



Article

# Perceiving AI as an Epistemic Authority or *Algority*: A User Study on the Human Attribution of Authority to AI

Frida Milella <sup>1,\*</sup> and Federico Cabitza <sup>1,2</sup>

<sup>1</sup> Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Viale Sarca 336, 20126 Milan, Italy; federico.cabitza@unimib.it

<sup>2</sup> Digital Health and Wellbeing Center, Fondazione Bruno Kessler (FBK), Via Sommarive 18, 38123 Trento, Italy

\* Correspondence: frida.milella@unimib.it

## Abstract

The increasing integration of artificial intelligence (AI) in decision-making processes has amplified discussions surrounding algorithmic authority—the perceived epistemic legitimacy of AI systems over human judgment. This study investigates how individuals attribute epistemic authority to AI, focusing on psychological, contextual, and sociotechnical factors. Existing research highlights the importance of trust in automation, perceived performance, and moral frameworks in shaping such attributions. Unlike prior conceptual or philosophical accounts of algorithmic authority, our study adopts a relational and empirically grounded perspective by operationalizing *algority* through psychometric measures and contextual assessments. To address knowledge gaps in the micro-level dynamics of this phenomenon, we conducted an empirical study using psychometric tools and scenario-based assessments. Here, we report key findings from a survey of 610 participants, revealing significant correlations between trust in automation (TiA), perceptions of automated performance (PAS), and the propensity to defer to AI, particularly in high-stakes scenarios like criminal justice and job-matching. Trust in automation emerged as a primary factor, while moral attitudes moderated deference in ethically sensitive contexts. Our findings highlight the practical relevance of transparency and explainability for supporting critical engagement with AI outputs and for informing the design of contextually appropriate decision support. This study contributes to understanding algorithmic authority as a multi-dimensional construct, offering empirically grounded insights for designing AI systems that are trustworthy and context-sensitive.

**Keywords:** algorithmic authority; human–AI interaction; decision-making; epistemic authority; *algority*



Academic Editors: Da Tao, Tingru Zhang and Hailiang Wang

Received: 23 December 2025

Revised: 24 January 2026

Accepted: 30 January 2026

Published: 5 February 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

In an episode of the TV show *Little Britain USA*, a hospital admissions clerk proposes a joint replacement surgery to a child who has come for a routine procedure; when the mother protests, the attendant replies with stolid certainty: “the computer says no.” The phrase has since become a shorthand for situations in which institutional decisions are justified by reference to computational outputs rather than to accountable human judgment, and it has reappeared in recent scholarly discussions of automated decision-making (e.g., [1–5]). We refer to this composite phenomenon as *algorithmic authority*: the influence of computational systems within human practices characterized by expert discretion, interpretation, and decision making.

A large part of the literature treats algorithmic authority as a sociotechnical construct whose legitimacy is produced through the interaction of technical artifacts, institutions, and users. In this paper we adopt an explicitly relational and empirically oriented stance and focus on *epistemic* authority, that is, the attribution of epistemic legitimacy to an AI system's outputs as convincing assertions about what is true and, consequently, what it is better to do. We call *algority* the propensity to confer such authority on algorithms in contexts where one might otherwise defer to human experts.

The empirical contribution of this study is to capture primary dimensions of algority through a psychometric and scenario-based survey instrument and to test how individual predispositions are associated with different, practically relevant expressions of deference to AI. Concretely, we combine (i) items drawn from established scales measuring trust in automation and beliefs about automation performance, as well as authority-relevant moral attitudes, with (ii) ad hoc items elicited through ordinary and high-stakes scenarios (e.g., navigation versus morally and legally sensitive judgments). We then examine associative patterns between these measures using correlation analysis and non-parametric group comparisons. In line with the exploratory design of the study, these analyses are intended as associative evidence (rather than causal inference) and do not constitute a full psychometric validation of a novel standalone scale.

To strengthen alignment between the theoretical framing and the subsequent analysis, we articulate the following research questions, formulated to match the constructs and statistical tests used in the study:

- **RQ1 (predispositions and algority).** How are literature-derived predispositions—specifically trust in automation and beliefs about automation performance—associated with respondents' algority-related responses across scenarios (e.g., expectation that AI makes fewer errors; acceptance of replacement; preference for AI judgment; deference to AI recommendations)?
- **RQ2 (moral-attitudinal correlates in sensitive contexts).** How are authority-relevant moral attitudes associated with algority-related responses in ethically sensitive scenarios (e.g., criminal judgment), and does this association differ from low-stakes or everyday contexts?
- **RQ3 (scenario dependence).** To what extent do the observed associations between predispositions (trust in automation; beliefs about automation performance; authority-relevant moral attitudes) and algority-related responses differ between low-stakes scenarios (e.g., route planning) and high-stakes scenarios (e.g., criminal justice, job interviewing, creditworthiness)?

The remainder of the paper proceeds as follows: Section 2 situates our approach within the broader conceptual literature on algorithmic authority and clarifies key definitional and theoretical boundaries; we then describe the study design and measures in Section 3, while the analysis strategy and the empirical results are discussed in Section 4. Section 5 summarizes the important findings and their significance, while Section 6 addresses the study's limits and suggests potential directions for future research. Section 7 presents a series of concluding remarks.

## 2. Related Work: Definitions, Boundaries, and Positions

### 2.1. From Algorithmic Authority to Algority

The expression *algorithmic authority* is commonly traced to Shirky's observation that, when seeking reputable information sources, people increasingly treat aggregators and filters as authoritative [6,7]. A core aspect of this shift is that epistemic reliance is conferred not on an identifiable individual but on a process: an unmanaged computational procedure

that derives value from heterogeneous sources [6–8]. Lustig and Nardi further articulated the pragmatic dimension of the construct by emphasizing that algorithmic authority becomes salient when it (i) directs human action and (ii) is *differentially trusted*, i.e., preferred over human authority in relevant circumstances [9]. Later refinements describe algorithmic authority as the power of algorithms to manage human action and influence what information is accessible to users, stressing that it does not reside solely in code but emerges from a diversity of sociotechnical actors [10]. In this paper, we adopt *algority* to name the individual-level propensity to confer authority on algorithmic outputs, especially in contexts where expert judgment and discretion are salient. (Hence, we prefer *algority* to *algocracy* [11], unless the latter is used to indicate governance *through* algorithms rather than governance *by* algorithms.)

## 2.2. Distinguishing Human and Algorithmic Authority

A recurring concern in this literature is whether users can recognize and differentiate between human and algorithmic authority, and what follows when they cannot—particularly as decision support systems become more prevalent and more tightly integrated into professional and institutional decisions [12]. This motivates empirical inquiry into the conditions under which epistemic legitimacy is conferred on algorithmic outputs, and into the micro-level correlates of deference, resistance, and preference for hybrid arrangements.

## 2.3. Agency, Mediation, and Scope: The Non-Agentiality Stance

Some accounts interpret algorithmic authority through the lens of mediation and distributed agency, noting that algorithmic outputs are embedded in sociotechnical arrangements in which individuals' judgment may remain a necessary supplement to algorithmic authority [10]. While such framings are theoretically suggestive, they also invite strong agentive language (e.g., “algorithms using humans”). In this study, we adopt a more cautious scope condition: we treat *algority* as a relational phenomenon grounded in human conferral of legitimacy rather than as an instance of algorithmic agency, a boundary we have previously discussed as a *non-agentiality* stance [13]. This positioning supports an operationalization that focuses on measurable attributions, expectations, and decision preferences without presupposing goals or intentions on the part of algorithms.

## 2.4. Macro-Level Accounts and Micro-Level Mechanisms

Algorithmic authority is also studied through sociological lenses that emphasize how computational systems reorder social relations, institutional practices, and epistemic arrangements [14,15]. Complementary critical work highlights structural dynamics such as the reinforcement of social inequalities and the propagation of bias [15–18]. Building on these perspectives, our contribution focuses on a complementary level of analysis: the micro-level correlates of epistemic authority attribution to AI advice and a psychometric operationalization that enables systematic empirical investigation across scenarios and domains. Table 1 summarizes the main theoretical concepts adopted in this study.

**Table 1.** Summary of the main theoretical concepts adopted in this study.

Concept	Definition
Algorithmic authority	The power of algorithms to manage human action and influence what information is accessible to users, stressing that it does not reside solely in code but emerges from a diversity of sociotechnical actors [10].

Table 1. Cont.

