptions of Counterfactual Explanations in Automated Decision-Making. In *ICML-21 Workshop on Algorithmic Recourse*.

[45] Ulrike Kuhl, André Artelt, and Barbara Hammer. 2022. Keep Your Friends Close and Your Counterfactuals Closer: Improved Learning From Closest Rather Than Plausible Counterfactual Explanations in an Abstract Setting. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, 2125–2137. https://doi.org/10.1145/3531146.3534630 arXiv:2205.05515

[46] Nathan R. Kuncel, David M. Klieger, Brian S. Connelly, and Deniz S. Ones. 2013. Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology* 98, 6 (2013), 1060–1072. https://doi.org/10.1037/a0034156

[47] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (jan 2018), 205395171875668. https://doi.org/10.1177/2053951718756684

[48] Martin Leo, Suneel Sharma, and K. Maddulety. 2019. Machine Learning in Banking Risk Management: A Literature Review. *Risks 2019, Vol. 7, Page 29* 7, 1 (mar 2019), 29. https://doi.org/10.3390/RISKS7010029

[49] Edwin A Locke and Gary P Latham. 2006. New Directions in Goal-Setting Theory. *Current Directions in Psychological Science* 15, 5 (oct 2006), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

[50] Steven Lockey, Nicole Gillespie, Daniel Holm, and Ida Asadi Someh. 2021. A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. 5463–5472. https://hdl.handle.net/10125/71284

[51] Henrietta Lyons, Tim Miller, and Eduardo Velloso. 2023. Algorithmic Decisions, Desire for Control, and the Preference for Human Review over Algorithmic Review. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 764–774. https://doi.org/10.1145/3593013.3594041

[52] Henrietta Lyons, Senuri Wijenayake, Tim Miller, and Eduardo Velloso. 2022. What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517606

[53] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–20. https://doi.org/10.1145/3613904.3642671

[54] Divyat Mahajan, Chenhao Tan, and Amit Sharma. 2019. Preserving Causal Constraints in Counterfactual Explanations for Machine Learning Classifiers. arXiv:1912.03277 https://arxiv.org/abs/1912.03277v3http://arxiv.org/abs/1912.03277

[55] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2022. A Survey on Bias and Fairness in Machine Learning. *Comput. Surveys* 54, 6 (jul 2022), 1–35. https://doi.org/10.1145/3457607

[56] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007 arXiv:1706.07269

[57] Tim Miller. 2021. Contrastive explanation: A structural-model approach. *The Knowledge Engineering Review* 36 (oct 2021), e14. https://doi.org/10.1017/S0269888921000102

[58] Lillio Mok, Sasha Nanda, and Ashton Anderson. 2023. People Perceive Algorithmic Assessments as Less Fair and Trustworthy Than Identical Human Assessments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (sep 2023), 1–26. https://doi.org/10.1145/3610100

[59] Vincenzo Moscato and Giancarlo Sperlì. 2022. A benchmark of credit score prediction using Machine Learning. In *CEUR Workshop Proceedings*. http://ceur-ws.org

[60] Ramaravind K. Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 607–617. https://doi.org/10.1145/3351095.3372850 arXiv:1905.07697

[61] Mahsan Nourani, Donald R Honeycutt, Jeremy E Block, Chiradeep Roy, Tahrima Rahman, Eric D Ragan, and Vibhav Gogate. 2020. Investigating the Importance of First Impressions and Explainable AI with Interactive Video Analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382967

[62] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *Proceedings of 26th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 340–350. https://doi.org/10.1145/3397481.3450639

[63] Claudio Novelli, Mariarosaria Taddeo, and Luciano Floridi. 2024. Accountability in artificial intelligence: what it is and how it works. *AI & SOCIETY* 39, 4 (aug 2024), 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

[64] Office of the Governor for Policy Planning, Tokyo Metropolitan Government. 2023. Public Opinion Survey on Automobile Usage and the Environment. https://www.metro.tokyo.lg.jp/tosei/hodohappyo/press/2023/03/29/documents/01_full.pdf Accessed: 2024-Sep-06 (only in Japanese).

[65] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3491102.3502104

[66] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and Measuring Model Interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–52. https://doi.org/10.1145/3411764.3445315

[67] Rafael Poyiadzi, Kacper Sokol, Raul Santos-Rodriguez, Tijl De Bie, and Peter Flach. 2020. FACE: Feasible and Actionable Counterfactual Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 344–350. https://doi.org/10.1145/3375627.3375850 arXiv:1909.09369

[68] Amy Rechkemmer and Ming Yin. 2022. When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models. In *Proceeding of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3501967

[69] Alexis Ross, Himabindu Lakkaraju, and Osbert Bastani. 2021. Learning Models for Actionable Recourse. In *Advances in Neural Information Processing Systems*. 18734–18746.

