

Trustworthy FinAInce: Unpacking How AI-Mediated Financial Advice is Judged

ARYAN RAMCHANDRA KAPADIA, University of Illinois Urbana-Champaign, USA

ESHWAR CHANDRASEKHARAN, University of Illinois Urbana-Champaign, USA

KOUSTUV SAHA, University of Illinois Urbana-Champaign, USA

As generative AI is increasingly used as a source of personal financial guidance, understanding how people appraise such advice is important for supporting appropriate reliance. We conducted a randomized vignette experiment with 285 U.S. adults across eight financial decisions, independently varying three advice styles—*AI*, *expert*, and *online community*—and displayed source labels while holding the underlying recommendation consistent. Advice style most strongly shaped message and safety appraisals. Expert labels selectively increased perceived source knowledge, and decision context primarily shaped risk and safety appraisals. These appraisals were associated with downstream judgments, with models explaining 69.2% of overall quality, 75.9% of trust, and 82.9% of intended reliance. Expert-style advice also remained most preferred when shown without source labels. Our findings have implications for understanding financial advice evaluation, distinguishing the roles of advice style and source labels, and designing financial AI that supports grounded evaluation rather than simply maximizing trust.

CCS Concepts: • **Human-centered computing** → *Empirical studies in collaborative and social computing; Social media*; • **Applied computing** → *Psychology*.

Additional Key Words and Phrases: Generative AI, trust, reliance, advice evaluation, advice style, human-AI decision making

1 Introduction

Generative AI is rapidly becoming a source of everyday financial guidance, from budgeting and saving to investing [16, 58]. General-purpose conversational AI systems such as *ChatGPT*, *Gemini*, and *Claude*, as well as finance-oriented tools such as *SoFi Coach* [12], make financial guidance available on demand, often at little to no direct cost, while conversational interfaces enable context disclosure and iterative refinement [58, 82]. This accessibility is already translating into substantial use: in a recent survey of 4,000 U.S. consumers, 51% reported using AI for financial advice or information [1]. Yet easier access does not guarantee reliability. LLMs can still provide unreliable or insufficiently contextualized financial guidance, alongside broader concerns around privacy, transparency, and accountability [29, 31, 36, 56]. Moreover, their performance can deteriorate as financial decisions become more complex and context-dependent [16, 29]. This is particularly consequential in personal finance, where plausible but imperfect recommendations can cause financial harm, compounded by overreliance [11, 40], growing AI dependency [24], and cognitive offloading of financial judgment [67, 87]. Understanding how users evaluate such advice—and how those appraisals relate to trust and intended reliance—is therefore important for designing financial AI that supports appropriately grounded decision-making.

However, the process of advice evaluation is much more complex than assessing substantive content alone. [9] Prior work shows that message-level characteristics such as tone, reasoning, and presentation can shape responses to advice [15, 53, 74, 76], while explicit *source labels* can activate expectations about expertise, credibility, and authority [23, 70, 73]. Advice evaluation may further vary with the *decision context* [9, 33] and with the knowledge, experiences, and predispositions of the person evaluating it [4, 28]. Yet these influences have largely been examined separately or only in partial combinations, leaving limited understanding of how they jointly shape advice evaluation and where differences in *trust* and intended *reliance* originate.

Authors' Contact Information: Aryan Ramchandra Kapadia, kapadia8@illinois.edu, University of Illinois Urbana-Champaign, Urbana, IL, USA; Eshwar Chandrasekharan, eshwar@illinois.edu, University of Illinois Urbana-Champaign, Urbana, IL, USA; Koustuv Saha, ksaha2@illinois.edu, University of Illinois Urbana-Champaign, Urbana, IL, USA.

Understanding how these influences shape advice evaluation becomes especially important in a multi-source financial advice ecosystem, where people may encounter guidance from AI systems, financial professionals, and online communities that differ in perceived expertise, credibility, and accountability. Yet financial-AI research has largely compared AI with professional advisors or examined AI advice in isolation, with comparatively little work examining online-community advice within the same evaluative framework [42, 76, 78, 83, 85]. These source perceptions may arise from both explicit and implicit cues: *source labels* directly indicate who or what produced the advice, whereas *advice style* conveys source-associated signals through tone, reasoning, and presentation. Distinguishing these cues matters because they have different implications for financial AI design and governance: advice style shapes how authority is communicated, whereas source labels make provenance explicit. Lastly, the decision context may influence how people weigh different cues when evaluating advice, with implications for subsequent *trust* and *reliance* [9, 33]. Motivated by these gaps, we ask the following research questions (RQs):

RQ1: What factors shape people’s appraisals of financial advice?

RQ2: How are these appraisals associated with *Overall Quality*, *Trust*, and *Intended Reliance*?¹

To answer these RQs, we conducted a preregistered randomized vignette experiment with 285 U.S. adults, yielding 1,140 repeated evaluations of financial advice. We constructed eight personal-finance scenarios spanning a $2 \times 2 \times 2$ decision-context taxonomy of *stakes*, *external uncertainty*, and *verifiability*. For each scenario, we constructed three source-associated *advice styles*—AI Financial Assistant, Certified Financial Planner, and Online Community—while keeping the underlying financial recommendation consistent across versions. We independently manipulated *source labels* by presenting advice with its corresponding label, without a label, or with an incorrect label. Guided by the Information Adoption Model (IAM) as an organizing lens for information evaluation and adoption [75], we examined advice style, source labels, decision context, and individual differences as potential influences on four appraisals—*Message Appraisal*, *Risk Acknowledgment*, *Safety Concern*, and *Source Knowledge*—and how these appraisals were associated with *Overall Quality*, *Trust*, and *Intended Reliance*. We additionally included a comparative task in which participants ranked all three advice styles for a previously unseen financial scenario.

Our results show that advice style, source labels, and decision context were associated with different aspects of financial advice evaluation. *Advice style* showed broader and more consistent associations across appraisals, particularly with *Message Appraisal* and *Safety Concern*, whereas *source-label* effects were more selective and most clearly reflected in perceived *Source Knowledge*. *Decision context* was associated primarily with risk- and safety-related appraisals. These appraisals were much more strongly associated with downstream outcomes than upstream factors (advice style, source labels, and decision context) alone: models incorporating the evaluation pathway explained 69.2% of the variance in *Overall Quality*, 75.9% in *Trust*, and 82.9% in intended *Reliance*, compared with 9.8–14.8% for upstream factors alone. *Overall Quality* was the strongest correlate of *Trust* ($\beta = .607$), while *Trust* was the strongest correlate of *Intended Reliance* ($\beta = .594$). We also find that the effects of Advice style and source labels were largely additive rather than interactive, while individual differences produced more selective associations with how appraisal cues related to *Trust* and *Intended Reliance*.

Overall, our work makes empirical, theoretical, and design contributions. **Empirically**, we compare AI, Expert, and Online Community financial advice within a common experimental framework, disentangling *advice style* from *source labels* while examining how decision context and individual differences relate to advice evaluation. **Theoretically**, we

¹ *Overall Quality* refers to participants’ overall evaluation of the advice; *Trust* refers to whether they would trust the advice if they were in the protagonist’s situation; and *Intended Reliance* refers to whether they would feel comfortable following the advice.

use information adoption as an organizing lens to show that *Trust* and intended *Reliance* are embedded within a broader evaluation process: different factors are associated with distinct appraisals, which are in turn strongly associated with *Overall Quality*, *Trust*, and intended *Reliance*. From a **design** perspective, we argue for an evaluation-first approach to financial AI: systems should help users assess the reasoning, risks, and provenance underlying advice, rather than simply making recommendations appear more trustworthy. Together, these contributions shift the design goal from maximizing *Trust* toward supporting grounded evaluation and appropriate *Reliance*.

2 Related Work

2.1 Financial Advice in the Age of AI

Recent advances in large language models have expanded their potential role in personal financial decision support. Newer systems demonstrate growing capabilities across financial analysis, investment-related tasks, portfolio support, and personal financial guidance [19, 20, 29, 39, 55, 71]. Their natural-language interfaces also make financial guidance highly accessible, allowing users to describe their circumstances, ask follow-up questions, and progressively refine recommendations [58, 82]. However, technical capability alone does not guarantee reliable personalized advice. LLMs can produce inaccurate information, operate with incomplete user context, and raise privacy, transparency, and compliance concerns [31, 36, 56]; their performance can also deteriorate as financial decisions become more complex and context-dependent [29]. Similarly, Choukhmane et al. [16] find that LLM recommendations broadly align with established financial principles while remaining less reliable for subtler, context-sensitive aspects of household financial planning. These limitations have motivated growing human-centered research on how people perceive and use AI-generated financial advice, including work on advisor preferences, trust, personalization, transparency, and interaction design [10, 52, 76, 85]. We build on this shift beyond technical performance toward understanding how people perceive, evaluate, and ultimately use AI-generated financial advice.

Financial advice, however, is rarely encountered in isolation from other sources. Recent nationally representative evidence shows that among the U.S. adults who sought financial guidance in the preceding year, 32% consulted professional advisors; informal and digitally mediated sources included family members (35%), friends (23%), internet research (73%), and news, media, or social media (26%); and 18% used AI tools such as ChatGPT or Claude [32]. At the same time, source use did not directly correspond to confidence, which was highest for professional advisors and substantially lower for AI systems [3, 32]. Financial decision-making is also socially embedded: peer effects can shape investment behavior [57], online communities such as Reddit influence how retail investors interpret financial information and develop strategies [54], and Trust in financial guidance can vary with both the advice source and users' surrounding social networks [10]. Yet research on AI-mediated financial advice has largely examined AI relative to professional advisors [10, 52], AI-human hybrid advisors [42, 83], or AI systems in isolation through factors such as personalization, preference-elicitation, agent personality, and human-like interaction [76, 85]. Online communities have received comparatively separate attention; for example, Thukral et al. [78] compare LLM-generated and human responses to financial questions on Reddit. What remains less understood is how people evaluate AI, professional, and online-community advice within the same financial decision setting. We build on this literature by examining these three prominent advice sources within a common evaluation framework.

2.2 Evaluating Financial Advice: Appraisals, Trust, and Reliance

The Information Adoption Model (IAM) provides a useful organizing lens for understanding how people evaluate and ultimately use external information. At its core, IAM proposes that people engage in a sequence of cognitive assessments when encountering information, including evaluations of *argument quality* and *source credibility*, which shape perceived *information usefulness* and subsequent *information adoption* [75]. This perspective has been widely applied in digital information environments [18, 79] and, more recently, to AI-mediated advice [13, 25, 33, 46]. In financial AI specifically, recent work has further incorporated constructs such as perceived trust and risk into this broader evaluation and adoption process [33]. In this study, we build on this perspective by examining these cognitive assessments at greater granularity in the context of financial advice. In addition to message- and source-related appraisals, we make *Risk Acknowledgment* and *Safety Concern* explicit because financial advice can appear clear and credible while still failing to acknowledge uncertainty or exposing users to potential harm.

These cognitive assessments ultimately matter because they shape whether advice is trusted and acted upon. While IAM focuses on information adoption as the downstream outcome of evaluation, human–AI research distinguishes *trust* from *reliance*. Trust generally reflects a user’s attitude toward the competence, reliability, or trustworthiness of an advisor or system, whereas reliance concerns whether its recommendation is incorporated into a decision [45, 64]. The two are related but not interchangeable: users may report Trust without following a recommendation, or rely on advice without correspondingly high reported trust [59, 63, 64]. Advice-taking research similarly emphasizes that receiving or valuing advice is distinct from the extent to which it is incorporated into one’s final judgment [9]. This motivates us to treat *Overall Quality*, *Trust*, and *Intended Reliance* as distinct downstream responses to the appraisals users form.

Moreover, this framework is particularly relevant to financial decision-making contexts [33]. Characteristics of decision context may shape which cues become salient and how people interpret them. For example, the magnitude of potential consequences can affect how carefully people approach a decision [37, 72], uncertainty can increase the value of external guidance [9], and differences in verifiability can affect how readily the quality of a recommendation can be assessed against established principles [9, 44]. Finance-specific work similarly shows that decision risk can alter how information quality contributes to perceived usefulness [33]. The same evaluative process may also vary across individuals. Financial knowledge, risk tolerance, familiarity with a decision, and perceived expertise can affect users’ ability or confidence in evaluating external advice [4, 26, 28, 84], while AI literacy, prior experience, and broader orientations toward AI can shape expectations about AI-mediated advice [5, 30, 35, 60]. Advice-taking and persuasion research similarly suggest that decision context and individual characteristics can shape how people weight different evaluative cues when considering external advice [9, 14, 43, 62]. We therefore consider these factors as potential boundary conditions on this information-evaluation process, examining whether they shape the appraisals users form and moderate the downstream associations of those appraisals with *Trust* and *Intended Reliance*.