Concept	Definition
Epistemic Authority	The attribution of epistemic legitimacy to an AI system's outputs as convincing assertions about what is true and, consequently, what it is better to do.
Algority	The propensity to confer such authority on algorithms in contexts where one might otherwise defer to human experts. We treat algority as a relational phenomenon grounded in human conferral of legitimacy rather than as an instance of algorithmic agency, a boundary we have previously discussed as a non-agentiality stance [13].
Epistemic reliance	Epistemic reliance is conferred not on an identifiable individual but on a process: an unmanaged computational procedure that derives value from heterogeneous sources [6–8].

### 3. Methods

#### 3.1. User Study

In this study, we focused on identifying the main factors making participants confident in considering AI advisors as authorities, thereby providing preliminary insights into the propensity to attribute epistemic authority to AI advisors, that is, *algority*. To explore this, a psychometric tool was developed to evaluate attitudes and preferences regarding algorithmic authority compared to human authority. This allowed us to test factors affecting algority by measuring them through a user study and then examining their association using Spearman's rank correlation coefficients and non-parametric tests, such as the Mann-Whitney U test (MW) and the Kruskal–Wallis test (KW). We used a stepwise methodological approach to confirm associative hypotheses between predisposing factors and potential enablers for recognizing epistemic authority in AI. This aligns with the exploratory nature of this study.

We designed the psychometric tool to incorporate both validated items from existing literature and ad hoc items into an online questionnaire, as described below (see Section 3.2). We initially asked respondents to complete a section of the questionnaire based on items taken from the existing literature. We administered items from three scales that have been suggested in the literature concerning authority and human interactions with technology, since we did not find any specific source focusing on determinants of algorithmic authority. Many factors or facilitating conditions of algorithmic authority likely overlap with traditional determinants of trust in automation, which has been extensively discussed in the literature. These determinants include perceived fairness [19], the reputation of the organisation behind the algorithms [20], transparency [21], and perceived efficiency and objectivity [22], as well as perceived effectiveness and decision accuracy [23]. In this study, we assumed relevant those capturing individual priorities in moral decision-making (i.e., the Moral Foundation Questionnaire—MFQ) [24], trust in automation (i.e., The Trust in Automation scale—TiA) [25], and perceptions regarding the performance of automated systems (i.e., the Perfect Automation Schema—PAS) [26].

Subsequently, we presented hypothetical scenarios to respondents in which the decisions and judgments of authorities may have legal or otherwise significant implications for subjects, aiming to investigate their attitudes and preferences towards automated decision-making. To do so, we designed ad hoc items that we grouped into 5 areas of concern based on the selected ordinary scenarios. A subset of the ad hoc items (i.e., Expectation on AI (EonAI) and Attitude towards replacement (ATR)—see Section 3.2) were grounded in the traditional epistemic relationship between laypeople and specialists, which remains the main instance of epistemic authority [27]. Experts, indeed, are typically epistemic

agents seen as epistemically superior to us concerning assertions within their areas of competence [27–29]. Thus, assessing the perceived importance of human experts in relevant and sensitive contexts facilitates an understanding of the degree to which users still place their trust in established human epistemic authorities. On the other hand, epistemic authority is not an inherent attribute but arises via contextual interactions with users, dependent on the user’s objectives, circumstances, and assessment criteria [30]. Consequently, the remaining ad hoc items (i.e., Trust on Prediction (ToP), Preference on Judgment (PoJ) and Deference for Action (DoA)—see Section 3.2) were designed to specifically assess users’ perceptions of AI systems as reliable advisors in making decisions regarding personal and professional matters. Recent works, indeed, are increasingly distinguishing between the dimensions of trust and epistemic authority, positing that trust is not a prerequisite for epistemic authority, but rather a concurrent perspective through which the interaction between humans and AI is framed and negotiated [30]. In this context, trust and authority are separate yet interrelated aspects of human–AI interaction, influenced by how users formulate their expectations in particular tasks and domains [30]. Therefore, the development of an independent scale for measuring algorithmic authority, focusing on its situational nature and the epistemic asymmetry between traditional experts and non-specialists, contributes in bridging the gap in recognizing epistemic authority [27] as a standalone construct. In doing so, we screened the literature on epistemic authority to find paradigmatic instances of alority. We were inspired by some instances, such as expectations on the errors in automated aids [26]; attitude towards human judgment replacements, with experts asserting themselves as epistemic authorities in their respective domains of expertise [27]; and algorithmic decisions in assessment domains not always being error- or bias-free (e.g., mortgage loans and job recruitment) [31]. We also opted to include instances of low-risk domains and sensitive scenarios to assess individuals’ attitudes regarding AI authority in predictive decision-making. Our ad hoc items were conceived to cover three decision-making approaches, that is, responders exhibiting a preference for human authority, those showing preference or deference towards AI as decision-makers, and responders favoring human–AI collaboration in the decision-making process.

Since the items taken from the literature were not available in Italian, this part of the questionnaire was translated from English to Italian by the authors independently. The discrepancies between the item translations were analyzed until a consensus was achieved. We directly conceived the ad hoc items in Italian, so no translation or back-translation was necessary to preserve the original meaning (e.g., [32]). The complete questionnaire was then administered via the online platform Lime Survey to students enrolled in the Computer Science courses at the University of Milano-Bicocca (Milano, Italy). Data collection was carried out at two time-points, from November 2022 to May 2023 and subsequently from November 2023 to January 2024. A snowball sampling method was employed to expand the survey sample until a convenient sample was obtained to compute statistics (e.g., [33,34]). Data were analyzed using R software version 4.3. Section 3.2 describes the measures employed in the questionnaire.

## 3.2. Measures

### 3.2.1. Literature-Derived Constructs

(1) The *Moral Foundation Questionnaire* (MFQ) [35,36], which comprises 32 items, ranges from perception about a set of issues in moral decision-making to agreement in regard to a variety of moral propositions, such as ‘All children must learn respect for authority’. In this study, we chose three items of the former kind above for which we proposed a 5-point differential semantic ranging from 1 (=completely irrelevant) to 5 (=completely relevant), and three items of the latter type, for which we adopted a 6-item differential semantic ranging from 1 (=strongly disagree) to 6 (=strongly agree). As an aggregate, these items express

some form of attitude towards authority. (2) The *Trust in Automation scale* (TiA) [25,37] consists of six sub-scales (namely, Reliability/Competence, Understanding/Predictability, Propensity to Trust, Intention of Developers, Familiarity, and Trust in Automation) with 2 to 4 items each, for a total of 19 items. For the purposes of our study, we chose only the five items that belonged to the sub-scales ‘Propensity to Trust’ and ‘Trust in automation’; for both sub-scales, the adopted response format was a 6-point rating scale ranging from 1 (=strongly disagree) to 6 (=strongly agree). These items, aggregated, express the participants’ trust in automation. Item 1 was assigned scores on a reverse scale. (3) A larger pool of items, i.e., 9 items, was taken from the *Perfect Automation Schema* (PAS) scale proposed by [26], for which we used the same 6-point scale mentioned above. Six items were aimed at assessing high expectations (e.g., ‘Automated systems have 100% perfect performance’), whilst three items were used to assess All-or-None thinking (e.g., ‘If an automated system makes a mistake, then it is completely useless’). These items grouped should express the uncompromising trust in the unquestionable authority of AI. Items 4 and 5 were assigned scores on a reverse scale. Table 2 presents the measures employed in the questionnaire.

**Table 2.** Questionnaire: items taken from the literature, namely, the Moral Foundation Questionnaire—MFQ, The Trust in Automation scale—TiA, and the Perfect Automation Schema—PAS. \* Item 1 from TiA scale and items 4 and 5 from PAS scale were on a reversed scale.