[70] Chris Russell. 2019. Efficient Search for Diverse Coherent Explanations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 20–28. https://doi.org/10.1145/3287560.3287569 arXiv:1901.04909

[71] Candice Schumann, Jeffrey S Foster, Nicholas Mattei, and John P Dickerson. 2020. We Need Fairness and Explainability in Algorithmic Hiring. In *Proceeings of the 19th International Conference on Autonomous Agents and Multiagent Systems*. 1716–1720. www.ifaamas.org

[72] Dale H. Schunk. 1990. Goal Setting and Self-Efficacy During Self-Regulated Learning. *Educational Psychologist* 25, 1 (1990), 71–86. https://doi.org/10.1207/S15326985EP2501_6

[73] Harini Suresh, Natalie Lao, and Ilaria Liccardi. 2020. Misplaced Trust: Measuring the Interference of Machine Learning in Human Decision-Making. In *Proceedings of the 12th ACM Conference on Web Science*. ACM, New York, NY, USA, 315–324. https://doi.org/10.1145/3394231.3397922 arXiv:2005.10960

[74] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, textual or hybrid: the effect of user expertise on different explanations. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 109–119. https://doi.org/10.1145/3397481.3450662

[75] Suzanne Tolmeijer, Ujwal Gadiraju, Ramya Ghantasala, Akshit Gupta, and Abraham Bernstein. 2021. Second chance for a first impression? Trust development in intelligent system interaction. *UMAP 2021 - Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (jun 2021), 77–87. https://doi.org/10.1145/3450613.3456817/SUPPL_FILE/UMAP21-LP4215.MP4

[76] Tomu Tominaga, Naomi Yamashita, and Takeshi Kurashima. 2024. Reassessing Evaluation Functions in Algorithmic Recourse: An Empirical Study from a Human-Centered Perspective. In *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, California, 7913–7921. https://doi.org/10.24963/ijcai.2024/876

[77] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable Recourse in Linear Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 10–19. https://doi.org/10.1145/3287560.3287566

[78] Peter M. VanNostrand, Dennis M Hofmann, Lei Ma, and Elke A Rundensteiner. 2024. Actionable Recourse for Automated Decisions: Examining the Effects of Counterfactual Explanation Type and Presentation on Lay User Understanding. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 1682–1700. https://doi.org/10.1145/3630106.3658997

[79] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (apr 2023), 1–38. https://doi.org/10.1145/3579605

[80] Suresh Venkatasubramanian and Mark Alfano. 2020. The philosophical basis of algorithmic recourse. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 284–293. https://doi.org/10.1145/3351095.3372876

[81] Oleksandra Vereschak, Fatemeh Alizadeh, Gilles Bailly, and Baptiste Caramiaux. 2024. Trust in AI-assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made. In *Proceedings of the CHI*

Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3613904.3642018

[82] Sahil Verma, Varich Boonsanong, Minh Hoang, Keegan E. Hines, John P. Dickerson, and Chirag Shah. 2020. Counterfactual Explanations and Algorithmic Recourses for Machine Learning: A Review. (oct 2020). arXiv:2010.10596 http://arxiv.org/abs/2010.10596

[83] Sahil Verma, Keegan Hines, and John P. Dickerson. 2022. Amortized Generation of Sequential Algorithmic Recourses for Black-Box Models. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 8 (jun 2022), 8512–8519. https://doi.org/10.1609/aaai.v36i8.20828 arXiv:2106.03962

[84] Julius von Kügelgen, Amir-Hossein Karimi, Umang Bhatt, Isabel Valera, Adrian Weller, and Bernhard Schölkopf. 2022. On the Fairness of Causal Algorithmic Recourse. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 9 (jun 2022), 9584–9594. https://doi.org/10.1609/aaai.v36i9.21192 arXiv:2010.06529

[85] Sandra Wachter, Brent Mittelstadt, and Chris Russel. 2018. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and thhe GDPR. *Harvard Journal of Law & Technology* 20, 3 (mar 2018), 842–887.