2.3 Explicit and Implicit Cues in Financial Advice: Source Labels vs. Advice Style

Advice is evaluated not only through its substantive content, but also through cues about who or what produced it. Source information can activate expectations about expertise, credibility, and competence [14, 62], and human–AI studies show that responses to otherwise comparable advice can differ depending on whether it is labeled as coming from a human or algorithmic source [47, 80]. Source identity, however, is not communicated through labels alone. We use *advice style* to refer to message-level cues such as voice, tone, reasoning organization, experiential grounding, expressions of authority, and recommendation framing. Professional financial advice commonly emphasizes client circumstances,

financial principles, trade-offs, and longer-term consequences [21, 50], whereas online financial communities more often incorporate clarification, multiple viewpoints, empathy, and personal experience [78]. Comparisons of AI- and human-authored responses similarly identify differences in linguistic and interactional patterns, including more structured and neutral presentation in AI responses and greater conversational or experiential grounding in human responses [68, 69, 78]. Recent work similarly shows that generative AI can reproduce source-associated identity cues, including language suggestive of lived experience, without possessing the underlying experience those cues imply [?]. Such stylistic differences can themselves affect reader preferences and perceptions [76, 81, 86]. We therefore treat these patterns not as fixed properties of AI, professional, or community sources, but as recognizable source-associated communication profiles that may provide evaluative cues independently of an explicit source label.

Distinguishing the effect of advice style from source labels matters because these mechanisms have distinct design implications for how trust and reliance are shaped. Prior work faces a trade-off between experimental control and ecological validity. Identical-text designs can isolate the effect of source labels but remove the communication cues through which source identities are often conveyed in practice [47, 80]; naturalistic comparisons preserve those cues but may simultaneously vary substantive content, reasoning, style, and advice quality [7, 41]. This makes it difficult to determine whether observed differences arise from *how* advice is communicated or from *who* users believe produced it. This distinction also matters for AI governance, because professional presentation can be reproduced independently of the expertise, review, or accountability that a source label may imply. We address this gap by standardizing financial facts, numerical values, core reasoning, and recommendation direction across AI-, Expert-, and Online Community-style advice, while independently varying the source labels displayed alongside it.

3 Study Design and Data

3.1 Study Overview

We conducted a preregistered randomized vignette experiment to examine how people evaluate personal financial advice as a function of *advice style* and *source labels* [38]. Across eight financial scenarios, we constructed advice in three styles associated with an AI Financial Assistant (AI), Certified Financial Planner (Expert), and Online Community (OC). AI-style advice was generated first and verified by the first author for factual accuracy and alignment with the intended recommendation. Corresponding Expert- and OC-style versions were then manually crafted while holding constant the scenario-specific facts, numerical values, recommendation direction, and core financial rationale, while varying source-associated communication style, tone, and reasoning format. Source labels were independently manipulated as matching (*labeled condition*), absent (*unlabeled condition*), or incorrect (*misabeled condition*). Participants first completed a repeated-measures evaluation task in which they rated four scenario–advice pairs on 10 items each. They subsequently completed a comparative ranking task in which all three advice styles were shown for a previously unseen scenario. Figure 1 summarizes the complete experimental design.

3.2 Participants, Recruitment, and Ethics

Participants were recruited through Prolific. Eligibility was restricted to U.S. residents aged 18–70 who were fluent in English, had completed at least 100 previous Prolific submissions, and had an approval rate of at least 95%. The final analytic sample consisted of $N = 285$ participants (*labeled condition*: $n = 100$, *unlabeled condition*: $n = 92$, and *misabeled condition*: $n = 93$) after excluding incomplete responses, failed attention checks, and submissions completed in under two minutes, yielding 1,140 vignette-level observations in the primary evaluation task and 285 observations in the

Trustworthy FinAIInce: Study Design Overview

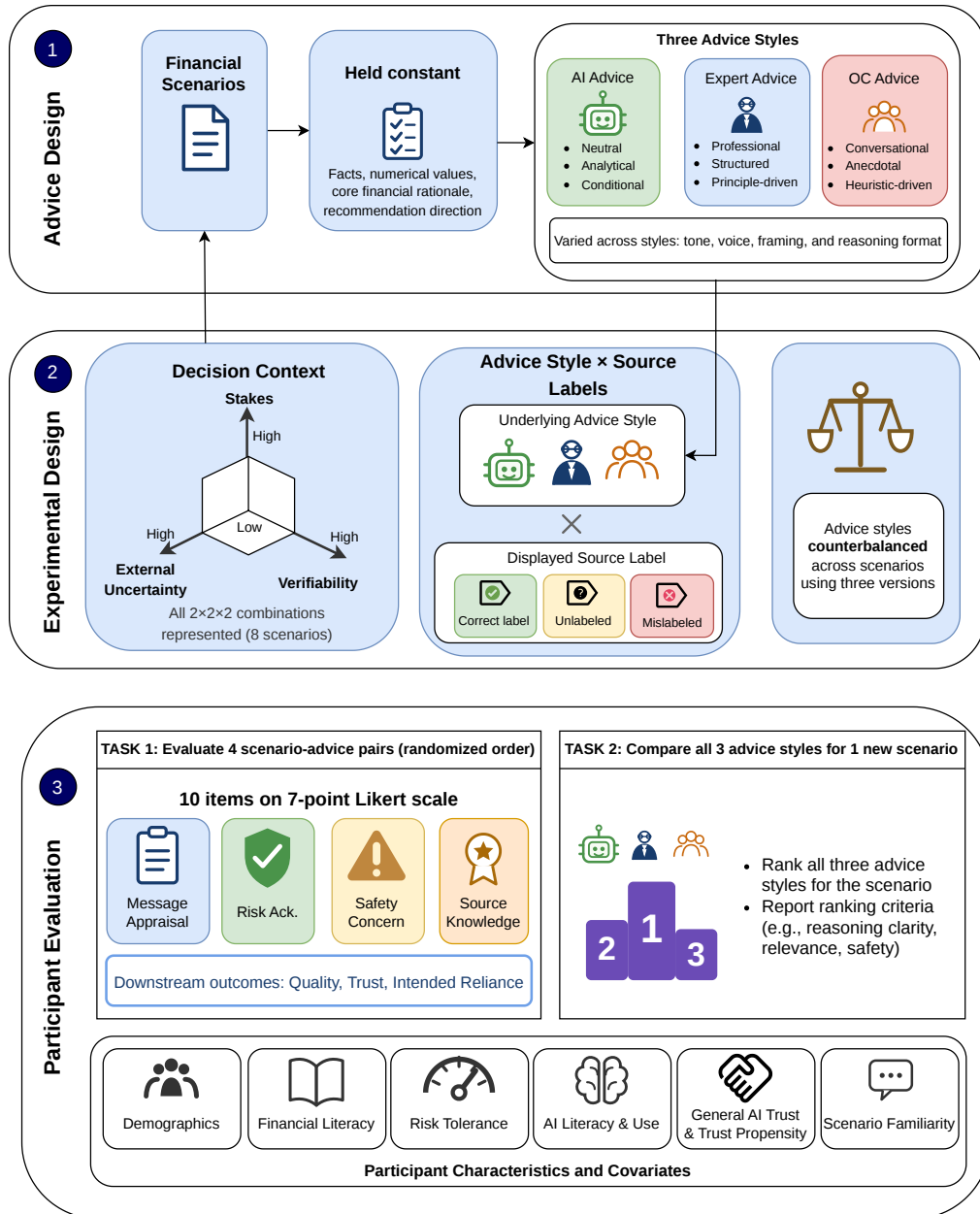


Fig. 1. Overview of the study design. Eight financial scenarios represent all combinations of stakes, external uncertainty, and verifiability. Advice was constructed in AI, Expert, and Online Community (OC) styles while preserving scenario-specific facts, numerical values, core financial rationale, and recommendation direction. Source labels were independently manipulated as *correctly labeled*, *unlabeled*, or *mislabelled*. Participants first evaluated four scenario–advice pairs and then ranked all three advice styles for one previously unseen scenario. Advice styles were counterbalanced across scenarios, and participant characteristics were also measured.

Table 1. Participant demographics and background characteristics ($N = 285$). Continuous variables are reported as mean (SD); categorical variables are reported as n (%).

Characteristic	Final analytic sample
Demographics	
Age	39.76 (12.22)
Sex	Male: 143 (50.2%); Female: 137 (48.1%); Prefer not to say: 3 (1.1%); Unavailable: 2 (0.7%)
Race/ethnicity	White: 178 (62.5%); Black: 41 (14.4%); Asian: 34 (11.9%); Mixed: 15 (5.3%); Other: 10 (3.5%); Prefer not to say: 3 (1.1%); Unavailable: 4 (1.4%)
Household income	Median category: \$50,000–\$74,999
Financial Background	
Objective financial literacy (0–3)	2.66 (0.68)
Self-rated financial literacy	Very low: 2 (0.7%); Low: 20 (7.0%); Moderate: 154 (54.0%); High: 90 (31.6%); Very high: 19 (6.7%)
Financial risk tolerance (1–4)	No financial risk: 36 (12.6%); Average risk: 161 (56.5%); Above-average risk: 77 (27.0%); Substantial risk: 11 (3.9%)
AI Experience and Attitudes	
AI use frequency	Never: 9 (3.2%); Rarely: 20 (7.0%); Occasionally: 42 (14.7%); Regularly: 80 (28.1%); Frequently: 134 (47.0%)
AI use for personal finance	No, and have not considered it: 53 (18.6%); No, but have considered it: 43 (15.1%); Used once or twice: 102 (35.8%); Used multiple times: 87 (30.5%)
General Trust in AI (1–7)	4.63 (1.50)
Human Trust Propensity (1–7)	4.61 (1.44)
Perceived AI Literacy (1–7)	5.16 (0.99)

comparative ranking task. Table 1 summarizes participant demographics and background characteristics. Participants received \$3 for completing the 15-minute survey. The study was reviewed and approved by the Institutional Review Board at our university.

3.3 Financial Decision Scenarios

3.3.1 Decision-Context Taxonomy. We constructed eight personal-finance scenarios using a $2 \times 2 \times 2$ factorial taxonomy spanning three complementary properties of financial decisions: *stakes*, *external uncertainty*, and *verifiability*. *Stakes* captures the magnitude and persistence of a decision’s financial consequences; high-stakes decisions involved larger or longer-term consequences than low-stakes decisions. *External uncertainty* captures whether outcomes depend on unpredictable future events, such as market movements or unexpected expenses; high-uncertainty decisions depended more strongly on such events. *Verifiability* captures whether the quality of a recommendation can be evaluated against an established financial principle; high-verifiability decisions had a clearer normative benchmark than low-verifiability decisions which were preference-driven. The eight scenarios represented all combinations of these three dimensions. Experimental vignette methodology enables systematic variation of theoretically relevant attributes while retaining concrete decision contexts [2, 6].

3.3.2 Scenario Construction. Each of the eight combinations of these dimensions was represented by a real-world financial scenario, covering decisions such as housing, credit card debt, travel insurance, emergency savings, graduate

education, debt repayment and budgeting, long-term investing, and concentrated stock risk. Scenarios were approximately 100–150 words and described a protagonist’s financial position, relevant numerical information, the competing considerations, and a concrete advice-seeking question. We operationalized the three dimensions before constructing the final advice stimuli and reviewed each scenario against the corresponding stakes, uncertainty, and verifiability criteria. Table A1 in the appendix summarizes the eight scenarios and their factorial assignments.

3.4 Advice Stimulus Construction

We first generated an AI-style advice response using a prompt that instructed the model (ChatGPT (OpenAI; GPT-5.3 Instant)) to act as a “neutral AI financial assistant” and respond to the financial decision described in the scenario. We then manually verified each response for factual accuracy and alignment with the intended financial recommendation. Corresponding Expert- and Online Community (OC)-style responses were then manually crafted to preserve the scenario-specific facts, numerical values, recommendation direction, and core financial rationale while varying source-associated advice style, tone, and reasoning format. Construction was informed by public financial-advice materials, including online financial communities, financial-planning resources, and conversational AI responses.

The three conditions were designed to capture recurring communication patterns associated with contemporary financial-advice sources rather than to represent all communication from any source category. *AI-style* advice used a neutral, impersonal voice with explicit analytical and conditional reasoning. *Expert-style* advice used a professional, measured tone with principle-based reasoning, explicit trade-offs, and longer-term planning considerations. *OC-style* advice used a conversational voice, experiential framing, informal language, and practical heuristics. Table A2 summarizes the style specifications.

3.4.1 Stimulus Validation. We conducted complementary linguistic, human, semantic, and external-reference checks to assess whether the constructed responses reflected the intended advice styles while preserving key substantive content. First, we compared 12 linguistic features associated with the intended style dimensions, including lexical complexity, formality, personal voice, conditional language, hedging, and contractions [8, 17, 68, 77]. Using Friedman tests across the eight matched scenario triplets with Benjamini–Hochberg FDR correction, 10 of 12 features differed across styles. The Categorical–Dynamic Index (CDI) [61] showed a clear gradient from AI ($M = 27.36$) to Expert ($M = 21.30$) to OC ($M = 12.03$), $\chi^2(2) = 14.25$, $p < .001$, with all pairwise contrasts remaining significant after FDR correction.