<b>Moral Foundation Questionnaire (MFQ)</b>
(1) Whether or not someone showed a lack of respect for authority
(2) Whether or not someone conformed to the traditions of society
(3) Whether or not an action caused chaos or disorder
(4) Respect for authority is something all children need to learn
(5) Men and women each have different roles to play in society
(6) If I were a soldier and disagreed with my commanding officer’s orders, I would obey anyway because that is my duty
<b>Trust in Automation (TiA)</b>
(1) One should be careful with unfamiliar automated systems *
(2) I rather trust a system than I mistrust it
(3) Automated systems generally work well
(4) I trust the system
(5) I can rely on the system
<b>Perfect Automation Schema (PAS)</b>
(1) Automated systems have 100% perfect performance
(2) Automated systems rarely make mistakes
(3) Automated systems can always be counted on to make accurate decisions
(4) Automated systems make more mistakes than people realize *
(5) People generally believe automated system work better than they do *
(6) People have no reason to question the decisions automated systems make
(7) If automated system makes error, then it is broken
(8) If an automated system makes a mistake, then it is completely useless
(9) Correctly functioning automated systems are perfectly reliable

### 3.2.2. Ad Hoc Constructs

(4) *Expectation on AI* comprises a simple item associated with the extent to which respondents believe AI is a reliable epistemic authority in delicate domains (EonAI a). (5) *Attitude towards replacement* groups two ad hoc items associated with the extent to which respondents believe AI could replace human authorities (ATR b) and experts in general (ATR a). (6) *Trust on Prediction* groups two ad hoc items associated with the extent to which respondents tend to consider algorithmic predictions about personally relevant future events (on love and job matters) reliable; to this aim, we selected two ordinary scenarios, like job seeking (ToP a) and date matching (ToP b) for the personal, as well as

commercial, implications of these two applications domains. (7) *Preference on Judgment* groups three ad hoc items associated with the extent to which respondents tend to prefer that decisions that can strongly impact their lives are made by AI rather than other humans (individually or acting in some form of collective); the scenarios chosen for these items are job interviewing (PoJ b), criminal judgment (PoJ a), and creditworthiness assessment (PoJ c). (8) *Deference for Action* comprises a simple item taken as paradigmatic of those cases in which AI can influence our action in daily, ordinary, and low-stakes situations; to this aim, we let respondents imagine a situation where the recommendations of a navigation system are compared with the route suggested by a friend of theirs (DoA a). Tables 3 and 4 show the ad hoc constructs chosen in this study and their answer options.

**Table 3.** Questionnaire: ad hoc items. Expectation on AI and Attitude towards replacement.

<b>Expectation on AI (EonAI)</b>
(a) In these areas, compared with renowned experts, an AI software or machine usually makes: (1) Far fewer errors (2) Slightly fewer errors (3) About the same number of errors (4) Slightly more errors (5) Many more errors (6) Can't make these decisions yet
<b>Attitude towards replacement (ATR)</b>
(a) If, in a given high-risk task, an AI system consistently shows higher accuracy than an average human expert: (1) It should support humans, but they must always have the last word and can disagree (2) It should support humans, who always have the last word and can disagree only if they bring solid evidence that they are right and the machine is wrong (3) It should replace humans in these tasks (4) This can never happen
(b) Nobel laureate Kahnemann once said that: "Whenever we can replace human judgment with an algorithm, we should at least consider it. [In fact] we should replace humans with algorithms whenever it is possible to: (1) Strongly disagree (2) Moderately disagree (3) Slightly disagree (4) Slightly agree (5) Moderately agree (6) Strongly agree

**Table 4.** Questionnaire: ad hoc items. Trust on Prediction, Preference on Judgment, and Deference for Action.

<b>Trust on Prediction (ToP)</b>
(a) If a dating app calculated a 98.5 percent affinity between you and a potential partner, how likely do you think you would be to have a long and rewarding love affair with that person: (1) Almost certainly (2) 50% (3) About 1 in 20 (5%) (4) Hardly possible
(b) If a computer program considered your exam grades, and your answers to a long psychometric questionnaire, and offered you a job position with a 97.8 percent probability that it would be the perfect job for you, what is the likelihood that you might actually feel fulfilled in doing that job? (1) Almost certainly (2) 50% (3) About 1 in 20 (5%) (4) Hardly possible

**Table 4.** *Cont.*

<b>Preference on Judgment (PoJ)</b>
<p>(a) You are falsely accused of a crime you did not commit, and you are assisted by a very good lawyer. You would rather be tried by:</p> <ol style="list-style-type: none"> <li>(1) A judge</li> <li>(2) An AI that calculates the probability that you are actually innocent or guilty</li> <li>(3) A popular jury</li> <li>(4) A popular jury including an AI that calculates the probability that you are actually innocent or guilty</li> <li>(5) A jury of judges</li> <li>(6) A jury of judges using an AI that calculates the probability that you are actually innocent or guilty</li> </ol>
<p>(b) For a very important job interview for a position you believe you deserve and that would change your life for the better, you would rather be judged by:</p> <ol style="list-style-type: none"> <li>(1) A human being</li> <li>(2) An artificial intelligence</li> <li>(3) A human supported by an artificial intelligence</li> <li>(4) A human committee</li> <li>(5) A human commission supported by an AI</li> </ol>
<p>(c) If you were to apply for a mortgage, and for that purpose you still had to fill out a lengthy questionnaire, knowing that the decision is final even though it is your right to know the reasons:</p> <ol style="list-style-type: none"> <li>(1) A human expert</li> <li>(2) An AI</li> <li>(3) A committee</li> <li>(4) A committee of which an AI is also a member</li> </ol>
<b>Deference for Action (DoA)</b>
<p>(a) Imagine the following situation: you have to drive to a friend who lives in an area you do not know. Getting into the car, you turn on the navigation system and enter your friend's correct address:</p> <ol style="list-style-type: none"> <li>(1) turn right and follow your friend's directions because... she will know exactly how to get to her house!</li> <li>(2) you turn left and follow the navigator's directions because it has just been updated and may want you to avoid a busy area or a temporarily closed road.</li> <li>(3) You are confused, you make the traffic circle a couple of times trying to figure out what to do but eventually you listen to the navigator who will still take you to your destination</li> <li>(4) You're confused, you go around the traffic circle a couple of times trying to figure out what to do but in the end you do what the friend told you trusting that the navigator will recalculate the route in a few seconds...</li> </ol>

## 4. Results

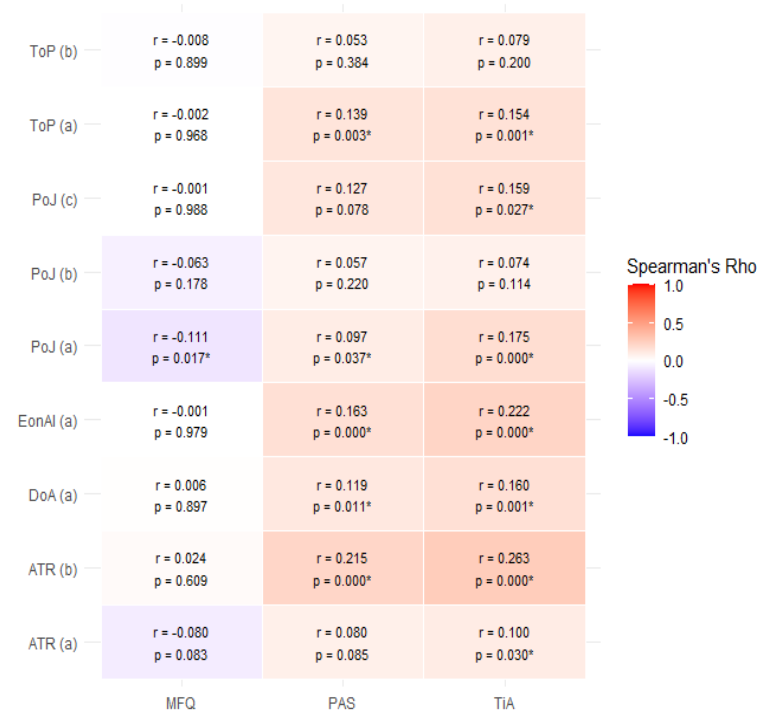
Six hundred and ten people answered the full questionnaire. The sample consisted of 50% males, 49% females, and 1% gender non-confirming. Over 60% of the sample was under the age of 26, and 39% declared not being students.

We first assessed the questionnaire's internal consistency using Cronbach's alpha to confirm its reliability against the recommended threshold of 0.7 [38]. The study found that TiA and PAS Cronbach's alpha values were higher than the suggested level of 0.70, indicating a moderate to high level of reliability (i.e., TiA Chronbach's alpha = 0.8; PAS Chronbach's alpha = 0.756), whereas MFQ Cronbach's alpha value was only 0.638. However, the range 0.6–0.7 is considered an acceptable level of reliability in the literature [39], so we also included the MFQ for the subsequent analysis. None of the remaining items were signalled as redundant, being below the critical upper bound of 0.95 [40], even though only the preference on judgment Cronbach's alpha value was higher than the critical threshold of 0.7 (i.e., preference on judgment Cronbach's alpha = 0.753). In this study, we therefore assumed priorities in moral decision-making, trust in automation, and the perceptions regarding the performance of automated systems as the predisposing factors of participants'

confidence or resistance to algorithmic decisions. We accordingly examined the correlation strength between each literature-derived construct (MFQ, TiA, and PAS) and each of the ad hoc items using Spearman’s rank correlation coefficients, as the Shapiro–Wilk normality tests for the data were unsuccessful [41].