[86] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376813 arXiv:2001.09604

[87] Zijie J. Wang, Jennifer Wortman Vaughan, Rich Caruana, and Duen Horng Chau. 2023. GAM Coach: Towards Interactive and User-centered Algorithmic Recourse. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3544548.3580816 arXiv:2302.14165

[88] Greta Warren, Ruth M J Byrne, and Mark T Keane. 2023. Categorical and Continuous Features in Counterfactual Explanations of AI Systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, Vol. 17. ACM, New York, NY, USA, 171–187. https://doi.org/10.1145/3581641.3584090

[89] Daniel S. Weld and Gagan Bansal. 2019. The challenge of crafting intelligible intelligence. *Commun. ACM* 62, 6 (may 2019), 70–79. https://doi.org/10.1145/3282486 arXiv:1803.04263

[90] I-Cheng Yeh. 2016. Dafeault of Credit Card Clients. https://doi.org/10.24432/C55S3H

[91] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300509

[92] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–21. https://doi.org/10.1145/3544548.3581161

[93] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 295–305. https://doi.org/10.1145/3351095.3372852 arXiv:2001.02114

A APPENDICES

A.1 ANALYTICAL MODELS

A.1.1 GAM Model (RQ1). The model we used for RQ1 is expressed as follows:

$$y_i = \beta_0 + f(x_i) = \beta_0 + \sum_{j=1}^{N} \beta_j s_j(x_i)$$

Here, y_i represents the i-th observation of the response variable, x_i denotes the explanatory variable, s refers to the basis functions, and β is the parameter. The nonlinear regression model $f()$ obtained by fitting this model to the experimental data is known as a smoothing spline function (or smoothing spline curve), while $s()$ denotes the smoothing terms. The index j identifies the smoothing terms, and N denotes their total number. The statistical significance of the explanatory variable's effect on the objective variable is tested using an F-test to verify whether $s()$, with x as input, is equal to zero.

A.1.2 GAMM Model (RQ2). For RQ2, we use a model with random intercepts and slopes, which includes mixed effects on the intercept (β_0) and coefficients of smoothing terms ($\beta_{j\geq 1}$).

Specifically, with u as the user identifier, i as the observation identifier, V as the set of explanatory variables, k as the explanatory variable identifier, and j as the smoothing term identifier, the relationship between the objective variable y and explanatory variables x is modeled using smoothing terms s, the total number of smoothing terms N, smoothing term coefficients β_j, and intercept β_0 as follows:

$$\begin{aligned} y_{iu} &= \beta_{0u} + \sum_{k=1}^{|V|} f_k(x_{k,iu}) \\ &= \beta_{0u} + \sum_{k=1}^{|V|} \sum_{j=1}^{N_k} \beta_{k,ju} s_{k,j}(x_{k,iu}) \\ \beta_{0u} &= \beta_{00} + \gamma_{0u}, \quad \beta_{k,ju} = \beta_{k,j0} + \gamma_{k,ju} \\ \gamma_{0u} &\sim \mathcal{N}(0, \sigma_0^2), \quad \gamma_{k,ju} \sim \mathcal{N}(0, \sigma_j^2) \end{aligned}$$

By hierarchically structuring the parameters of the intercept and smoothing term coefficients with mixed effects γ, we can isolate individual-specific effects and assess the impact of explanatory variables on the objective variable. The mixed effects γ follow the distributions with mean 0 and variance σ.

A.2 RECOURSE PROPERTIES

Table A1: Recourse properties and their distributions. The NP or PN in the column ID indicates that acceptance attitudes toward AI decisions change from negative to positive (NP group) or from positive to negative (PN group), respectively.

Theme	ID	Recourse Property	N	%
Clarity	NP1	Explaining rejection reasons through feature changes	84	26.2
Clarity	NP2	Highlighting one's shortcomings through feature changes	21	6.5
Transparency	NP3	Explaining decision criteria through feature changes	74	23.1
Transparency	NP4	Explaining out-of-control rules through feature changes	15	4.7
Feasibility	NP5	Proposing realistic plans via feature changes	41	12.8
Feasibility	NP6	Proposing unrealistic plans via feature changes (acceptable)	66	20.6
		(Others)	20	6.2
			321	100.0
Clarity	PN1	Failing to explain rejection reasons through feature changes	44	14.3
Transparency	PN2	Failing to explain decision criteria through feature changes	80	26.1
Transparency	PN3	Making rejection reasons or decision criteria unclear through feature changes	9	2.9
Feasibility	PN4	Proposing unrealistic plans via feature changes (unacceptable)	68	22.1
Feasibility	PN5	Proposing unnecessary plans via feature changes	30	9.8
Feasibility	PN6	Proposing plans inconsiderate of personal contexts	20	6.5
Concerns	PN7	Involving one's undesired items in feature changes	36	11.7
Concerns	PN8	Triggering AI-averse attitudes	6	2.0
		(Others)	14	4.6
			307	100.0