Three independent annotators, blind to the intended style labels, then classified each of the 24 advice texts and evaluated substantive consistency within each scenario. The intended style was identified in 91.7% of individual judgments, with majority classification matching all 24 stimuli (Fleiss’ $\kappa = .749$). Across all eight scenario triplets, majority judgments indicated preservation of the recommendation direction, scenario-specific facts, and core financial rationale. Annotators flagged potentially material additional information in five triplets, primarily in OC responses, typically reflecting the anecdotal or experiential framing intentionally associated with that style. As a complementary semantic check, Sentence-Transformers embeddings (all-mpnet-base-v2) [65] showed substantially greater similarity within than across scenarios (AI–Expert: .844 vs. .304; AI–OC: .728 vs. .282; Expert–OC: .798 vs. .293); across 24 directed cross-style comparisons, the nearest semantic counterpart always came from the same financial scenario (100% top-1 retrieval).

Finally, we compared the constructed stimuli with a topic-matched reference corpus of 72 naturally occurring financial-advice texts from AI systems, financial professionals, and online communities. Several broad linguistic patterns reproduced across the constructed and reference texts, including differences in lexical complexity, first-person language,

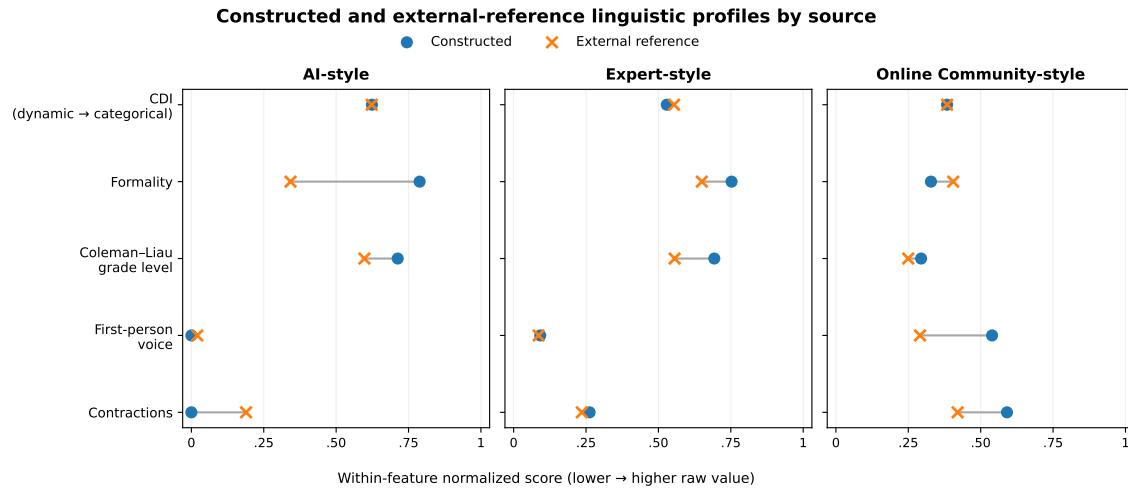


Fig. 2. Correspondence between constructed and external-reference linguistic profiles across AI, Expert, and Online Community advice. Source-level means are normalized within each linguistic dimension using the combined distribution of constructed and reference texts. Circles denote constructed stimuli, crosses denote external-reference texts, and connecting-line length reflects the discrepancy between the two profiles within each feature. Higher values indicate more categorical language for CDI and higher raw values for the remaining measures. Normalized positions should be interpreted within, rather than across, linguistic dimensions. Exact values and statistical comparisons are reported in Appendix A.

contractions, and CDI, although individual cues such as formality, conditional language, and hedging did not always follow the same ordering. We therefore interpret the manipulation as capturing controlled *source-associated advice styles*, rather than literal representations of all communication from each source category. Figure 2 summarizes the external comparison.

3.5 Experimental Manipulation and Assignment

Participants were randomly assigned between subjects to one of three source-label conditions. In the *correctly labeled* condition, advice was displayed with its corresponding source label (AI, Expert, or Online Community); in the *unlabeled* condition, no source label was shown; and in the *misabeled* condition, advice was displayed with an incorrect source label. Two counterbalanced mislabeling schemas were used. In Schema A, AI-, Expert-, and OC-style advice were displayed with OC, AI, and Expert labels, respectively, in Task 1, and with Expert, OC, and AI labels in Task 2. Schema B reversed these mappings across the two tasks. When present, source labels appeared before the advice.

The eight scenarios were divided into two groups of four. Participants were randomly assigned to a scenario group and one of three counterbalanced advice-style versions, such that each scenario appeared in each advice style across participants. Each participant evaluated all three advice styles across their four scenarios, with one style appearing twice, and scenario order was randomized. For the comparative ranking task, participants received a randomly selected, previously unseen scenario from the other scenario group and viewed all three advice-style versions side by side.

3.6 Procedure

After providing informed consent, participants completed background questions on financial advice seeking and participant characteristics. They then completed four randomized scenario evaluations. For each, participants read a

financial vignette, viewed one advice response under their assigned source-label condition, reported scenario familiarity, and rated the advice on ten evaluation items. After an attention check, participants completed the comparative ranking task using a previously unseen scenario from the other scenario group. They viewed all three advice-style versions side by side, ranked them from most to least preferred, and indicated the criteria that influenced their ranking. The survey concluded with the remaining participant measures and debriefing.

3.7 Measures

3.7.1 Advice Evaluation. After each scenario–advice pair, participants rated ten items on seven-point Likert scales assessing readability, reasoning clarity, situational fit, risk acknowledgment, financial-harm risk, misleadingness, source knowledge, overall quality, Trust, and intended Reliance. For analysis, readability, reasoning clarity, and situational fit were combined as *Message Appraisal*; financial-harm risk and misleadingness were combined as *Safety Concern*; and *Risk Acknowledgment* and *Source Knowledge* were retained as single-item appraisals. *Overall Quality*, *Trust*, and *Intended Reliance* were retained as distinct downstream outcomes. Trust captured whether participants would trust the advice in the protagonist’s situation, whereas intended Reliance captured whether they would feel comfortable following it. Exact item wording appears in Table A4.

3.7.2 Participant Characteristics. We measured participant characteristics that could shape financial-advice evaluation. Objective financial literacy was assessed using the “Big Three” questions on compound interest, inflation, and diversification [48, 49], alongside self-rated financial literacy. Financial risk tolerance was measured using the Survey of Consumer Finances item [26]. AI-related measures included AI-use frequency, prior AI use for personal finance, perceived AI literacy (PAILQ-6) [27], and General Trust in AI [51]. We additionally measured Human Trust Propensity [22], prior use of financial advice sources, and scenario-specific familiarity.

3.7.3 Comparative Ranking Task. The comparative ranking task provided a direct within-scenario comparison of the three advice styles. Participants viewed the AI, Expert, and OC advice for one previously unseen scenario and ranked the three responses from most to least preferred. They then indicated the considerations that informed their ranking, including readability, clarity of reasoning, safety, situational relevance, overall reliability, and alignment with their own thinking or preferences. Participants also reported their confidence in their ranking on a 5-point scale.

3.8 Analysis Strategy

Because each participant evaluated four pieces of advice, our primary analyses used linear mixed-effects models with participant-level random intercepts to account for repeated observations. Advice-evaluation outcomes were standardized before modeling. We first examined which upstream factors—advice style, source labels, decision context, and scenario familiarity—were associated with the four appraisals (RQ1). We then examined how these appraisals were associated with *Overall Quality*, *Trust*, and *Intended Reliance*, sequentially incorporating preceding evaluation variables to characterize the broader evaluative pathway (RQ2). Table 2 summarizes the model specifications, inferential tests, and multiple-comparison corrections used throughout.

We conducted two sets of complementary analyses. First, to disentangle advice style from source labels, we tested whether their associations were additive or interactive and decomposed their appraisal-carried pathways using participant-level bootstrap resampling. Second, we examined potential boundary conditions by testing both the direct associations of participant characteristics and whether individual differences or decision context moderated downstream evaluation pathways. Where multiple related hypotheses were tested, we applied the correction procedures specified in

Table 2. Summary of the analysis strategy.

Analysis	Model / Method	Inference
RQ1: What factors shape people’s appraisals of financial advice?		
Appraisals	Appraisal $\sim$ Upstream Factors + (1 Participant)	LMM; β , p
Participant characteristics	Appraisal $\sim$ Upstream Factors + Participant Characteristics + (1 Participant)	Joint model; BH-FDR
RQ2: How are appraisals associated with Overall Quality, Trust, and intended Reliance?		
Upstream baseline	Outcome $\sim$ Upstream Factors + (1 Participant)	R_m^2 , R_c^2
Overall Quality	Overall Quality $\sim$ Upstream Factors + Four Appraisals + (1 Participant)	LMM; β , p , R^2
Trust	Trust $\sim$ Upstream Factors + Four Appraisals + Overall Quality + (1 Participant)	LMM; β , p , R^2
Intended Reliance	Reliance $\sim$ Upstream Factors + Four Appraisals + Overall Quality + Trust + (1 Participant)	LMM; β , p , R^2
Disentangling Advice Style and Source Labels		
Style $\times$ Label	Outcome $\sim$ Base Model + Advice Style $\times$ Source Label	LRT; BH-FDR for contrasts
Pathway decomposition	Advice Style / Source Label $\rightarrow$ Appraisals $\rightarrow$ Overall Quality $\rightarrow$ Trust $\rightarrow$ Reliance	2,000 bootstraps; 95% CI
Boundary Conditions		
Participant characteristics	Trust $\sim$ RQ2 Trust Model + Participant Characteristics Reliance $\sim$ RQ2 Reliance Model + Participant Characteristics	β , p ; BH-FDR
Individual-difference moderation	Trust $\sim$ Upstream Factors + Evaluation Predictors + Moderator + Moderator $\times$ Evaluation Predictors + (1 Participant) Reliance $\sim$ Upstream Factors + Evaluation Predictors + Trust + Moderator + Trust $\times$ Moderator + (1 Participant)	LRT χ^2 ; BH-FDR
Decision-context moderation	Trust $\sim$ Upstream Factors + Evaluation Predictors + Context $\times$ Evaluation Predictors + (1 Participant) Reliance $\sim$ Upstream Factors + Evaluation Predictors + Trust + Context $\times$ (Evaluation Predictors + Trust) + (1 Participant)	LRT χ^2 ; BH-FDR
Other Analyses		
Comparative ranking	First-choice frequencies and complete within-participant rankings	χ^2 ; Friedman; Kendall’s W ; Wilcoxon + Holm
Robustness	Within-between decomposition; GEE; scenario fixed effects; constituent-item models	Sensitivity checks

Note. Upstream Factors comprise Advice Style, Source Label, Stakes, External Uncertainty, Verifiability, and Scenario Familiarity. Four Appraisals comprise Message Appraisal, Risk Acknowledgment, Safety Concern, and Source Knowledge. Evaluation Predictors comprise the Four Appraisals and Overall Quality. Participant Characteristics comprise the seven individual-difference measures entered jointly. Moderator denotes one focal participant characteristic, and Context denotes one focal decision-context dimension (Stakes, External Uncertainty, or Verifiability). Base Model denotes the corresponding additive model before the focal interaction. All mixed-effects models include participant-level random intercepts. BH-FDR denotes Benjamini–Hochberg false-discovery-rate correction; LRT denotes likelihood-ratio test.

Table 2. Because appraisals and downstream judgments were measured within the same advice-evaluation task, we interpret the pathway analyses as associational rather than as evidence of causal mediation.

Finally, we analyzed the side-by-side ranking task separately as convergent evidence using first-choice and complete-ranking comparisons. We also conducted a series of robustness checks addressing within-person variation, model specification, scenario parameterization, and composite construction. Additional statistical details and robustness results are provided in the appendix.

4 Results

4.1 Organizing Advice-Evaluation Measures into Appraisals

We organized the 10 advice-evaluation items according to the role they play in the advice-evaluation process. Building on the Information Adoption Model (IAM), which distinguishes evaluations of the message from evaluations of its source [75], we averaged *readability*, *reasoning clarity*, and *situational fit* into a *Message Appraisal* index ($\alpha = .887$). These items capture how understandable, well-reasoned, and applicable the advice appears. We retained *Source Knowledge* separately as a source-credibility judgment. Because financial advice can appear clear and credible while still presenting potential downside or epistemic risk, we evaluated risk and safety separately, consistent with prior financial-AI work that

Table 3. Construction and internal consistency of the four advice appraisals across the 1,140 advice evaluations. Higher *Safety Concern* scores indicate greater perceived risk of financial harm or misleadingness. Dashes indicate statistics that are not applicable to single-item appraisals.