4.1. Correlations

The results showed positive significant correlations between trust in automation (i.e., TiA) and the majority of the ad hoc items, such as Expectation on AI ( $r = 0.22, p < 0.05$ ), date matching ( $r = 0.15, p < 0.05$ ), criminal judgment ( $r = 0.18, p < 0.05$ ), creditworthiness assessment ( $r = 0.16, p < 0.05$ ), attitude towards expert replacement ( $r = 0.26, p < 0.05$ ), and Deference for Action ( $r = 0.16, p < 0.05$ ). Similarly, our findings revealed positive significant correlations between perceptions regarding the performance of automated systems (i.e., PAS) and Expectation on AI ( $r = 0.16, p < 0.05$ ), date matching ( $r = 0.14, p < 0.05$ ), attitude towards expert replacement ( $r = 0.22, p < 0.05$ ), and Deference for Action ( $r = 0.12, p < 0.05$ ). The findings also showed a negative significant correlation between priorities in moral decision-making (i.e., MFQ) and criminal judgment ( $r = -0.11, p < 0.05$ ), whereas a positive non-significant correlation was found between perceptions regarding the performance of automated systems (i.e., PAS) and creditworthiness assessment ( $r = 0.13, p > 0.05$ ) (Figure 1). The analysis showed that trust in automation and a strong belief in AI’s unquestionable authority are both linked to the willingness to let AI make decisions in different real-life situations, whereas the attitude toward authority exhibits a negative correlation with the inclination to delegate human decision-making to AI, though this association was statistically significant in only one real scenario. So, we looked into these initial associations, examining whether there were any statistically significant differences in the hypothesized dimensions of algorithmic authority based on how respondents felt about authority, how much they trust automation, and how strong is their faith in automation perfection.



**Figure 1.** Spearman’s rank correlation coefficients between literature-derived constructs (MFQ, TiA, and PAS) and ad hoc constructs. EonAI = Expectation on AI. ATR = Attitude towards replacement. ToP = Trust on Prediction. PoJ = Preference on Judgment. DoA = Deference on Action. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. \*  $p$  value  $< 0.05$

#### 4.2. Non-Parametric Tests

Mann–Whitney (MW) U tests [42] were conducted on two groups to determine if statistically significant differences exist across the groups of respondents depending on their high versus low attitude towards authority, high versus low trust in automation, and high versus low faith in automated system performance.

##### 4.2.1. Trust in Automaton Scale (TiA)

By categorizing respondents based on their levels of trust in automation, we found a statistically significant difference across the groups towards AI as reliable epistemic authorities in delicate domains ( $W = 33,622.50$ ; effect size  $r = 0.19$ ;  $Z = 4.39$ ;  $p$ -value  $< 0.001$ ), with a significant difference in those with medium-to-high levels of trust in automation ( $W = 33,622.50$ ; effect size  $r = 0.19$ ;  $Z = 4.39$ ;  $p$ -value  $< 0.001$ ) compared to the ones having medium-to-low trust in automation ( $W = 33,622.50$ ; effect size  $r = 0.19$ ;  $Z = 4.39$ ;  $p$ -value  $> 0.05$ ). Similarly, a statistically significant difference was observed between the groups regarding their attitudes towards replacing human authority ( $W = 32,402$ ; effect size  $r = 0.19$ ;  $Z = 4.43$ ;  $p$ -value  $< 0.001$ ), with the one-tailed test confirming this significant difference towards individuals with medium-to-high trust in automation ( $W = 32,402$ ; effect size  $r = 0.19$ ;  $Z = 4.43$ ;  $p$ -value  $< 0.001$ ). On the other hand, the groups generally showed no statistically significant difference in their feelings about replacing expert authority ( $W = 29,522.5$ ; effect size  $r = 0.07$ ;  $Z = 1.70$ ;  $p$ -value  $= 0.09$ ), although those with medium-to-high trust in automation were more likely to notice this difference ( $W = 29,522.5$ ; effect size  $r = 0.07$ ;  $Z = 1.70$ ;  $p$ -value  $= 0.045$ ). In terms of the effect sizes ( $r$ ), the results observed indicated that the disparity between the groups in terms of their attitude towards the replacement of human authority, both in general and in sensitive areas, exists in relation to the different levels of perceived trustworthiness of AI systems, but this difference, although statistically significant, is not substantial enough to be distinctly observable in practical applications. Regardless of the expressed level of trust in automation, there was not a statistically significant difference among the groups in terms of how they consider algorithmic predictions of personal events, like job seeking (two-tailed test:  $W = 9618$ ; effect size  $r = 0.06$ ;  $Z = 1.39$ ;  $p$ -value  $= 0.164$ ; one-tailed test:  $W = 9618$ ; effect size  $r = 0.06$ ;  $Z = 1.39$ ;  $p$ -value  $= 0.082$ ), and in their attitude towards automated decisions concerning real-life scenarios, like creditworthiness assessment (two-tailed test:  $W = 5160$ ; effect size  $r = 0.07$ ;  $Z = 1.60$ ;  $p$ -value  $= 0.109$ ; one-tailed test:  $W = 5160$ ; effect size  $r = 0.07$ ;  $Z = 1.60$ ;  $p$ -value  $= 0.055$ ). Conversely, we found a statistically significant difference between groups concerning automated predictions about date matching ( $W = 30,410$ ; effect size  $r = 0.13$ ;  $Z = 3.07$ ;  $p$ -value  $= 0.002$ ) and the preference for judgments made by AI versus humans, such as criminal judgments ( $W = 30,300.50$ ; effect size  $r = 0.13$ ;  $Z = 3.13$ ;  $p$ -value  $= 0.002$ ) and job interviews ( $W = 29,169.50$ ; effect size  $r = 0.09$ ;  $Z = 2.012$ ;  $p$ -value  $= 0.034$ ), with the strongest difference in those who had medium-to-high trust in automation (data matching:  $W = 30,410$ ; effect size  $r = 0.13$ ;  $Z = 3.07$ ;  $p$ -value  $= 0.001$ ; criminal judgment:  $W = 30,300.50$ ; effect size  $r = 0.13$ ;  $Z = 3.13$ ;  $p$ -value  $< 0.001$  and job interviewing:  $W = 29,169.50$ ; effect size  $r = 0.09$ ;  $Z = 2.012$ ;  $p$ -value  $= 0.017$ ). Similarly, we detected a statistically significant difference in the respondents' willingness to receive AI support in their daily tasks, regardless of their degree of trust in automation ( $W = 29,135.5$ ; effect size  $r = 0.11$ ;  $Z = 2.62$ ;  $p$ -value  $= 0.009$ ). This difference was stronger for those who showed high-to-medium trust compared to those who had medium-to-low trust ( $W = 29,135.5$ ; effect size  $r = 0.11$ ;  $Z = 2.62$ ;  $p$ -value  $= 0.004$ ). Although the results indicate that there are statistically significant differences among groups with varying levels of trust regarding their perception of AI as a reliable source of advice for personal and professional decision-making, the magnitude

of the observed difference is too small to suggest a meaningful change in the decision to delegate decision-making to AI in practical operational contexts.

Overall, the results indicate that categorizing respondents according to their perceived trustworthiness of AI systems reveals statistically significant differences among groups in their propensity to defer epistemic authority to these systems. However, the systematically narrowed values of the effect sizes are unlikely to produce a practically perceivable change in practice. Table 5 shows the results.

**Table 5.** Mann–Whitney U tests grouped by TiA. W represents the value of the distribution test. r represents the effect size. EonAI = Expectation on AI. ATR = Attitude towards replacement. ToP = Trust on Prediction. PoJ = Preference on Judgment. DoA = Deference on Action. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. (1) Two-tailed test. (2) One-tailed test (“greater”). \*  $p$  value < 0.05.