Table A2: Descriptions of recourse properties from decision subjects' perspective with examples of their responses. P* shows participant IDs and → indicates the change in evaluation scores from the initial acceptance attitude to the final acceptance attitude.**

ID	Decision Subjects' Perspective	Example quotes
NP1	The decision subjects can justify the rejection decision based on changes in the recourse items.	*"It's convincing"* (P951, 3→7), *"I understand that a loan approval will not be granted unless the position is at least a manager"* (P025, 2→7), *"I think that the length of service is important"* (P663, 3→5), *"I understood the reason for the rejection and could infer that a similar process is followed in bank evaluations"* (P538, 3→5)
NP2	The decision subjects become aware of their shortcomings through changes in the recourse items.	*"Insufficient years of service"* (P363, 1→6), *"No home ownership, a high school graduate, and not a section manager"* (P696, 3→6), *"A doctoral degree with short working hours seems like non-regular employment"* (P762, 3→5)
NP3	The decision subjects believe the changes in the recourse items are related to the decision criteria.	*"The idea that increasing income by working longer hours will improve repayment ability is convincing"* (P498, 2→5), *"Owning a home indicates having fixed assets, which suggests a more stable situation than the current one"* (P159, 2→6), *"Because the reliability has increased"* (P694, 2→5), *"These are the conditions for someone who can be reliably trusted"* (P514, 2→6)
NP4	The decision subjects recognize that the decision criteria are beyond their control based on changes in the recourse items.	*"If it's said to be a rule, there's no room for disagreement"* (P742, 2→5), *"When the result is given, it can't be helped"* (P563, 2→6)
NP5	The decision subjects are recommended plans that are realistically actionable.	*"Even though I didn't pass the evaluation, it feels like I am close to making it"* (P126, 3→5), *"It's within reach if I put in the effort"* (P646, 2→7), *"I think it's something I can fix with enough effort"* (P836, 3→5)
NP6	The decision subjects are recommended plans that are not realistically actionable.	*"Because it's not a level that I can achieve"* (P037, 3→6), *"With such a wide range of conditions, it is understandable"* (P089, 3→6), *"The profile presents an excessively high hurdle"* (P012, 1→5)
	(Others)	*"None in particular"*, *"I'm not sure"*, *"Because it can be compared"*, etc.
PN1	The decision subjects cannot justify the rejection decision based on changes in the recourse items.	*"Unacceptable"* (P236, 6→3), *"I can't understand why I must be in Tokyo"* (P345, 5→2), *"I don't understand why having a degree means the application will be accepted"* (P436, 5→3), *"While working longer hours might slightly increase income, I wonder why it needs to be home office hours. It would make more sense to work overtime at the office"* (P668, 5→3)
PN2	The decision subjects believe the changes in the recourse items are not related to the decision criteria.	*"Education level and working hours are unrelated to the loan"* (P087, 6→2), *"I don't believe working hours impact the loan assessment"* (P358, 5→2), *"These changes are not considered relevant to the disapproval criteria"* (P518, 5→3), *"The necessity of overseas experience is not understandable"* (P407, 5→2)
PN3	The decision subjects' understanding of the rejection reason or decision criteria becomes unclear due to changes in the recourse items.	*"Looking at the specific details, sometimes stability is valued and sometimes it isn't, which made it even less clear why it was rejected"* (P531, 5→3), *"The criteria for job titles are vague"* (P077, 5→3), *"The criteria are not clear"* (P141, 7→2)
PN4	The decision subjects are recommended plans that are not realistically actionable.	*"I feel that the content is not realistic"* (P381, 5→1), *"Changing job positions or educational levels, especially while pursuing a PhD, seems unrealistic and impossible"* (P212, 7→1), *"I believe that only a limited number of people can meet the criteria for job position and management experience"* (P911, 5→2), *"The required levels for approval are too high for each process"* (P672, 5→1)
PN5	The decision subjects are recommended plans with no significant difference from counterfactual samples.	*"The difference is not significant"* (P077, 5→3), *"I cannot accept that there is a difference in the evaluation results based on such a small discrepancy"* (P542, 5→2), *"A university degree should be enough"* (P015, 5→2), *"I don't see that the current profile indicates a lack of payment ability"* (P325, 6→1)
PN6	The decision subjects are recommended plans that do not align with their real-world context, such as lifestyle, work, or social environment.	*"The current lifestyle is supported by various interconnected factors. It's not possible to easily change just one aspect, such as the workplace"* (P751, 6→1), *"I do not work from home"* (P193, 6→1), *"The proposed daily working hours are illegal"* (P542, 5→3), *"We are no longer in an era where working for the same company for many years is common"*
PN7	The decision subjects express concerns about fairness or discrimination based on changes in the recourse items.	*"I don't want to be evaluated based on the items listed above"* (P231, 6→2), *"It's not feasible to judge solely by job position and education levels"* (P468, 5→3), *"I believe this seems like discrimination based on place of origin"* (P958, 5→2), *"Assessments based on residence or job changes feel unfair to me"* (P765, 5→2)
PN8	The decision subjects develop an aversion to AI.	*"I cannot accept being told by AI"* (P082, 5→1), *"Everyone has a different life, and it should not be decided by AI"* (P407, 5→2), *"I think it's unreasonable to be judged by AI"* (P544, 5→1)
	(Others)	*"None"*, *"Don't know"*, *"Lost interest"*, *"Flexible"*