Appraisal	Items	<i>k</i>	α	Mean inter-item <i>r</i>
Message Appraisal	Readability; reasoning clarity; situational fit	3	.887	.72
Risk Acknowledgment	Risk acknowledgment	1	–	–
Safety Concern	Financial-harm risk; misleadingness	2	.712	.55
Source Knowledge	Perceived source knowledge	1	–	–

distinguishes information quality from perceived risk in advice evaluation [33]. *Risk of financial harm* and *misleadingness* were averaged into a *Safety Concern* index ($\alpha = .712$; inter-item $r = .55$), with higher values indicating greater concern. *Risk Acknowledgment* was retained separately because recognizing potential risks in the advice is conceptually different from judging the advice itself as harmful or misleading. Table 3 presents information on the four appraisals. Finally, we retained *Overall Quality*, *Trust*, and *Intended Reliance* as distinct downstream judgments, allowing us to examine how these initial appraisals relate to broader evaluation, trust, and willingness to follow the advice.

4.2 Advice Style, Source Labels, and Decision Context Shape Different Appraisals

Advice style, source labels, and decision context were associated with different appraisal profiles (Table 4).

Advice style primarily shaped Message Appraisal and Safety Concern. Compared with AI-style advice, both Expert- and Online Community (OC)-style advice received higher *Message Appraisal* ratings ($\beta_{\text{Expert}} = .334, p < .001$; $\beta_{\text{OC}} = .223, p < .001$) and lower *Safety Concern* ($\beta_{\text{Expert}} = -.245, p < .001$; $\beta_{\text{OC}} = -.127, p = .026$). The styles also differed in *Source Knowledge*: Expert-style advice was perceived as coming from a more knowledgeable source than AI-style advice ($\beta = .141, p = .025$), whereas OC-style advice was perceived as less knowledgeable ($\beta = -.174, p = .006$), producing an Expert–AI–OC ordering on perceived source knowledge. *Risk acknowledgment* showed a different pattern: Expert and AI styles did not significantly differ ($\beta = -.061, p = .340$), while OC-style advice was rated lower than AI-style advice ($\beta = -.147, p = .021$).

Source labels produced more selective differences. Relative to unlabeled advice, an Expert label was associated with greater *Source Knowledge* ($\beta = .225, p = .017$) and higher *Message Appraisal* ($\beta = .190, p = .035$). AI and OC labels did not differ significantly from the unlabeled condition across the four appraisals. Thus, source labels were most clearly reflected in perceived source expertise.

Decision context primarily shaped risk and safety evaluations. High-stakes scenarios increased *Safety Concern* ($\beta = .436, p < .001$) and reduced *Message Appraisal* ($\beta = -.241, p < .001$). Greater external uncertainty increased *Risk Acknowledgment* ($\beta = .310, p < .001$), while higher verifiability reduced both *Risk Acknowledgment* ($\beta = -.199, p = .010$) and *Safety Concern* ($\beta = -.292, p < .001$). Scenario familiarity was associated with more favorable evaluations across several dimensions: greater *Message Appraisal* ($\beta = .319, p < .001$), greater *Risk Acknowledgment* ($\beta = .140, p < .001$), greater perceived *Source Knowledge* ($\beta = .206, p < .001$), and lower *Safety Concern* ($\beta = -.125, p < .001$).

Individual differences shaped selected appraisals. Using the same base model, we examined participant characteristics jointly and found selective associations with the four appraisals (Table 5). Greater AI Literacy was associated with higher *Message Appraisal* ($\beta = .189, p_{\text{FDR}} < .001$) and *Source Knowledge* ($\beta = .178, p_{\text{FDR}} < .001$). General AI Trust ($\beta = .168, p_{\text{FDR}} = .001$) and Objective Financial Literacy ($\beta = .086, p_{\text{FDR}} = .038$) were also positively associated with

Table 4. Mixed-effects models predicting the four initial appraisals of financial advice. Coefficients are standardized fixed-effect estimates (β). Models include a participant-level random intercept. AI-style advice and the unlabeled condition are the reference categories. Higher *Safety Concern* indicates greater perceived potential for financial harm or misleadingness.

	Message Appraisal	Risk Ack.	Safety Concern [†]	Source Knowledge
<i>Advice style</i>				
Expert vs. AI	+0.334***	-0.061	-0.245***	+0.141*
OC vs. AI	+0.223***	-0.147*	-0.127*	-0.174**
<i>Source labels</i>				
AI label vs. Unlabeled	+0.075	-0.088	-0.141	+0.049
Expert label vs. Unlabeled	+0.190*	-0.026	-0.166	+0.225*
OC label vs. Unlabeled	+0.168	-0.044	-0.119	+0.091
<i>Decision context</i>				
High stakes	-0.241***	+0.083	+0.436***	+0.010
High external uncertainty	-0.072	+0.310***	+0.026	-0.005
High verifiability	+0.099	-0.199**	-0.292***	+0.036
<i>Covariate</i>				
Scenario familiarity	+0.319***	+0.140***	-0.125***	+0.206***
Marginal R^2 (R_m^2)	0.170	0.054	0.115	0.067
Conditional R^2 (R_c^2)	0.405	0.298	0.458	0.312

Note. $N = 1,140$ evaluations from 285 participants. Scenario familiarity is standardized; binary experimental predictors are coded 0/1. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5. Participant characteristics associated with initial advice appraisals.

Participant characteristic	Appraisal	β	p_{FDR}
AI Literacy	Message Appraisal	.189	<.001
AI Literacy	Source Knowledge	.178	<.001
General AI Trust	Source Knowledge	.168	.001
Objective Financial Literacy	Source Knowledge	.086	.038
Risk Tolerance	Safety Concern	.145	.015
Risk Tolerance	Source Knowledge	-.117	.011

Note. Each appraisal was modeled separately using a linear mixed-effects model with participant-level random intercepts. Models included the same base predictors (advice style, source labels, decision context, and scenario familiarity), with all seven participant characteristics entered simultaneously. Benjamini–Hochberg FDR correction was applied across the seven participant-characteristic coefficients within each appraisal. Only associations surviving $p_{FDR} < .05$ are shown.

Source Knowledge. Greater Risk Tolerance was associated with higher *Safety Concern* ($\beta = .145$, $p_{FDR} = .015$) and lower *Source Knowledge* ($\beta = -.117$, $p_{FDR} = .011$). Overall, individual differences affected particular appraisals rather than shifting advice appraisal uniformly.

Overall, these results reveal differentiated entry points into advice appraisal: advice style was most broadly and consistently associated with message and safety appraisals, Expert source labels with perceived source knowledge, decision context with risk and safety appraisals, and individual differences with selected appraisal dimensions.

4.3 From Appraisals to Quality, Trust, and Intended Reliance

We next examined how the four appraisals were associated with *Overall Quality*, *Trust*, and intended *Reliance*. For each outcome, we compared an upstream model containing advice style, source labels, decision context, and scenario familiarity with a model that additionally included the preceding appraisal variables (Table 6).

Appraisals were strongly associated with Overall Quality. All four intermediate appraisals were independently associated with *Overall Quality*. *Source Knowledge* showed the strongest association ($\beta = .489, p < .001$), followed by *Message Appraisal* ($\beta = .325, p < .001$). Greater *Safety Concern* predicted lower *Overall Quality* ($\beta = -.088, p < .001$), while greater *Risk Acknowledgment* predicted somewhat higher quality ratings ($\beta = .075, p < .001$). Adding the appraisals also increased marginal explained variance from 9.8% to 69.2%. We also observe that initial Expert-style and Expert-label associations with *Overall Quality* were substantially attenuated after accounting for these appraisals.

Overall Quality was the strongest correlate of Trust. Once the appraisals and *Overall Quality* were included, *Overall Quality* was by far the strongest correlate of *Trust* ($\beta = .607, p < .001$). *Message Appraisal* ($\beta = .132, p < .001$), lower *Safety Concern* ($\beta = -.125, p < .001$), and *Source Knowledge* ($\beta = .109, p < .001$) retained smaller independent associations with *Trust*, whereas *Risk Acknowledgment* did not. Marginal explained variance increased from 11.2% in the upstream model to 75.9% in the appraisal- and quality-adjusted model. The initial Expert-style association with *Trust* was also no longer detectable after including these variables.

Trust was the strongest correlate of Intended Reliance. *Trust* was the dominant correlate of *Intended Reliance* ($\beta = .594, p < .001$), followed by *Overall Quality* ($\beta = .206, p < .001$); the remaining appraisals showed substantially smaller associations: *Message Appraisal* ($\beta = .079, p < .001$), lower *Safety Concern* ($\beta = -.050, p < .001$), and *Source Knowledge* ($\beta = .047, p = .020$). Adding the full evaluation pathway increased marginal explained variance from 14.8% to 82.9%. After accounting for these variables, initial advice-style and source-label differences in *Reliance* were substantially reduced. That said, because we measured these intermediate judgments within the same advice evaluation, we interpret this structure as an *evaluative pathway* rather than evidence of causal mediation.

Across all three downstream outcomes, models including the appraisal pathway explained substantially more variance than upstream factors alone, while the initial advantages associated with Expert advice style and Expert source labels were largely attenuated after accounting for these appraisals. Taken together, the results are consistent with a progressively organized evaluative pathway in which advice style, source labels, and decision context first shape distinct message, risk, safety, and source-related appraisals; these appraisals are strongly associated with *Overall Quality*; *Overall Quality* is in turn strongly associated with *Trust*; and *Trust* is the strongest correlate of *Intended Reliance*.

4.4 Advice Style and Source Labels Operate Additively Through Different Appraisals

In a subsequent exploratory analysis, we tested whether the associations of advice style depended on the source label displayed with the advice. Across the four appraisals and three downstream outcomes, adding Advice Style $\times$ Source Label interactions did not improve model fit (all $p > .29$; Table A5). Thus, these results support treating advice style and displayed source labels as largely additive rather than interactive effects.

We then examined whether these two cues were reflected differently through the appraisal pathway. Participant-level bootstrap decompositions showed that advice style was relatively more strongly reflected through *Message Appraisal*, whereas source labels were relatively more strongly reflected through *Source Knowledge*. For Expert advice, the Message-versus-Knowledge pathway was more pronounced for advice style than for the corresponding Expert label ($\Delta\Delta = .081$, 95% CI [.014, .148]); the same pattern appeared for Online Community advice ($\Delta\Delta = .124$, 95% CI [.045, .212]). These results suggest that how advice is written is reflected more strongly in appraisals of the message itself, whereas explicit source labels shift appraisal relatively toward perceived source expertise. Full pathway decompositions and simple contrasts are reported in Table A6 and Table A7, respectively.

Table 6. Upstream and pathway-adjusted models of Overall Quality, Trust, and Intended Reliance.

Predictor	Overall Quality		Trust		Intended Reliance	
	Upstream	Pathway-adjusted	Upstream	Pathway-adjusted	Upstream	Pathway-adjusted
<i>Advice style</i>						
Expert vs. AI	+0.205***	+0.011	+0.185**	-0.028	+0.173**	-0.019
OC vs. AI	-0.004	+0.005	+0.009	-0.009	+0.023	+0.002
<i>Source labels</i>						
AI label vs. Unlabeled	+0.126	+0.074	+0.066	-0.038	+0.103	+0.029
Expert label vs. Unlabeled	+0.248**	+0.064	+0.176	-0.046	+0.210*	+0.013
OC label vs. Unlabeled	+0.170	+0.060	+0.154	+0.004	+0.189*	+0.040
<i>Decision context</i>						
High stakes	-0.040	+0.069*	-0.107*	-0.000	-0.155**	-0.048
High external uncertainty	-0.013	-0.008	+0.001	+0.014	+0.017	+0.030
High verifiability	+0.152	+0.091*	+0.166*	+0.026	+0.178*	+0.022
<i>Covariate</i>						
Scenario familiarity	+0.256***	+0.038*	+0.280***	+0.040*	+0.324***	+0.049***
<i>Appraisal-to-reliance pathway</i>						
Message Appraisal	—	+0.325***	—	+0.132***	—	+0.079***
Risk Acknowledgment	—	+0.075***	—	+0.024	—	-0.015
Safety Concern [†]	—	-0.088***	—	-0.125***	—	-0.050***
Source Knowledge	—	+0.489***	—	+0.109***	—	+0.047*
Overall Quality	—	—	—	+0.607***	—	+0.206***
Trust	—	—	—	—	—	+0.594***
<i>Model fit</i>						
Marginal R^2 (R_m^2)	0.098	0.692	0.112	0.759	0.148	0.829
Conditional R^2 (R_c^2)	0.373	0.742	0.351	0.795	0.362	0.829

Note. Cells report standardized fixed-effect coefficients (β) from linear mixed-effects models with participant-level random intercepts. Upstream models include advice style, source labels, decision context, and scenario familiarity. Pathway-adjusted models additionally include the four initial appraisals; the Trust model further includes Overall Quality, and the Reliance model further includes Overall Quality and Trust. AI-style advice and the unlabeled condition are the reference categories. Models include $N = 285$ participants and 1,140 advice evaluations. R_m^2 and R_c^2 denote marginal and conditional R^2 , respectively. Orange and blue shading indicate significant positive and negative coefficients, respectively; shading intensity reflects statistical significance ($p < .05$, $p < .01$, and $p < .001$), not effect magnitude. Unshaded coefficients are not statistically significant. [†] Higher values indicate greater perceived financial harm or misleadingness. * $p < .05$, ** $p < .01$, *** $p < .001$.