	W	r	Z-Score	$p$ -Value (1)	$p$ -Value (2)
EonAI	33,622.50	0.19	4.39	<0.001 *	<0.001 *
ATR (a)	29,522.5	0.07	1.70	0.090	0.045 *
ATR (b)	32,402	0.19	4.43	<0.001 *	<0.001 *
ToP (b)	9618	0.06	1.39	0.164	0.082
PoJ (c)	5160	0.07	1.60	0.109	0.055
ToP (a)	30,410	0.13	3.07	0.002 *	0.001 *
PoJ (a)	30,300.50	0.13	3.13	0.002 *	<.001 *
PoJ (b)	29,169.50	0.09	2.12	0.034 *	0.017 *
DoA	29,135.5	0.11	2.62	0.009 *	0.004 *

#### 4.2.2. Perfect Automation Schema (PAS)

Using the faith in automated system performance to stratify the respondents, we noticed that a statistically significant difference exists between groups regardless of the level of faith in how they consider AI as reliable epistemic authorities in delicate domains ( $W = 30,439.5$ ; effect size  $r = 0.09$ ;  $Z = 2.00$ ;  $p$ -value = 0.046) as well as in the level of agreement about the statement on the replacement of AI with humans whenever possible ( $W = 31,282$ ; effect size  $r = 0.15$ ;  $Z = 3.46$ ;  $p$ -value < 0.001). However, using a one-tailed test, individuals with medium-to-high belief in AI’s unquestionable authority exhibited a significant preference towards AI as an epistemic authority in sensitive areas ( $W = 30,439.5$ ; effect size  $r = 0.09$ ;  $Z = 2.00$ ;  $p$ -value = 0.023), as well as in the substitution of human judgments with automated decisions ( $W = 31,282$ ; effect size  $r = 0.15$ ;  $Z = 3.46$ ;  $p$ -value < 0.001) relative to the other group. Regarding effect sizes ( $r$ ), our findings indicated that, while the observed difference between groups is statistically significant, it may not reflect a substantial divergence in their inclination to confer epistemic authority upon AIs in practice. Conversely, we observed no statistically significant difference in the propensity to substitute expert authority between the two groups ( $W = 27,625.50$ ; effect size  $r = 0.00$ ;  $Z = 0.08$ ;  $p$ -value = 0.939), nor was there any significant difference by secondly exploring the one-tailed test ( $W = 27,625.50$ ; effect size  $r = 0.00$ ;  $Z = 0.08$ ;  $p$ -value = 0.470). No statistically significant difference was found among groups regarding their perceptions of algorithmic predictions related to personal events, such as job seeking (two-tailed test:  $W = 9360$ ; effect size  $r = 0.04$ ;  $Z = 0.85$ ;  $p$ -value = 0.396; one-tailed test:  $W = 9360$ ; effect size  $r = 0.04$ ;  $Z = 0.85$ ;  $p$ -value = 0.198) nor on their attitude towards automated judgments in real-life settings, such as creditworthiness assessment (two-tailed test:  $W = 5255$ ; effect size  $r = 0.07$ ;  $Z = 1.60$ ;  $p$ -value = 0.109; one-tailed test:  $W = 5160$ ; effect size  $r = 0.07$ ;  $Z = 1.60$ ;

$p$ -value = 0.055). Additionally, no statistically significant difference was observed between the groups concerning the respondents’ intent to accept AI assistance in their everyday tasks (two-tailed test:  $W = 27,786.50$ ; effect size  $r = 0.06$ ;  $Z = 1.47$ ;  $p$ -value = 0.141; one-tailed test:  $W = 27,786.50$ ; effect size  $r = 0.06$ ;  $Z = 1.47$ ;  $p$ -value = 0.070) or their preference for judgments made by AI compared to humans, such as in job interviews (two-tailed test:  $W = 27,904$ ; effect size  $r = 0.04$ ;  $Z = 0.97$ ;  $p$ -value = 0.334; one-tailed test:  $W = 27,904$ ; effect size  $r = 0.04$ ;  $Z = 0.97$ ;  $p$ -value = 0.167). Conversely, the groups generally showed statistically significant differences concerning the automated predictions about date matching ( $W = 30,350.50$ ; effect size  $r = 0.12$ ;  $Z = 2.89$ ;  $p$ -value = 0.004), with those who had medium-to-high faith in automation performance more likely to notice this difference ( $W = 30,350.50$ ; effect size  $r = 0.12$ ;  $Z = 2.89$ ;  $p$ -value = 0.002). Similarly, in the group exhibiting medium-to-high faith in automation performance, the observed difference regarding the respondents’ preference for judgments made by AI compared to humans, such as in criminal cases, was statistically significant, thereby confirming a greater likelihood of identifying this directional effect in a one-tailed test (two-tailed test:  $W = 29,065$ ; effect size  $r = 0.08$ ;  $Z = 1.96$ ;  $p$ -value = 0.05; one-tailed test:  $W = 29,065$ ; effect size  $r = 0.08$ ;  $Z = 1.96$ ;  $p$ -value = 0.025). Nonetheless, the minimal and small-to-medium impact sizes suggest that both groups may exhibit comparable behavior in conferring epistemic status to AIs, regardless of their differing perceptions of faith in automated system performance.

Overall, the findings suggest that classifying respondents based on their confidence in automated system performance uncovers statistically significant variations in their propensity to delegate epistemic authority to these systems, although these differences are so slight that they may not result in behavior that is substantially different in practical terms. Table 6 shows the results.

**Table 6.** Mann–Whitney U tests grouped by PAS. W represents the value of the distribution test. r represents the effect size. EonAI = Expectation on AI. ATR = Attitude towards replacement. ToP = Trust on Prediction. PoJ = Preference on Judgment. DoA = Deference on Action. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. (1) Two-tailed test. (2) One-tailed test (“greater”). \*  $p$  value < 0.05.

	W	r	Z-Score	$p$ -Value (1)	$p$ -Value (2)
EonAI	30,439.5	0.09	2.00	0.046 *	0.023 *
ATR (a)	27,625.50	0.00	0.08	0.939	0.470
ATR (b)	31,282	0.15	3.46	<0.001 *	<0.001 *
ToP (b)	9360	0.04	0.85	0.396	0.198
PoJ (c)	5255	0.07	1.60	0.109	0.055
ToP (a)	30,350.50	0.12	2.89	0.004 *	0.002 *
PoJ (a)	29,065	0.08	1.96	0.050 *	0.025 *
PoJ (b)	27,904	0.04	0.97	0.334	0.167
DoA	27,786.50	0.06	1.47	0.141	0.070

#### 4.2.3. Moral Foundation Questionnaire (MFQ) and Gender

By categorizing the respondents based on their attitudes towards authority, we identified only a statistically significant difference between the groups regarding their preference for being evaluated by AI versus humans in morally relevant cases, such as criminal judgment, irrespective of their attitude levels towards authority (two-tailed test:  $W = 24224.50$ ; effect size  $r = 0.09$ ;  $Z = -2.14$ ;  $p$ -value = 0.033; one-tailed test:  $W = 24224.50$ ; effect size  $r = 0.09$ ;  $Z = -2.14$ ;  $p$ -value = 0.984). Similarly, using the gender variable to group the

respondents, we observed a significant difference in the respondents’ trust in automated predictions of ordinary scenarios, like job seeking, in the females compared to males (two-tailed test:  $W = 9890$ ; effect size  $r = 0.08$ ;  $Z = 1.92$ ;  $p$ -value = 0.055; one-tailed test:  $W = 9890$ ; effect size  $r = 0.08$ ;  $Z = 1.92$ ;  $p$ -value = 0.027), while in the Attitude towards replacement of human authority, the difference was significant regardless of the gender group (two-tailed test:  $W = 23,385.50$ ; effect size  $r = 0.11$ ;  $Z = -2.95$ ;  $p$ -value = 0.008; one-tailed test:  $W = 23,385.50$ ; effect size  $r = 0.11$ ;  $Z = -2.95$ ;  $p$ -value = 0.996). Therefore, regardless of whether the sample is stratified based on the respondents’ moral orientation towards authority or categorized by gender, the magnitude of the observed effect is fundamentally minimal, indicating that the disparities between the groups may not result in significantly varied behaviors in the attribution of epistemic superiority to AI systems in practical terms. Tables 7 and 8 show the results.

**Table 7.** Mann–Whitney U tests grouped by MFQ. W represents the value of the distribution test. r represents the effect size. EonAI = Expectation on AI. ATR = Attitude towards replacement. ToP = Trust on Prediction. PoJ = Preference on Judgment. DoA = Deference on Action. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. (1) Two-tailed test. (2) One-tailed test (“greater”). \*  $p$  value < 0.05.