A.3 PERSPECTIVES FOR ACTIONABILITY EVALUATIONS

Table A3: Perspectives for evaluating actionability of recourses and their themes

Theme	Perspective for actionability evaluation	NP group		PN group	
Feasibility of change	Proposing plans that are easy to implement the feature changes	17	5.3%	16	5.2%
	Proposing plans that are difficult to implement the feature changes	156	48.6%	137	44.6%
Motivation	Proposing motivating plans via feature changes	37	11.5%	4	1.3%
	Proposing unmotivating plans via feature changes	6	1.9%	16	5.2%
Rationality	Proposing rational plans via feature changes	25	7.8%	6	2.0%
	Proposing irrational plans via feature changes	10	3.1%	32	10.4%
Time or financial costs	Proposing plans with time or financial costs via feature changes	35	10.9%	30	9.8%
External constraints	Proposing plans beyond one's effort or discretion via feature changes	22	6.9%	42	13.7%
Unpleasantness	Involving one's undesired items in feature changes	0	0.0%	10	3.3%
	(Others)	13	4.0%	14	4.6%
		321	100.0%	307	100.0%

Table A4: Example quotes corresponding to perspectives for evaluating actionability of recourses. P* shows participant IDs and → indicates the change in evaluation scores from the initial acceptance attitude to the final acceptance attitude.**

Theme	Perspective for actionability evaluation	Example quotes
Feasibility of change	Proposing plans that are easy to implement the feature changes	*"Relatively easy to implement."* (P469, 1→5), *"Just changing how I work from home."* (P417, 5→2)
	Proposing plans that are difficult to implement the feature changes	*"Learning English is tough."* (P183, 2→6), *"Way too hard."* (P648, 7→1)
Motivation	Proposing motivating plans via feature changes	*"Doesn't seem that different, so if I really tried, I think I could do it."* (P656, 3→6), *"Feels like it could be doable."* (P350, 5→1)
	Proposing unmotivating plans via feature changes	*"Improving academic skills after becoming an adult seems tough. I wouldn't go out of my way to get a degree for this."* (P159, 2→6), *"I don't want to work more hours."* (P809, 6→1)
Rationality	Proposing rational plans via feature changes	*"Seems like a reasonable suggestion."* (P580, 3→5), *"Whether or not one owns a home seems to be the basis for assessment."* (P503, 5→2)
	Proposing irrational plans via feature changes	*"Not really sure."* (P284, 1→5), *"Aligning job type and lifestyle with the AI's judgment criteria doesn't necessarily lead to higher or more stable income."* (P799, 5→1)
Time or financial costs	Proposing plans with time or financial costs via feature changes	*"It takes years to get a higher education or a better job, and owning a home needs a lot of money."* (P706, 3→7), *"Takes too long and it's just not practical."* (P434, 5→1)
External constraints	Proposing plans beyond one's effort or discretion via feature changes	*"My company forbids side jobs."* (P458, 3→5), *"Can't change it just by my own effort."* (P523, 5→1)
Unpleasantness	Involving one's undesired items in feature changes	*"Feels like discrimination."* (P236, 6→3), *"Concerns about being influenced by educational background and job position."* (P087, 6→2)
	(Others)	*"Just a feeling."*, *"Nothing special."*, *"No particular reason."*