4.5 Boundary Conditions of the Evaluation Pathway

We next examined whether the relationships among appraisals, *Overall Quality*, *Trust*, and *Intended Reliance* varied across decision contexts or participant characteristics.

Decision context did not substantially alter the downstream pathway. Although stakes, external uncertainty, and verifiability shaped the initial appraisals participants formed, we found no robust evidence that any of the three decision-context dimensions moderated how those appraisals translated into *Trust* or intended *Reliance* after FDR correction. Thus, decision context appeared to matter primarily at the appraisal stage rather than in the subsequent evaluation pathway.

Individual differences produced selective moderation. Objective Financial Literacy strengthened the associations of *Source Knowledge* with *Trust* ($\beta_{\text{int}} = .054$, $p_{\text{FDR}} = .017$) and of *Trust* with *Intended Reliance* ($\beta_{\text{int}} = .039$, $p_{\text{FDR}} = .002$). Human Trust Propensity selectively moderated the association between *Safety Concern* and *Trust* ($\beta_{\text{int}} = .056$, $p_{\text{FDR}} =$

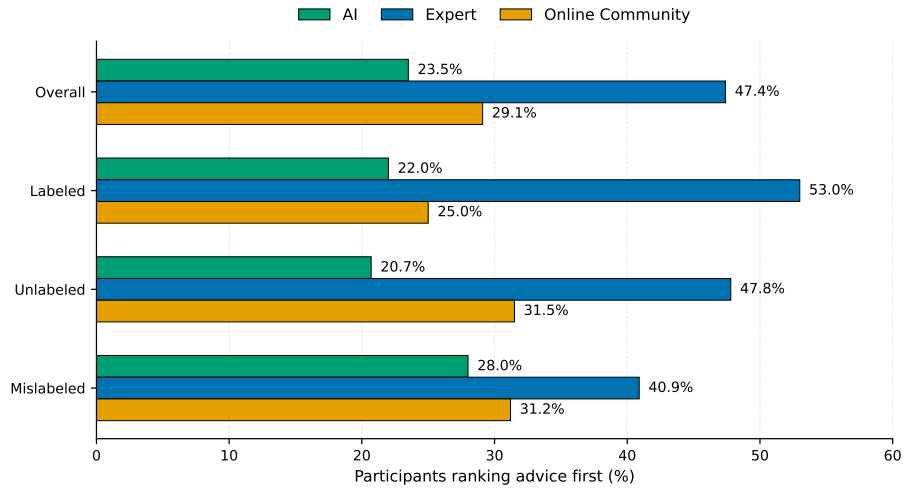


Fig. 3. Percentage of participants ranking each advice style first in the comparative ranking task, overall and by source label condition. Expert-style advice was most frequently ranked first across all conditions. Although its descriptive advantage was smaller in the Mislabeled condition, the overall Advice Style $\times$ Source Label Arm interaction was not statistically significant.

.002), with safety concerns weighing more strongly among participants lower in trust propensity. General AI Trust was associated with higher overall *Trust* ($\beta = .202$, $p_{\text{FDR}} < .001$) and *Intended Reliance* ($\beta = .227$, $p_{\text{FDR}} < .001$), but these associations were substantially reduced after accounting for the appraisal pathway.

Overall, the downstream pathway was comparatively stable across decision contexts, while individual differences altered only selected parts of how appraisals translated into Trust and Reliance. More details are reported in Appendix B.2.

4.6 Comparative Ranking Provides Convergent Evidence

As a final comparative task, participants ranked AI-, Expert-, and Online Community (OC)-style advice side by side for a previously unseen financial scenario. Expert-style advice was ranked first by 47.4% of participants, compared with 29.1% for OC advice and 23.5% for AI advice ($\chi^2(2) = 26.61$, $p < .001$). Across the complete rankings, the three advice styles also differed significantly (Friedman $\chi^2(2) = 37.16$, $p < .001$, Kendall's $W = .065$). Pairwise Wilcoxon signed-rank tests with Holm correction showed that Expert-style advice was preferred over both AI- and OC-style advice (both $p_{\text{Holm}} < .001$), whereas AI and OC did not differ. Notably, the Expert-style preference remained in the Unlabeled condition, where 47.8% ranked it first, compared with 31.5% for OC and 20.7% for AI (Figure 3). We found no evidence that the relative ranking of the three styles systematically differed across source-label conditions ($\chi^2(4) = 6.65$, $p = .155$).

Participants' self-reported ranking criteria further emphasized properties of the advice itself. Clear reasoning (65.6%), ease of reading and understanding (60.0%), situational relevance (53.3%), and acknowledgment of risks and uncertainties (50.9%) were the most frequently selected criteria, whereas only 7.0% reported that the displayed source influenced their ranking (Figure 4). Although these self-reports do not identify what causally determined participants' rankings, the comparative task provides convergent evidence for the primary results: participants preferred Expert-style advice and predominantly described their preferences in terms of message-level appraisal cues rather than explicit source information.

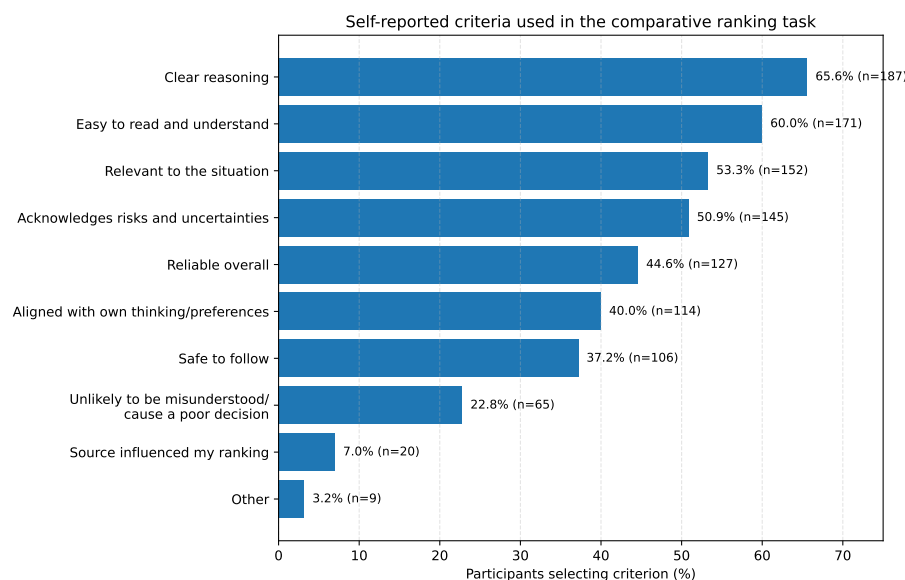


Fig. 4. Self-reported criteria used in the comparative ranking task. Participants could select multiple criteria. Percentages indicate the proportion of participants who selected each criterion. Clear reasoning, readability, situational relevance, and acknowledgment of risks and uncertainties were the most frequently reported considerations, whereas explicit source influence was comparatively rare.

4.7 Robustness Checks

We conducted four robustness checks to assess whether the appraisal pathway depended on stable participant response tendencies, the repeated-measures estimator, the representation of decision context, or the construction of the appraisal composites. First, within-between participant decompositions showed that the principal associations remained at the within-participant level, indicating that the pathway was not driven solely by participants who generally rated all advice more favorably. Second, generalized estimating equations closely reproduced the mixed-effects estimates; for example, the *Overall Quality–Trust* association was $\beta = .608$ (vs. $.607$ in the primary model) and the *Trust–Reliance* association was $\beta = .595$ (vs. $.594$). Third, replacing the decision-context dimensions with scenario fixed effects produced nearly identical pathway estimates. Finally, replacing *Message Appraisal* and *Safety Concern* with their constituent items preserved the main conclusions. Overall, the appraisal–Quality–Trust–Reliance pattern was stable across alternative specifications and measurement choices. More details are reported in Appendix C.

5 Discussion

The landscape of personal financial decision-making is rapidly evolving as generative AI becomes an increasingly accessible source of financial guidance. As these systems become more prevalent, the central challenge is not simply whether people trust AI-generated advice, but whether that trust and subsequent reliance are grounded in an appropriate evaluation of the advice. Existing trust-calibration approaches commonly rely on cues such as disclaimers, source labels, communication choices, and anthropomorphic design, but examining their direct relationship with Trust provides only a partial account of how users arrive at these judgments. Our work instead frames trust and reliance as part of a broader

evaluation process, showing that advice style, source labels, and decision context are associated with distinct appraisals of the message, its risks and safety, and the source's perceived knowledge. We discuss the implications of this work.

5.1 Extending Information Adoption for Financial Advice Evaluation

A growing body of work on AI-generated information and human-AI advisory systems has approached users' responses through the lens of information evaluation and adoption [13, 25, 33, 46]. Classical IAM accounts, as discussed previously, organize this process around evaluations of *argument quality* and *source credibility*, which shape perceived usefulness and subsequent adoption [75]. Our work extends this perspective by opening up the process through which financial advice becomes trusted and relied upon. Rather than treating *Trust* as a direct response to an advice source or interface cue, our results characterize it as part of a broader evaluation process in which people form distinct appraisals of the message, its risks and safety, and its source, which are associated with *Overall Quality*, *Trust*, and ultimately *Intended Reliance*. This more granular view matters because similar levels of trust can mask very different underlying appraisal profiles. For example, Online-Community-style advice was appraised more favorably than AI-style advice on message and safety dimensions, but less favorably on perceived *Source Knowledge*; these differences largely offset downstream, leaving trust in the two styles statistically indistinguishable. Looking beneath aggregate trust therefore provides a more diagnostic account of *why* advice is trusted and which message-, risk-, safety-, or source-related cues may warrant design attention, consistent with work on appropriate reliance that does not treat higher Trust as inherently desirable [45, 63].

Our findings also elaborate this evaluation process for financial decisions by making risk- and safety-related appraisals explicit. Financial advice can appear clear, relevant, and credible while still failing to acknowledge uncertainty or exposing users to potential harm. Recent work on AI-assisted financial decision-making similarly highlights perceived risk and uncertainty as important to advice adoption [33]. Our results show that these appraisals were sensitive to characteristics of the decision itself: greater external uncertainty was associated with greater *Risk Acknowledgment* ($\beta = .310$), while higher stakes were associated with greater *Safety Concern* ($\beta = .436$); higher verifiability was associated with lower levels of both. These patterns suggest that perceptions of risk and safety are closely associated with the characteristics of the financial decision in which the advice is encountered. Decision context therefore appears to enter the evaluation process by influencing which aspects of uncertainty and potential harm become salient to users, extending information-evaluation accounts beyond properties of the message and its source alone.

Finally, our findings reinforce the distinction between *Trust* and *Intended Reliance*. Trust is commonly treated as an antecedent of reliance on automated systems, but the two are not equivalent: Lee and See [45] emphasize that the central design problem is whether trust supports *appropriate reliance*, while advice-taking research similarly distinguishes attitudes toward a source from whether its recommendation is incorporated into a decision [9]. In our results, *Trust* was the strongest correlate of *Intended Reliance*, yet *Overall Quality*, *Message Appraisal*, and *Safety Concern* retained smaller independent associations after accounting for Trust. Participant-level variation further supported this distinction: scenario familiarity was directly associated with greater Reliance ($\beta = .089$, $p < .001$) despite a weaker association with Trust, while objective financial literacy strengthened the Trust-Reliance relationship ($\beta_{\text{int}} = .039$, $p_{\text{FDR}} = .002$). Thus, similar levels of *Trust* need not translate into similar intentions to act. This aligns with prior work suggesting that studies of AI-supported decision-making should measure Trust and Reliance separately and, where possible, complement stated intentions with behavioral measures of whether users accept, reject, or verify advice [63].

5.2 Advice Style and Source Labels as distinct epistemic cues

Prior HCI research shows that how AI communicates can shape how users evaluate and trust it. Authoritative language can increase Trust even when system limitations are disclosed [53], while social and personality cues such as friendliness, extroversion, and stance adaptation can similarly influence responses to AI [74, 76]. Parallel work shows that displayed source information can act as a separate credibility cue: identical information may be evaluated differently depending on whether it is attributed to AI, a human, or an expert [34, 66, 70, 73]. Our findings bring these two strands together by independently varying advice style and source labels, showing that they were associated with distinct appraisal profiles. Advice style was reflected more broadly in message- and safety-related appraisals. Consistent with this pattern, Expert-style advice remained the most frequently preferred in the comparative ranking task even when no source label was displayed. Expert source labels, in contrast, were most clearly associated with perceived *Source Knowledge*. Their associations were also largely additive rather than interactive, suggesting that sounding like an expert and being identified as an expert provide distinct rather than interchangeable signals.