	W	r	Z-Score	p-Value (1)	p-Value (2)
EonAI	27,799	0.00	0.05	0.960	0.480
ATR (a)	25,457.50	0.07	−1.69	0.090	0.955
ATR (b)	27,668	0.03	0.66	0.511	0.255
ToP (b)	8585.50	0.03	−0.71	0.478	0.762
PoJ (c)	4541.50	0.01	−0.33	0.746	0.628
ToP (a)	26,570.5	0.01	−0.18	0.859	0.571
PoJ (a)	24,224.50	0.09	−2.14	0.033 *	0.984
PoJ (b)	25,961	0.03	−0.67	0.502	0.749
DoA	27,064	0.04	0.83	0.408	0.204

**Table 8.** Mann–Whitney U tests grouped by gender. W represents the value of the distribution test. r represents the effect size. EonAI = Expectation on AI. ATR = Attitude towards replacement. ToP = Trust on Prediction. PoJ = Preference on Judgment. DoA = Deference on Action. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. (1) Two-tailed test. (2) One-tailed test (“greater”). \*  $p$  value < 0.05.

	W	r	Z-Score	p-Value (1)	p-Value (2)
EonAI	29,220	0.04	0.98	0.330	0.165
ATR (a)	28,889.5	0.04	0.91	0.364	0.184
ATR (b)	23,385.5	0.11	−2.65	0.008 *	0.996
ToP (b)	9890	0.08	1.92	0.055	0.027 *
PoJ (c)	4677.50	0.01	0.33	0.741	0.371
ToP (a)	27,295	0.01	0.30	0.763	0.381
PoJ (a)	27,111.50	0.01	0.17	0.869	0.434
PoJ (b)	26,604	0.01	−0.25	0.806	0.598
DoA	26,604	0.02	0.48	0.635	0.317

#### 4.2.4. Decision-Making Approaches

The Kruskal–Wallis (KW) tests [42] were performed to investigate statistically significant differences among three decision-making preferences: individuals favoring human authority, those preferring AI as a decision-maker, and those endorsing collaboration between humans and AI, in relation to the respondents' attitude towards authority, trust in automation, and confidence in automation's performance, respectively. The analysis focused on three real-life scenarios (job interviewing, criminal judgment, and credit-worthiness assessment), which were hypothesized to have an influence on respondents' preferences for AI-made judgments over human ones. The Kruskal–Wallis results revealed a statistically significant difference between the three groups of decision-making supports concerning trust automation in the criminal judgment scenario ( $\chi^2 = 14.142$ ,  $p = 0.0008$ ). The subsequent two-group Mann–Whitney tests, which were used to compare the decision-making supports, identified a significant difference between the group expressing a preference for relying on hybrid human–AI decision support and the group who deferred to human authority ( $p\text{-value}_{adj} = 0.0016$ ), albeit the effect size was small ( $\eta^2 = 0.0226$ ). This indicates that the perceived trustworthiness of the AI system as a decision-maker is a crucial aspect in deciding whether to fully depend on AI, advocate for a hybrid human–AI decision-making approach, or acknowledge epistemic authority exclusively in human expertise. Nonetheless, despite the statistically significant differences between groups favoring a collaborative human–AI approach and those who continue to depend solely on traditional human authority for decisions pertinent to personal matters (such as criminal judgment scenarios), these groups may not exhibit a pronounced behavioral difference in practice. Similarly, a statistically significant difference was found between the three groups of decision-making supports concerning the attitude towards authority in the criminal judgment scenario ( $\chi^2 = 7.9952$ ,  $p = 0.01836$ ,  $\eta^2 = 0.0112$ ). The two-group Mann–Whitney tests, on the other hand, did not show any statistically significant differences when comparing the groups pairwise. This suggests that there is at least one group exhibiting a different attitude toward authority in the context of criminal judgment on the basis of the chosen decision-making support. In the same way, a significant difference between trusting AI to make decisions, working together with AI, or listening to human authority in the case of job seeking tasks was observed when comparing faith in the automation performance in the three decision support groups ( $\chi^2 = 6.5179$ ,  $p = 0.03843$ ,  $\eta^2 = 0.00841$ ). However, multiple pairwise comparisons between groups did not show a significant difference relative to the degree of AI's unquestionable authority. Table 9 shows the results.

Overall, the findings confirmed the significance of the preliminary associations between the predisposing factors and the hypothesized dimensions (especially in the case of trust in automation—TiA, and faith in automation performance—PAS), suggesting that humans could likely consider AI advisors as authorities, even though the effect size of the statistical tests ranged from small to small-to-medium. Therefore, this prompts us to recognize that the practical implications of these associative connections are minimal. The observed low and medium-low effect sizes suggest that while there are distinctions in the perception of TiA and PAS factors between the two groups, these differences are not enough to establish a conclusive preference for individuals exhibiting higher trust and expectations in automated systems compared to those with lower confidence in automated decision support concerning the attribution of epistemic authority to AI systems in practice.

**Table 9.** Kruskal–Wallis tests grouped by decision support. Pairwise comparisons between group levels with Bonferroni corrections for multiple testing using the Wilcoxon run sum test with continuity correction function in R [41,43].  $\chi^2$  statistic represents the Kruskal–Wallis test statistic.  $\eta^2$  is the effect size [44]. <sup>1</sup>  $p$ -value-adjusted Bonferroni. \*  $p$  value < 0.05. PoJ = Preference on Judgment. (a), (b), (c) are the ad hoc items—see Tables 3 and 4. PMW test is the pairwise Mann–Whitney test. H-H is Hybrid–Human.

	$\chi^2$	$p$ -Value	$\eta^2$	PMW Test	$p$ -adj <sup>1</sup>
TiA-PoJ (a)	14.142	0.0008492 *	0.0226	H-H	0.0016 *
PAS-PoJ (a)	4.4416	0.1085	0.00455	-	-
MFQ-PoJ (a)	7.9952	0.01836 *	0.0112	-	-
TiA-PoJ (b)	2.5591	0.2782	0.00104	-	-
PAS-PoJ (b)	6.5179	0.03843 *	0.00841	-	-
MFQ-PoJ (b)	2.9134	0.233	0.00170	-	-
TiA-PoJ (c)	5.6236	0.0601	0.00675	-	-
PAS-PoJ (c)	3.3988	0.1828	0.00260	-	-
MFQ-PoJ (c)	0.15968	0.9233	0.00	-	-

## 5. Discussion

This study explored the factors influencing the attribution of epistemic authority to AI systems, that is, *alority*, through an empirical investigation leveraging psychometric tools and scenario-based assessments. The results offer critical insights into the interplay between trust, perceived performance in automation, and human preference in delegating decision-making authority to automated systems.

### 5.1. Is Expert Judgment Being Replaced by AI as Epistemic Authority?

The findings suggest that trust in automation (TiA) and the perceived performance of AI systems (PAS) significantly influence participants' readiness to accept AI as epistemic authorities. Firstly, the stepwise analysis indicates that both trust in automation (TiA) and belief in its effectiveness (PAS) might stimulate individuals' attitudes about substituting human judgment with automated decision-making. Nevertheless, we did not find that individual trust in automation (TiA) and the cognitive patterns associated with human views regarding the performance of automated systems (PAS) had a significant role in the propensity to replace expert judgment, except weakly for those who had expressed a medium-to-high trust in automated decision-making. This is interesting in light of the recent study by [27], which posits that the concept of epistemic authority is significantly more fine-grained than that of expertise, indicating that the criteria for expertise should not be encompassed within the definition of epistemic authority. In a similar vein, ref. [45] contend that the notions of expertise and epistemic authority are not coextensive, concluding that there are significant distinctions between the issues of identifying authority and identifying expertise, particularly regarding the challenges that laypersons encounter in addressing them. Furthermore, scholars observed that when experts disagree, laypersons must discern the most credible source within a field, despite their lack of expertise, while navigating an environment where two or more sources, seemingly equally competent to them, present conflicting perspectives [46]. On the other hand, the work by [47] indicates that deference to scientific authority and trust in scientists are separate constructs, with deference notably forecasting elevated trust in scientific figures regarding science-related matters, and epistemic trust which pertains to the readiness to regard new information from others as credible, applicable, and pertinent [48]. In this view, ref. [49] conclude that while

deference theoretically arises from a confidence in the authority of scientific knowledge, various other conditions convert this conviction into actual deference to that authority in diverse settings. Notably, recent research has shown that users confer authority to AI systems based on sources that are frequently unrelated to the system, and do not reflect the reliability, utility, or efficacy of the AI in question [50]. However, ref. [51] argued that the proficiency exhibited by Large Language Models in question-answering tasks may create a perception that these AI systems hold competencies that categorize them as epistemic experts, thereby suggesting an epistemic obligation to rely on the predictions made by these AI systems [51,52]. In this view, further analysis is needed to explore whether other factors may elucidate the underlying reasons for the preference of replacing expert judgment as epistemic authority.