From an AI-governance perspective, this distinction raises concerns around provenance and accountability. Generative AI can readily reproduce the linguistic markers of professional advice—such as structured reasoning, measured language, and explicit trade-offs—without necessarily carrying the expertise, review, or responsibility associated with a qualified professional. This gap between *presentational authority* and *epistemic authority* may encourage unwarranted confidence when polished communication is mistaken for evidence of expertise. Our findings therefore suggest that advice style and source information should be treated as distinct layers of evaluation and governance in AI-assisted financial advice. Systems should make clear who produced, reviewed, or endorsed a recommendation and where responsibility lies, rather than allowing professional presentation alone to signal authority. This distinction is already relevant in deployed systems: *SoFi* reports that its financial planners were involved in evaluating responses from Coach, illustrating how AI-generated advice may incorporate human professional review without being directly authored by a human advisor [12].

5.3 Design Implications for AI-Mediated Financial Advice

Our findings suggest three considerations for the design and evaluation of AI-mediated financial advice systems.

Design for grounded evaluation and appropriate Reliance. Financial AI systems should help users assess *why* a recommendation deserves Trust and whether it warrants action, rather than optimizing Trust through surface-level cues such as professionalized advice style alone. This challenge is already visible in deployed financial AI systems: in testing its AI financial guide, Coach, *SoFi* reported that effective financial guidance depends not only on the underlying financial logic, but also on how recommendations and trade-offs are communicated to users [12]. Interfaces could make relevant assumptions, reasoning, trade-offs, uncertainty, and potential downside consequences easier to inspect—for example, through expandable assumptions or a brief “Why this recommendation?” view. Such support should ultimately help users decide when to follow, verify, reject, or escalate a recommendation, including by facilitating comparison with other sources or suggesting professional consultation when appropriate. Evaluations of these systems should likewise distinguish *Trust* from *Intended Reliance* and, where possible, examine actual advice-taking and verification behavior rather than positive perceptions alone.

Separate professional presentation from verifiable epistemic authority. Clear and professional communication can improve how advice is evaluated, but it should not substitute for evidence of expertise. Financial advice systems should

distinguish whether advice was *AI-generated*, *professionally reviewed*, or *directly authored by a qualified advisor*; make the relevant expertise or credentials behind that involvement visible; expose the evidence or financial principles supporting the recommendation; and clarify where responsibility lies and what limitations apply. Provenance could function as an independently inspectable layer alongside the advice itself, rather than as a decorative “Expert” badge. This separation can preserve the benefits of clear communication without allowing presentational authority alone to stand in for warranted epistemic authority.

Adapt evaluation support to the decision and the user. Decision context was reflected primarily in the appraisals users formed; interfaces could therefore foreground different information depending on what makes a financial decision difficult to evaluate. For example, high-stakes decisions might emphasize potential downside consequences, uncertain decisions might make assumptions and contingencies explicit, and difficult-to-verify decisions could make trade-offs and alternative courses of action more explicit rather than presenting a single recommendation as definitive. User-sensitive support could similarly help people evaluate cues such as source expertise or safety concerns, rather than assuming that all users interpret these cues in the same way. Lastly, because greater familiarity was consistently associated with more favorable appraisals of trust and Reliance, financial AI systems may need to guard against familiarity reducing scrutiny. Even when a scenario feels familiar to the user, interfaces should preserve lightweight checks on assumptions, risks, and verifiability so reliance is based on evaluating the advice rather than familiarity with the situation alone.

5.4 Limitations and Future Work

Our work has several limitations that also suggest interesting future directions. First, *Reliance* was measured as behavioral intention rather than observed financial behavior. Real financial decisions often unfold over time, involve monetary consequences, and include consultation with other people or information sources. Future work should therefore examine whether the evaluation patterns observed here predict actual advice taking, verification, information seeking, and decision outcomes, including whether users appropriately accept, reject, or escalate recommendations.

Second, although our results are consistent with an evaluative organization linking appraisals, *Overall Quality*, *Trust*, and *Intended Reliance*, these constructs were measured within the same advice- evaluation task. The observed associations therefore should not be interpreted as establishing a causal or temporal pathway. Future work could test this ordering more directly through longitudinal or experimental designs that manipulate specific appraisals and observe subsequent changes in Trust and Reliance.

Third, we used controlled financial scenarios and deliberately constructed advice styles. This allowed us to hold the underlying recommendation relatively constant while varying its presentation and source label, but no small set of styles can capture the diversity of AI, professional, or online-community advice. Real-world AI advice is also interactive: users ask follow-up questions, introduce new constraints, challenge recommendations, and observe how systems respond. Extending this work to multi-turn and longitudinal interactions is therefore an important next step.

Fourth, our decision-context dimensions were instantiated through eight specific financial scenarios. Although robustness analyses suggest that the downstream pattern was not driven by individual scenarios, other financial decisions may involve factors not represented here, including urgency, irreversibility, regulatory complexity, social consequences, or ambiguity in user goals. Similarly, our sample consisted of U.S. adults recruited through Prolific; financial norms, access to professional advice, and expectations of AI and expertise may differ across populations and institutional settings. Replication across financial domains, countries, and real advisory settings will therefore be important for establishing generalizability.

Finally, individual characteristics such as financial literacy, AI literacy, and risk tolerance were measured rather than experimentally manipulated. Accordingly, their associations and moderation patterns should not be interpreted causally. Future work could test whether interventions that improve financial knowledge, source-verification skills, or risk comprehension change how users appraise advice and translate Trust into Reliance over time. Longitudinal studies could also examine how these relationships evolve as users accumulate experience with an AI system and observe the consequences of its advice.

6 Conclusion

As generative AI becomes an increasingly common source of personal finance advice, the challenge is not simply whether people trust its advice, but whether that trust is grounded in how they evaluate it. In a randomized experiment with 285 U.S. adults, we show that advice style, source labels, and decision context shaped distinct appraisals of the message, its risks and safety, and its source. These appraisals were strongly associated with *Overall Quality*, which was in turn closely associated with *Trust*, while Trust was the strongest correlate of *Intended Reliance*. Advice style and source labels contributed different signals: style was associated more broadly with message- and safety-related appraisals, whereas Expert labels were most clearly associated with perceived *Source Knowledge*. Decision context and individual differences further shaped selected parts of this evaluation process. Together, these findings shift the design goal from making financial AI appear trustworthy toward supporting grounded evaluation: helping users assess the reasoning, risks, source expertise, and context of advice so that Trust and Reliance better reflect what the recommendation warrants.

Acknowledgments

Need to add this: AI declaration, annotators, etc

References

- [1] 2026. As More U.S. Consumers Struggle with Rising Prices, Many Turn to Artificial Intelligence for Financial Advice. <https://www.jdpower.com/business/resources/more-us-consumers-struggle-rising-prices-many-turn-artificial-intelligence/>
- [2] Herman Aguinis and Kyle J. Bradley. 2014. Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods* 17, 4 (Oct. 2014), 351–371. doi:10.1177/1094428114547952
- [3] Tom Akana and Larry Santucci. 2025. Trust and Credibility: Comparing Sources of Financial Advice. (2025). <https://www.philadelphiafed.org/consumer-finance/trust-and-credibility-comparing-sources-of-financial-advice>
- [4] Sam Allgood and William B. Walstad. 2016. THE EFFECTS OF PERCEIVED AND ACTUAL FINANCIAL LITERACY ON FINANCIAL BEHAVIORS. *Economic Inquiry* 54, 1 (Jan. 2016), 675–697. doi:10.1111/ecin.12255
- [5] Reshaa F. Alruwaili, Abdullah A. Alasmari, Hussien Tash Niyazi, I. K. Youssef, Hanan Aifan, and Somia A. Asklany. 2025. Modeling public trust in AI cognitive capabilities using statistical and machine learning approaches. *Scientific Reports* 15, 1 (Nov. 2025), 39856. doi:10.1038/s41598-025-23447-4
- [6] Christiane Atzmüller and Peter M. Steiner. 2010. Experimental Vignette Studies in Survey Research. *Methodology* 6, 3 (Jan. 2010), 128–138. doi:10.1027/1614-2241/a000014
- [7] John W. Ayers, Adam Poliak, Mark Dredze, Eric C. Leas, Zechariah Zhu, Jessica B. Kelley, Dennis J. Faix, Aaron M. Goodman, Christopher A. Longhurst, Michael Hogarth, and Davey M. Smith. 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine* 183, 6 (June 2023), 589–596. doi:10.1001/jamainternmed.2023.1838
- [8] Nikolay Babakov, David Dale, Ilya Gusev, Irina Krotova, and Alexander Panchenko. 2023. Don't Lose the Message While Paraphrasing: A Study on Content Preserving Style Transfer. In *Natural Language Processing and Information Systems: 28th International Conference on Applications of Natural Language to Information Systems, NLDB 2023, Derby, UK, June 21–23, 2023, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 47–61. doi:10.1007/978-3-031-35320-8_4
- [9] Silvia Bonaccio and Reeshad S. Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101, 2 (Nov. 2006), 127–151. doi:10.1016/j.obhdp.2006.07.001
- [10] Prince Bosompim, Richmond Essieku, and Jodi Letkiewicz. 2026. In whom we trust? Social networks and trust in AI vs. human experts financial advice. *Finance Research Letters* 91 (March 2026), 109437. doi:10.1016/j.frl.2025.109437

- [11] Zana Bućinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 188:1–188:21. doi:10.1145/3449287
- [12] Caitlin Mullen. 2026. What SoFi learned testing its AI adviser Coach | Banking Dive. <https://www.bankingdive.com/news/sofi-bank-ai-adviser-coach-financial-planning/822329/>
- [13] Mark Anthony Camilleri and Adriana Caterina Camilleri. 2024. The Acceptance and Usage of ChatGPT: An Information Adoption Model perspective. In *2024 8th International Conference on Communications and Future Internet (ICCFI)*. 61–66. doi:10.1109/ICCFI63302.2024.00017
- [14] Shelly Chaiken. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39, 5 (1980), 752–766. doi:10.1037/0022-3514.39.5.752
- [15] Xusen Cheng, Xiaoping Zhang, Jason Cohen, and Jian Mou. 2022. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management* 59, 3 (May 2022), 102940. doi:10.1016/j.ipm.2022.102940
- [16] Taha Choukmane, Tim de Silva, Weidong Lin, and Matthew Akuzawa. 2026. AI Financial Advice: Supply, Demand, and Life Cycle Implications. doi:10.2139/ssrn.6446286
- [17] Meri Coleman and Ta Lin Liao. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology* 60, 2 (1975), 283.
- [18] Ismail Erkan and Chris Evans. 2016. The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior* 61 (Aug. 2016), 47–55. doi:10.1016/j.chb.2016.03.003
- [19] Douglas (DJ) Fairhurst and Daniel Greene. 2025. How Much Does ChatGPT Know about Finance? *Financial Analysts Journal* 81, 1 (Jan. 2025), 12–32. doi:10.1080/0015198X.2024.2411941
- [20] George Fatouros, Kostas Metaxas, John Soldatos, and Dimosthenis Kyriazis. 2025. Can Large Language Models beat wall street? Evaluating GPT-4's impact on financial decision-making with MarketSenseAI. *Neural Computing and Applications* 37, 30 (Oct. 2025), 24893–24918. doi:10.1007/s00521-024-10613-4
- [21] Financial Planning Standards Board. [n. d.]. Standards for the profession. <https://fpsb.org/standards-for-the-profession/>
- [22] M. Lance Frazier, Paul D. Johnson, and Stav Fainshmidt. 2013. Development and validation of a propensity to trust scale. *Journal of Trust Research* 3, 2 (Oct. 2013), 76–97. doi:10.1080/21515581.2013.820026
- [23] Isabel O. Gallegos, Chen Shani, Weiyan Shi, Federico Bianchi, Izzy Gainsburg, Dan Jurafsky, and Robb Willer. 2026. Labeling messages as AI-generated does not reduce their persuasive effects. *PNAS nexus* 5, 2 (Feb. 2026), pgag008. doi:10.1093/pnasnexus/pgag008
- [24] Adalia Y. H. Goh, Andree Hartanto, and Nadyanna M. Majeed. 2025. Generative artificial intelligence dependency: Scale development, validation, and its motivational, behavioral, and psychological correlates. *Computers in Human Behavior Reports* 20 (Dec. 2025), 100845. doi:10.1016/j.chbr.2025.100845
- [25] Nannan Gong, Joohwan Seo, and Huanzhen Liu. 2026. When algorithms speak: Boundary conditions of information adoption in AI- and user-generated content. *Journal of Retailing and Consumer Services* 93 (Aug. 2026), 104926. doi:10.1016/j.jretconser.2026.104926
- [26] John Grable and Ruth H. Lytton. 1999. Financial risk tolerance revisited: the development of a risk assessment instrument. *Financial Services Review* 8, 3 (Sept. 1999), 163–181. doi:10.1016/S1057-0810(99)00041-4
- [27] Simone Grassini. 2024. A Psychometric Validation of the PAILQ-6: Perceived Artificial Intelligence Literacy Questionnaire. In *Nordic Conference on Human-Computer Interaction*. ACM, Uppsala Sweden, 1–10. doi:10.1145/3679318.3685359
- [28] Matthew J. Grawitch, Steven L. Winton, and Srikanth P. Mudigonda. 2026. Who listens to ChatGPT and when should they? A two-study examination of AI-assisted decision making. *Computers in Human Behavior: Artificial Humans* 7 (March 2026), 100283. doi:10.1016/j.chbah.2026.100283
- [29] Oudom Hean, Utsha Saha, and Binita Saha. 2025. Can AI help with your personal finances? *Applied Economics* 57, 60 (Dec. 2025), 11219–11227. doi:10.1080/00036846.2025.2450384
- [30] Kuo-Ting Huang and Christopher Ball. 2024. The Influence of AI Literacy on User's Trust in AI in Practical Scenarios: A Digital Divide Pilot Study. *Proceedings of the Association for Information Science and Technology* 61, 1 (Oct. 2024), 937–939. doi:10.1002/pra2.1146
- [31] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2025. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems* 43, 2 (Jan. 2025), 42:1–42:55. doi:10.1145/3703155
- [32] Gallup Inc. 2026. Where Americans and Canadians Turn for Financial Guidance. <https://news.gallup.com/poll/712952/americans-canadians-turn-financial-guidance.aspx> Section: Economy.
- [33] Athina Ioannou, M. Mahdi Tavalaei, Dorteia Vatn, and Patrick Mikalef. 2026. The Role of Explainability In AI-Driven Financial Decision Making. *Information Systems Frontiers* (April 2026). doi:10.1007/s10796-026-10727-1
- [34] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T. Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300469
- [35] Sarah A. Jessup, Tamera R. Schneider, Gene M. Alarcon, Tyler J. Ryan, and August Capiola. 2019. The Measurement of the Propensity to Trust Automation. In *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, Jessie Y.C. Chen and Gino Fragomeni (Eds.). Springer International Publishing, Cham, 476–489. doi:10.1007/978-3-030-21565-1_32
- [36] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Ho Shu Chan, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* 55, 12 (Dec. 2023), 1–38. doi:10.1145/3571730 arXiv: 2202.03629 [cs].

- [37] Barbara E. Kahn and Jonathan Baron. 1995. An Exploratory Study of Choice Rules Favored for High-Stakes Decisions. *Journal of Consumer Psychology* 4, 4 (Jan. 1995), 305–328. doi:10.1207/s15327663jcp0404_01
- [38] Aryan Ramchandra Kapadia, Eshwar Chandrasekharan, and Koustuv Saha. 2026. How People Evaluate AI-, Expert-, and Peer-Style Financial Advice. *arXiv preprint arXiv:2608.09019* (2026).
- [39] Kemal Kirtac and Guido Germano. 2024. Sentiment trading with large language models. *Finance Research Letters* 62 (April 2024), 105227. doi:10.1016/j.frl.2024.105227
- [40] Artur Klingbeil, Cassandra Grützner, and Philipp Schreck. 2024. Trust and reliance on AI – An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior* 160 (Nov. 2024), 108352. doi:10.1016/j.chb.2024.108352
- [41] Harsh Kumar, Jasmine Chahal, Yinuo Zhao, Zeling Zhang, Annika Z Wei, Louis Tay, and Ashton Anderson. 2026. When AI Gives Advice: Evaluating AI and Human Responses to Online Advice-Seeking for Well-Being. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3772318.3791233
- [42] Saurav Sathish Kumta, Niyaz Panakaje, S. M. Riha Parvin, and Ata Mohammad Nikzad. 2026. Can you smell what AI is advising? Human, machine, or both: A comparative experimental study on financial advice. *Acta Psychologica* 264 (April 2026), 106530. doi:10.1016/j.actpsy.2026.106530
- [43] Juliane E. Kämmer, Shoham Choshen-Hillel, Johannes Müller-Trede, Stephanie L. Black, and Jürgen Weibler. 2023. A systematic review of empirical studies on advice-based decisions in behavioral and organizational research. *Decision* 10, 2 (2023), 107–137. doi:10.1037/dec0000199
- [44] Patrick R. Laughlin and John Adamopoulos. 1980. Social combination processes and individual learning for six-person cooperative groups on an intellectual task. *Journal of Personality and Social Psychology* 38, 6 (1980), 941–947. doi:10.1037/0022-3514.38.6.941
- [45] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (March 2004), 50–80. doi:10.1518/hfes.46.1.50_30392
- [46] Chenxi Li, Yixun Lin, Ruqing Chen, and Jing (Elaine) Chen. 2025. How do users adopt AI-generated content (AIGC)? An exploration of content cues and interactive cues. *Technology in Society* 81 (June 2025), 102830. doi:10.1016/j.techsoc.2025.102830
- [47] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (March 2019), 90–103. doi:10.1016/j.obhdp.2018.12.005
- [48] Annamaria Lusardi and Olivia S. Mitchell. 2011. Financial literacy around the world: an overview. *Journal of Pension Economics and Finance* 10, 4 (Oct. 2011), 497–508. doi:10.1017/S1474747211000448
- [49] Annamaria Lusardi and Olivia S. Mitchell. 2014. The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature* 52, 1 (March 2014), 5–44. doi:10.1257/jel.52.1.5
- [50] Kirsten L. MacDonald, Ellana Loy, Mark Brimble, and Karen Wildman. 2023. The value of personal professional financial advice to clients: A systematic quantitative literature review. *Accounting & Finance* 63, 4 (2023), 4399–4429. doi:10.1111/acfi.13099
- [51] Melanie J. McGrath, Oliver Lack, James Tisch, and Andreas Duenser. 2025. Measuring trust in artificial intelligence: validation of an established scale and its short form. *Frontiers in Artificial Intelligence* 8 (May 2025), 1582880. doi:10.3389/frai.2025.1582880
- [52] Christoph Merkle. 2025. AI Appreciation and Financial Advice. doi:10.2139/ssrn.5186803
- [53] Luise Metzger, Linda Miller, Martin Baumann, and Johannes Kraus. 2024. Empowering Calibrated (Dis-)Trust in Conversational Agents: A User Study on the Persuasive Power of Limitation Disclaimers vs. Authoritative Style. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3613904.3642122
- [54] Markus Münster, Felix Reichenbach, and Martin Walther. 2024. Robinhood, Reddit, and the news: The impact of traditional and social media on retail investor trading. *Journal of Financial Markets* 71 (Nov. 2024), 100929. doi:10.1016/j.finmar.2024.100929
- [55] Paweł Niszczota and Sami Abbas. 2023. GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. *Finance Research Letters* 58 (Dec. 2023), 104333. doi:10.1016/j.frl.2023.104333
- [56] OECD. 2026. *Artificial intelligence and personal finance*. OECD Artificial Intelligence Papers. doi:10.1787/2858fdf4-en Edition: 62 Series: OECD Artificial Intelligence Papers.
- [57] Paige Ouimet and Geoffrey Tate. 2020. Learning from Coworkers: Peer Effects on Individual Investment Decisions. *The Journal of Finance* 75, 1 (2020), 133–172. doi:10.1111/jofi.12830
- [58] Tae-Young Pak. 2026. How individuals use generative AI for personal financial management. *Journal of Behavioral and Experimental Finance* 49 (March 2026), 101145. doi:10.1016/j.jbef.2026.101145
- [59] Olivia Pal, Veda Duddu, Agam Goyal, Drishti Goel, and Koustuv Saha. 2026. Do We Know What They Know We Know? Calibrating Student Trust in AI and Human Responses through Mutual Theory of Mind. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3772363.3799386
- [60] Zilong Pan, Ozias A. Moore, Antigoni Papadimitriou, and Jiayan Zhu. 2025. AI literacy and trust: A multi-method study of Human-GAI team collaboration. *Computers in Human Behavior: Artificial Humans* 4 (May 2025), 100162. doi:10.1016/j.chbah.2025.100162
- [61] James W. Pennebaker, Cindy K. Chung, Joey Frazee, Gary M. Lavergne, and David I. Beaver. 2014. When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE* 9, 12 (Dec. 2014), e115844. doi:10.1371/journal.pone.0115844
- [62] Richard E. Petty and John T. Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, Richard E. Petty and John T. Cacioppo (Eds.). Springer, New York, NY, 1–24. doi:10.1007/978-1-4612-4964-1_1
- [63] Muhammad Raees, Vassilis-Javed Khan, Ioanna Lykourantzou, and Konstantinos Papangelis. 2026. Do People Appropriately Rely on AI-Advice? An Analytical Review of HCI Research on Human-AI Decision-Making. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing*

- Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3772318.3791467
- [64] Muhammad Raees and Konstantinos Papangelis. 2026. Trust to Reliance: Measurement Constructs for Human-AI Appropriate Reliance. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3772363.3798835
- [65] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. doi:10.48550/arXiv.1908.10084 arXiv:1908.10084 [cs].
- [66] Moritz Reis, Florian Reis, and Wilfried Kunde. 2024. Influence of believed AI involvement on the perception of digital medical advice. *Nature Medicine* 30, 11 (Nov. 2024), 3098–3100. doi:10.1038/s41591-024-03180-7
- [67] Evan F. Risko and Sam J. Gilbert. 2016. Cognitive Offloading. *Trends in Cognitive Sciences* 20, 9 (Sept. 2016), 676–688. doi:10.1016/j.tics.2016.07.002
- [68] Koustuv Saha, Yoshee Jain, Chunyu Liu, Sidharth Kaliappan, and Ravi Karkar. 2026. AI vs. Humans for Online Support: Comparing the Language of Responses from LLMs and Online Communities of Alzheimer's Disease. *ACM Transactions on Computing for Healthcare* 7, 3 (June 2026), 61:1–61:39. doi:10.1145/3709366
- [69] Koustuv Saha, Yoshee Jain, Violeta J. Rodriguez, and Munmun De Choudhury. 2026. Linguistic comparison of AI- and human-written responses to online mental health queries. *npj Artificial Intelligence* (May 2026). doi:10.1038/s44387-026-00099-x
- [70] Siavosh Sahebi, Paul Formosa, and Sarah Banks. 2026. The AI penalty and disclosure paradox: Trust, authenticity and knowledge uptake in AI-mediated communication. *Computers in Human Behavior: Artificial Humans* 8 (May 2026), 100304. doi:10.1016/j.chbah.2026.100304
- [71] Minh Tam Tammy Schlosky and Sterling Raskie. 2025. ChatGPT as a Financial Advisor: A Re-Examination. *Journal of Risk and Financial Management* 18, 12 (Dec. 2025), 664. doi:10.3390/jrfm18120664
- [72] P. Slovic. 1987. Perception of risk. *Science* 236, 4799 (April 1987), 280–285. doi:10.1126/science.3563507
- [73] Xin Sun, Rongjun Ma, Shu Wei, Pablo Cesar, Jos A. Bosch, and Abdallah El Ali. 2026. Understanding trust toward human versus AI-generated health information through behavioral and physiological sensing. *International Journal of Human-Computer Studies* 209 (Feb. 2026), 103714. doi:10.1016/j.ijhcs.2025.103714
- [74] Yuan Sun and Ting Wang. 2026. Be Friendly, Not Friends: How LLM Sycophancy Shapes User Trust. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3772318.3791079
- [75] Stephanie Watts Sussman and Wendy Schneider Siegal. 2003. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. *Information Systems Research* 14, 1 (March 2003), 47–65. doi:10.1287/isre.14.1.47.14767
- [76] Takehiro Takayanagi, Kiyoshi Izumi, Javier Sanz-Cruzado, Richard McCreadie, and Iadh Ounis. 2025. Are Generative AI Agents Effective Personalized Financial Advisors?. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25)*. Association for Computing Machinery, New York, NY, USA, 286–295. doi:10.1145/3726302.3729897
- [77] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (March 2010), 24–54. doi:10.1177/0261927X09351676
- [78] Sachin Thukral, Suyash Sangwan, Vipul Chauhan, Arnab Chatterjee, and Lipika Dey. 2024. Generating insights about financial asks from Reddit posts and user interactions. In *Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '23)*. Association for Computing Machinery, New York, NY, USA, 294–299. doi:10.1145/3625007.3627313
- [79] Deepak Verma, Prem Prakash Dewani, Abhishek Behl, and Yogesh K. Dwivedi. 2023. Understanding the impact of eWOM communication through the lens of information adoption model: A meta-analytic structural equation modeling perspective. *Computers in Human Behavior* 143 (June 2023), 107710. doi:10.1016/j.chb.2023.107710
- [80] Kailas Vodrahalli, Roxana Daneshjou, Tobias Gerstenberg, and James Zou. 2022. Do Humans Trust Advice More if it Comes from AI?: An Analysis of Human-AI Interactions. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Oxford United Kingdom, 763–777. doi:10.1145/3514094.3534150
- [81] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3411764.3445645
- [82] Jing Jian Xiao and Kexin Meng. 2025. AI and personal finance: evaluating rules of thumb generated by ChatGPT. *International Journal of Bank Marketing* 43, 9 (2025), 1969–1989.
- [83] Cathy (Liu) Yang, Kevin Bauer, Xitong Li, and Oliver Hinz. 2026. My Advisor, Her AI, and Me: Evidence from a Field Experiment on Human-AI Collaboration and Investment Decisions. *Management Science* 72, 1 (2026), 242–264. https://ideas.repec.org/a/inm/ormnsc/v72y2026i1p242-264.html
- [84] Ilan Yaniv and Eli Kleinberger. 2000. Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes* 83, 2 (Nov. 2000), 260–281. doi:10.1006/obhd.2000.2909
- [85] Alex Zarifis and Xusen Cheng. 2024. How to Build Trust in Answers Given by Generative AI for Specific and Vague Financial Questions. doi:10.1108/JEBDE-11-2023-0028
- [86] Jiawei Zhou, Kritika Venkatachalam, Minje Choi, Koustuv Saha, and Munmun De Choudhury. 2026. Communication styles and reader preferences of LLM- and human-authored COVID-19 information explanations: a case study. *BMC Artificial Intelligence* 2, 1 (June 2026), 10. doi:10.1186/s44398-026-00030-z
- [87] Qiuhan Zhu, Xiangnan Li, Yiang Dong, Pengcheng Chang, and Mengmeng Fan. 2026. Not all cognitive offloading is equal: distinguishing dependent and autonomous offloading to generative AI. *Frontiers in Psychology* 17 (July 2026). doi:10.3389/fpsyg.2026.1878629

A Study Design Details

Table A1. Decision-context assignments for the eight financial scenarios.