### *5.2. The Importance of Operational Domain and Task Nature in Granting Epistemic Authority to AI*

As expected, we found a remarkable association between both trust in automation (TiA) and perceptions of automated system performance (PAS) with users' attitudes towards errors in assistance, especially among those with elevated expectations of automated decision-making efficacy and trust in automation. This is consistent with the literature in which scholars argue that people have higher performance expectations toward algorithms than humans [31] and a faulty human-automation interaction tends to induce disuse in the automation aids even though users know the aid is more accurate than they are (e.g., [26,53,54]). However, recent works suggest that AI authority tends to influence users' attitudes towards AI, including trust in AI skills, self-blame for negative and biased results, forgiveness, and thankfulness towards AI, resulting in increased tolerance for AI alerts and diminished acknowledgment of AI biases [50]. In this vein, ref. [55] noticed that first-time failures may result in a lack of reliance in an algorithm, but increased familiarity may foster excessive trust, particularly in fields like automated driving [56] and healthcare. Research indicates that, although previous beliefs affect the initial trust in AI, continuing interactions and the resultant familiarity also have a significant role [57]. Interestingly, when we tested differences between groups on the basis of their degree of faith in automation performance (PAS), we found no significant association between automated aids in route planning tasks and individuals' perception of AI accuracy, despite a preliminary association in the correlation analysis. This may lead to the consideration that faith in automated performance (PAS) appears to be a predisposing factor in acknowledging AI's epistemic authority mainly in high-stakes scenarios, which are situations with potentially far-reaching consequences for an individual's future [50]. Trust in automation (TiA), instead, appears to be a fundamental factor that predisposes people to recognize AI as an authority. For instance, participants displayed acceptance of AI in predictive domains like dating or job-matching and in areas requiring moral discretion, such as criminal justice. This suggests that individuals perceive AI as reliable advisors in domains defined both by quantifiable outcomes and subjective or ethical nuances, reinforcing the hypothesis that perceived task suitability is pivotal in determining AI's acceptance as an authority. This is especially significant considering recent research that has established a strong connection between trust in AI and the perceived risk associated with decision-making, indicating that individuals are more inclined to trust AI as the perceived risk of the decision escalates [55]. Furthermore, our findings appear to align with the evidence gathered that the operational domain is a crucial driver of trust [55] as well as the nature of the task [57]. Ref. [58] observed that participants exhibited a greater reliance on human judgment than on algorithmic input when tasked with predicting the funniness of jokes. Conversely, ref. [59] found that participants relied more on algorithms than on humans for numerical tasks with an objectively correct answer. In this perspective, in consumer research, ref. [60] states that consumers exhibit varying degrees of algorithmic

inclination or aversion, contingent upon the nature of the task for which the algorithm is employed and the perception of that task. The authors conclude that enhancing the perceived objectivity of a task pairs with an increase in trust and utilization of algorithms associated with that task, thereby suggesting that improving the perceived emotional similarity of algorithms to humans effectively encourages their application in subjective tasks [60].

### *5.3. The Role of Individuals' Moral Foundations in Acknowledging Authority in AI*

While trust in AI (TiA) positively aligns with preferences for its deployment in high-stakes scenarios, traditional moral frameworks (as captured by MFQ scores) inversely correlate with reliance on algorithmic authority in sensitive decisions, such as criminal judgments. These outcomes reveal a tension between inherent trust in the technological capabilities of AI and cultural or ethical predispositions emphasizing human oversight. This might align with the current discourse around AI deference and human oversight. Ref. [29] endorses the assertion that if AI cannot establish a reliably superior viewpoint in the pertinent domain, or if it does not address counterarguments acknowledged by a competent human, then reliance on it becomes illogical. However, a recent study by [61] on the potential associations between individual moral foundations, assessed via the Moral Foundations Questionnaire (MFQ), and participants' perceptions and evaluations of morally contentious AI behavior, found that sensitivity to care/harm violations was the most significant predictor of the perceived wrongness of AI behavior in relation to moral foundations. The study concludes that an individual's perception of AI systems is variable and influenced by elements beyond their moral roots [61]. Therefore, additional studies are required to investigate how individual variations in moral underpinnings and evaluations of AI systems may influence individuals' inclination to acknowledge authority in AI. However, when we compared attitudes toward authority in the three different decision support groups, no clear dominant category of decision support emerged in terms of predisposition toward judgment or moral relevance of authority, contrary to our initial expectations. Interestingly, subjects who preferred a collaborative approach with AI, as opposed to relying solely on human judgment, showed significantly greater trust in automation in particularly sensitive decision-making choices, such as criminal judgment. Some studies argue that knowing more about the technical side of AI systems might change how people judge the behavior of AIs [61]. Other studies suggest that there is a clear difference in how people see moral violations committed by AIs compared to human offenders [62], which means that more research is needed, especially since AIs of the future will perform a wider range of actions that are morally significant [62]. This might suggest that, in contexts of particular ethical and moral relevance, personal moral judgment can play a decisive role in decision support choices, as long as there is greater awareness of the potential effects of using automated support while recognizing its potential for support. Further research is required to confirm this point by simultaneously examining whether all of TiA, MFQ and PAS affect humans' propensity to ascribe epistemic authority to AI advisors in high vs. low critical contexts, including the task's complexity for which AI provides guidance.

### *5.4. Gender and Epistemic Authority of AI*

The findings on gender-based differences warrant further investigation, as they may indicate underlying sociocultural or experiential factors shaping perceptions of automation. Studies indicate that gender influences both aversion and appreciation of technology utilization [63,64] and can induce stereotypes or biases in human-AI teaming [65], rendering the subject of gender disparities in algorithm acceptance to be fully explored [64]. Unexpectedly,

the association between trust in automation (TiA) and individuals' confidence in considering AI as epistemic authorities in personal domains (e.g., creditworthiness assessments) was weaker than anticipated. This may be due to heightened awareness of potential biases and fairness issues in AI systems, a growing concern in public discourse. This highlights the need for enhanced transparency and accountability in algorithmic decision-making [66,67]. This is noteworthy because our study primarily focuses on younger respondents, which allows us to identify significant demographic influences, such as those who are more familiar with technology and exhibit greater openness to algorithmic decision-making. These patterns should resonate with the "digital-native" hypothesis, which posits generational shifts in technology adoption and trust, as confirmed in the literature (e.g., [68]).

## 6. Limitations and Further Research

Notwithstanding our interesting results, we believe that several limits of our study must be acknowledged.

Firstly, the stepwise method is based on associative hypotheses derived from correlation analysis and subsequent non-parametric tests which impede inferences of causal relationships. We should extend this initial result, which identifies hypothetical predictors of alority and their presumed relationship to the variable, by testing the various relationships simultaneously. This will help us understand whether there is a basis for modeling complex theoretical causal relationships between constructs, exploring how consistent the resulting model is with the data.

Secondly, ad hoc constructs cannot be deemed to be comprehensive of the several real-world scenarios pertinent for evaluating the propensity towards algorithmic versus human authority. Future study should enhance and consolidate the psychometric instrument with the aim to operationalise the algorithmic construct. In this vein, this study paves the way for further experimental and qualitative research. As an example, our associative findings might be investigated in an interview-based study design to elucidate and unfold the most fuzzified associations related to conferring epistemic authority to AI. Similarly, additional experimental research ought to be devised to evaluate our early findings in pertinent contexts, such as healthcare and law, that are grounded in ethical foundations and the epistemic superiority of human expertise. Both scenarios may be appropriate for enhancing the understanding of the preference we observed regarding human–AI collaborative solutions in making decisions; maybe this would also allow for the investigation of whether "AI is successfully appropriated by a human agent" [51], meaning that it enhances users' epistemic skills and effectively helps them achieve their epistemic goals.

Thirdly, the sample was mainly restricted to university students under 30 years of age, hence constraining the generalizability and comparability of our findings across diverse user groups. Future research should investigate how aged respondents engage with algorithmic deference tasks and the epistemic role of AI to attain a more nuanced understanding of how reliance on AI, expectations of AI performance, and authority-related moral attitudes may serve as potential predictors for measuring algorithmic authority. Furthermore, our sample comprised respondents only in Italy. Future research should incorporate respondents from diverse geographical locations to assess the impact of cultural differences on trust in AI and the influence on individuals' moral foundations when dealing with AI within ethically sensitive contexts.

As for further research, our findings have important implications for the design and implementation of AI systems. Our findings suggested that trust is the underlying predictor of acknowledging AI as an epistemic authority. Nevertheless, although respondents exhibit confidence in automated decision-making when AI's output may mitigate errors, their assessment of automation's efficacy was deemed insufficient to supplant expert judgment.