Scenario	Stakes	External Uncertainty	Verifiability
Renting: Shared vs. Private	Low	Low	Low
Credit Card vs. New Laptop	Low	Low	High
Travel Insurance	Low	High	Low
Emergency Fund vs. Investing Bonus	Low	High	High
Graduate School vs. Continuing Work	High	Low	Low
High-Interest Debt Repayment and Budgeting	High	Low	High
Long-Term Investment Planning	High	High	Low
Concentrated Stock Position Risk	High	High	High

Table A2. Operationalization of the three source-associated advice styles.

Feature	AI advice	Expert advice	OC advice
Voice	Impersonal; no first-person identity	Mild first-person professional authority	First-person peer perspective
Tone	Neutral and analytical	Professional and measured	Informal and conversational
Reasoning	Explicit analytical and conditional reasoning	Principle-based reasoning with explicit trade-offs and longer-term planning	Experiential reasoning supported by practical heuristics
Framing	Balances available options and financial risk exposure	Connects the decision to financial foundations and broader goals	Uses lived experience, subjective judgment, and relatable consequences
Structure	Situation and trade-off; analysis; conditional recommendation	Quantified framing; trade-off explanation; actionable principle	Position or anecdote; experiential lesson; direct recommendation

Note. These profiles were designed as controlled, source-associated styles rather than literal representations of all communication from each source category. Across versions of a scenario, scenario-specific facts, numerical values, recommendation direction, and core financial rationale were preserved while style, tone, and reasoning presentation varied.

Table A3. Selected linguistic profiles of the constructed stimuli and the external reference corpus. Values are means.

Measure	AI		Expert		Online Community	
	Constructed	Reference	Constructed	Reference	Constructed	Reference
Categorical–Dynamic Index	27.36	27.31	21.30	22.92	12.03	12.05
Formality	.824	.530	.801	.733	.521	.572
Coleman–Liau grade level	13.67	12.04	13.39	11.45	7.74	7.11
First-person singular / 100 words	0.00	0.14	0.62	0.59	3.66	1.97
Contractions / 100 words	0.00	1.15	1.59	1.43	3.58	2.55

Note. Constructed values summarize eight stimuli per advice style; reference values summarize 24 topic-matched texts per source category. Higher CDI values indicate more categorical/analytical language, whereas lower values indicate more dynamic/interpersonal language. The main text reports the inferential validation tests and the dimensions that did not reproduce the same ordering across constructed and reference texts.

Table A4. Advice-evaluation items and corresponding analytic constructs.

No.	Measurement Item (7-point Likert scale)	Item Construct	Analytic Construct
Appraisals			
1	The advice is easy to read and follow.	Readability	Message Appraisal
2	The advice provides clear reasoning for its recommendation.	Reasoning Clarity	Message Appraisal
3	The advice fits well with X's situation.	Situational Fit	Message Appraisal
4	The advice acknowledges potential risks and uncertainties.	Risk Acknowledgment	Risk Acknowledgment
5	Following this advice could put X at financial risk.	Financial Harm Risk	Safety Concern†
6	Someone with limited financial knowledge could misunderstand this advice.	Misleadingness	Safety Concern†
7	The source of this advice appears knowledgeable about financial decisions.	Source Knowledge	Source Knowledge
Downstream Judgments			
8	Overall, this is high-quality financial advice for X.	Overall Quality	Overall Quality
9	I would trust this advice if I were in X's situation.	Trust	Trust
10	If I were in X's situation, I would feel comfortable following this advice.	Intended Reliance	Intended Reliance

Note. X denotes the protagonist named in each financial scenario. Readability, Reasoning Clarity, and Situational Fit were combined to form *Message Appraisal*; Financial Harm Risk and Misleadingness were combined to form *Safety Concern*. *Risk Acknowledgment* and *Source Knowledge* were retained as single-item appraisals. Items marked with a dagger (†) are negatively valenced, such that higher ratings indicate greater safety concern.

B Supplementary Statistical Results

B.1 Advice Style and Source Labels

Table A5. Likelihood-ratio tests comparing additive mixed-effects models with models including the Advice Style × Source Label interaction.

Outcome	$\chi^2(6)$	<i>p</i>
Message Appraisal	3.71	.716
Risk Acknowledgment	7.32	.293
Safety Concern	2.20	.901
Source Knowledge	3.29	.772
Overall Quality	4.83	.566
Trust	1.48	.961
Reliance	2.17	.904

Table A6. Bootstrap decomposition of appraisal-carried pathways from advice style and source labels to intended Reliance. Bolded estimates have 95% bootstrap confidence intervals excluding zero.

Upstream contrast	Message Appraisal	Risk Ack.	Safety Concern	Source Knowledge	Total
<i>Advice style</i>					
Expert vs. AI	.114 [.069,.166]	-.003 [-.009,.002]	.043 [.023,.068]	.055 [.009,.106]	.210 [.115,.309]
OC vs. AI	.076 [.033,.123]	-.006 [-.017,.001]	.022 [.003,.045]	-.068 [-.129,-.014]	.023 [-.083,.129]
<i>Source labels</i>					
AI vs. Unlabeled	.025 [-.033,.090]	-.003 [-.014,.004]	.024 [-.009,.061]	.018 [-.053,.095]	.064 [-.072,.209]
Expert vs. Unlabeled	.064 [.001,.129]	-.001 [-.011,.008]	.028 [-.006,.066]	.088 [.014,.164]	.179 [.032,.325]
OC vs. Unlabeled	.058 [.001,.118]	-.002 [-.011,.006]	.020 [-.015,.056]	.038 [-.036,.114]	.114 [-.022,.256]

Note. Cells report bootstrap mean pathway estimates with 95% percentile confidence intervals from 2,000 participant-level bootstrap samples. AI is the advice-style reference; Unlabeled is the source-label reference. These estimates are descriptive pathway decompositions rather than causal mediation effects.

Table A7. Bootstrap contrasts comparing the Message Appraisal and Source Knowledge pathways to intended Reliance.

Pathway contrast	Estimate	95% CI
<i>Expert</i>		
Expert style: Message Appraisal – Source Knowledge	.060	[.006,.112]
Expert label: Message Appraisal – Source Knowledge	-.021	[-.081,.035]
Style–label contrast-of-contrasts	.081	[.014,.148]
<i>Online Community</i>		
OC style: Message Appraisal – Source Knowledge	.144	[.101,.190]
OC label: Message Appraisal – Source Knowledge	.020	[-.043,.082]
Style–label contrast-of-contrasts	.124	[.045,.212]

Note. Positive values indicate a relatively stronger Message Appraisal pathway than Source Knowledge pathway. Bolded intervals exclude zero.

B.2 Boundary Conditions

Table A8. Supplementary tests of boundary conditions for Trust and intended Reliance.

Boundary condition	Tested association	Statistic	df	p_{FDR}
<i>Decision-context moderation of Trust</i>				
Stakes	Context $\times$ evaluation-pathway block	$\chi^2 = 11.52$	5	.104
External Uncertainty	Context $\times$ evaluation-pathway block	$\chi^2 = 10.23$	5	.104
Verifiability	Context $\times$ evaluation-pathway block	$\chi^2 = 5.94$	5	.312
<i>Decision-context moderation of intended Reliance</i>				
Stakes	Context $\times$ evaluation-pathway block	$\chi^2 = 5.22$	6	.737
External Uncertainty	Context $\times$ evaluation-pathway block	$\chi^2 = 13.17$	6	.121
Verifiability	Context $\times$ evaluation-pathway block	$\chi^2 = 3.55$	6	.737
<i>Selected participant-level associations and moderation</i>				
General AI Trust	Main association with Trust (upstream)	$\beta = .202$	–	< .001
General AI Trust	Main association with Trust (pathway-adjusted)	$\beta = .046$	–	.109
General AI Trust	Main association with Reliance (upstream)	$\beta = .227$	–	< .001
General AI Trust	Main association with Reliance (pathway-adjusted)	$\beta = .053$	–	.009
Human Trust Propensity	Safety Concern $\times$ moderator $\rightarrow$ Trust	$\beta_{int} = .056$	–	.002
Objective Financial Literacy	Source Knowledge $\times$ moderator $\rightarrow$ Trust	$\beta_{int} = .054$	–	.017
Objective Financial Literacy	Trust $\times$ moderator $\rightarrow$ Reliance	$\beta_{int} = .039$	–	.002

Note. Decision-context entries are likelihood-ratio tests comparing the additive model with a joint interaction model. For Trust, the interaction block included Message Appraisal, Risk Acknowledgment, Safety Concern, Source Knowledge, and Overall Quality; for Reliance, Trust was additionally included. Benjamini–Hochberg correction was applied within the prespecified test families. Participant characteristics were standardized in moderation models.

C Robustness Analyses

We conducted four sensitivity analyses focused on the central appraisal–Quality–Trust–Reliance pathway.

Within- and between-participant variation. We decomposed each time-varying predictor into a participant mean and a within-participant deviation and re-estimated the Quality, Trust, and Reliance models using both components. The principal associations remained at the within-participant level. Message Appraisal ($\beta_W = .329$), Risk Acknowledgment ($\beta_W = .073$), Safety Concern ($\beta_W = -.106$), and Source Knowledge ($\beta_W = .444$) retained their expected associations with Overall Quality (all $p \leq .001$). Overall Quality remained strongly associated with Trust ($\beta_W = .573$, $p < .001$), and Trust remained the dominant within-participant correlate of Reliance ($\beta_W = .533$, $p < .001$).

Alternative repeated-measures estimation. Generalized estimating equations with an exchangeable working correlation closely reproduced the mixed-effects estimates. The Overall Quality–Trust association was $\beta = .608$ under GEE (vs. .607 in the primary model), and the Trust–Reliance association was $\beta = .595$ (vs. .594). The only notable difference was the small Source Knowledge–Reliance association, which remained similar in magnitude ($\beta = .048$) but was not statistically distinguishable from zero ($p = .084$).

Scenario fixed effects. Replacing Stakes, External Uncertainty, and Verifiability with fixed effects for the eight scenarios produced nearly identical estimates. Message Appraisal predicted Overall Quality at $\beta = .320$ (vs. .325), Source Knowledge at $\beta = .491$ (vs. .489), Overall Quality predicted Trust at $\beta = .603$ (vs. .607), and Trust predicted Reliance at $\beta = .593$ (vs. .594).

Constituent-item sensitivity. Replacing Message Appraisal and Safety Concern with their constituent items preserved the conclusions. The three message-appraisal items were jointly associated with Overall Quality ($\chi^2(3) = 211.04$, $p < .001$), Trust ($\chi^2(3) = 35.79$, $p < .001$), and Reliance ($\chi^2(3) = 16.67$, $p < .001$). The two safety-concern items were likewise jointly associated with Overall Quality ($\chi^2(2) = 23.70$, $p < .001$), Trust ($\chi^2(2) = 56.55$, $p < .001$), and Reliance ($\chi^2(2) = 12.45$, $p = .002$). The central downstream estimates also remained nearly unchanged: Overall Quality predicted Trust at $\beta = .598$ (vs. $.607$), while Overall Quality and Trust predicted Reliance at $\beta = .208$ and $\beta = .594$ (vs. $.206$ and $.594$, respectively).