This implies that AI systems should be designed to strengthen users' perception of engaging with a skilled counterpart using appropriate cues of their effectiveness as epistemic authority. As an example, ref. [69] concludes the necessity of incorporating "professional credibility indicators" into AI to bolster physicians' trust in AI assistance. This underscores the pressing necessity for innovative design approaches to capture the epistemic authority ascribed to AI systems. Conversely, when users receive outputs that are precise, coherent, and contextually relevant, they may reasonably regard them with the same trust as they would experts [30]. In this view, our findings concerning respondents' preferences for AI accuracy in error assistance need significant considerations regarding the design of AI as an epistemic technology, which is intended to function on epistemic content through processes such as inference, prediction, and analysis [70]. This indicates that the behavior of such technologies is significantly influenced by training data [51], hence underscoring the necessity of including design approaches that are strongly ethical by design. Moreover, developers and policymakers must consider the contextual and psychological dimensions of trust to enhance user acceptance. Incorporating mechanisms for greater transparency, such as explainable AI, and promoting hybrid human–AI decision frameworks may bridge gaps in trust and perceived efficacy. For instance, scenarios blending human discretion with AI's computational strength (e.g., human judges supported by AI in courtrooms) emerged as a preferred model in this study, suggesting a pathway for integrating AI into socially sensitive domains. The observed preference towards hybrid collaboration between humans and AI in sensitive decision-making contexts calls for a comprehensive understanding of authority and competence in shared decision-making processes, to ascertain new evaluative parameters for measuring the effects and quality of decisions in collaborative human–AI decisions. Future research should build on these findings by exploring longitudinal changes in trust dynamics as exposure to and reliance on AI systems increase. Lastly, investigating cross-cultural variations could also provide a more global perspective on algorithmic authority. Additionally, there is a need to examine how real-world deployments of AI, beyond hypothetical scenarios, shape user trust and acceptance over time.

## 7. Conclusions

This exploratory study aims at identifying the main factors that contribute to individuals' confidence in viewing AI advisors as authorities, offering preliminary insights into the tendency to attribute epistemic authority to AI advisors, referred to as *algority*. Overall, our findings showed that there exists a tendency to acknowledge epistemic authority to AI, although further analysis is needed to test facilitators making algorithmic authority predominant over human authority. While faith in automated performance (PAS) appears to be a predisposing factor in acknowledging AI's epistemic authority, particularly in high-stakes scenarios with potentially far-reaching consequences for an individual's future, trust in automation (TiA) has emerged as a fundamental element in acknowledging AI as an epistemic authority. However, personal and moral judgment toward authority (as captured by MFQ) may moderate individuals' propensity to rely on automated decision-making in settings with moral tensions (e.g., criminal judgment) or affect users' sensitivity when AI systems suggest options (e.g., job interviewing). While trust in automation (TiA) and perceptions of automated system performance (PAS) have been confirmed to be meaningfully associated with users' attitudes towards errors in assistance, it appears that neither trust in automation (TiA) nor belief in its effectiveness (PAS) significantly influences the propensity to replace expert judgment. On the other hand, our research sheds light on the multiple dimensions of *algority* as a construct and paves the way for further research on *algority* as a standalone construct. These insights contribute to the theoretical discourse on algorithmic authority and offer practical implications for designing AI systems that are

not only trustworthy but also empower users to engage critically with algorithmic outputs (e.g., [71]).

**Author Contributions:** Conceptualization, F.M. and F.C.; methodology, F.M. and F.C.; software, F.M.; formal analysis, F.M.; data curation, F.M.; writing—original draft preparation, F.M. and F.C.; writing—review and editing, F.M. and F.C.; supervision, F.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work of Frida Milellawas supported by the co-funding of European Union – Next Generation EU, in the context of the National Recovery and Resilience Plan, PE8 “Conseguenze e sfide dell’invecchiamento”, Project Age-It (AGE—IT—A Novel Public-private Alliance to Generate Socioeconomic, Biomedical and Technological Solutions for an Inclusive Italian Ageing Society—Ageing Well in an Ageing Society)—AGE-IT-PE00000015 CUP: H43C22000840006.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We are grateful to Massimo Airoidi, Chiara Natali and Luca Marconi for their assistance in emphasising the reference topics in the previous version of the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Grimmelikhuijsen, S. Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Adm. Rev.* **2023**, *83*, 241–262. [CrossRef]
- Rhoen, M.; Feng, Q.Y. Why the ‘Computer says no’: Illustrating big data’s discrimination risk through complex systems science. *Int. Data Priv. Law* **2018**, *8*, 140–159. [CrossRef]
- Tingle, J. The computer says no: AI, health law, ethics and patient safety. *Br. J. Nurs.* **2021**, *30*, 870–871. [CrossRef] [PubMed]
- Vik, P. ‘The computer says no’: The demise of the traditional bank manager and the depersonalisation of British banking, 1960–2010. *Bus. Hist.* **2017**, *59*, 231–249. [CrossRef]
- Wihlborg, E.; Larsson, H.; Hedström, K. “The Computer Says No!”—A Case Study on Automated Decision-Making in Public Authorities. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 2903–2912.
- Sundin, O.; Haider, J.; Andersson, C.; Carlsson, H.; Kjellberg, S. The search-ification of everyday life and the mundane-ification of search. *J. Doc.* **2017**, *73*, 224–243. [CrossRef]
- Shirky, C. A Speculative Post on the Idea of Algorithmic Authority. 2017. Available online: <https://www.bibsonomy.org/url/a4f71c8404afbb43b64a2c03196fe5e5> (accessed on 24 January 2026).
- Forte, A. The new information literate: Open collaboration and information production in schools. *Int. J. Comput.-Support. Collab. Learn.* **2015**, *10*, 35–51. [CrossRef]
- Lustig, C.; Nardi, B. Algorithmic authority: The case of Bitcoin. In Proceedings of the 2015 48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 5–8 January 2015; pp. 743–752.
- Lustig, C.; Pine, K.; Nardi, B.; Irani, L.; Lee, M.K.; Nafus, D.; Sandvig, C. Algorithmic authority: The ethics, politics, and economics of algorithms that interpret, decide, and manage. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 1057–1062.
- Danaher, J. The threat of algocracy: Reality, resistance and accommodation. *Philos. Technol.* **2016**, *29*, 245–268. [CrossRef]
- Ståhl, T.; Sormunen, E.; Mäkinen, M. Epistemic beliefs and internet reliance—Is algorithmic authority part of the picture? *Inf. Learn. Sci.* **2021**, *122*, 726–748. [CrossRef]
- Cabitzza, F.; Campagner, A.; Simone, C. The need to move away from agential-AI: Empirical investigations, useful concepts and open issues. *Int. J. Hum.-Comput. Stud.* **2021**, *155*, 102696. [CrossRef]
- Schwarz, O. *Sociological Theory for Digital Society: The Codes That Bind Us Together*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
- Burrell, J.; Fourcade, M. The Society of Algorithms. *Annu. Rev. Sociol.* **2021**, *47*, 213–237. [CrossRef]
- Beer, D. Power Through the Algorithm? Participatory Web Cultures and the Technological Unconscious. *New Media Soc.* **2009**, *11*, 985–1002. [CrossRef]
- Noble, S.U. *Algorithms of Oppression*; New York University Press: New York, NY, USA, 2018. [CrossRef]
- Cheney-Lippold, J. *We Are Data*; New York University Press: New York, NY, USA, 2017. [CrossRef]

19. Araujo, T.; Helberger, N.; Kruijkemeier, S.; De Vreese, C.H. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI Soc.* **2020**, *35*, 611–623. [[CrossRef](#)]
20. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [[CrossRef](#)]
21. Kizilcec, R.F. How much information? Effects of transparency on trust in an algorithmic interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2390–2395.
22. Lee, M.K. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data Soc.* **2018**, *5*, 2053951718756684. [[CrossRef](#)]
23. Wiegmann, D.A.; Rich, A.; Zhang, H. Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theor. Issues Ergon. Sci.* **2001**, *2*, 352–367. [[CrossRef](#)]
24. Zhang, H.; Hook, J.N.; Johnson, K.A. Moral Foundations Questionnaire. In *Encyclopedia of Personality and Individual Differences*; Zeigler-Hill, V., Shackelford, T.K., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 1–3. [[CrossRef](#)]
25. Kohn, S.C.; De Visser, E.J.; Wiese, E.; Lee, Y.C.; Shaw, T.H. Measurement of trust in automation: A narrative review and reference guide. *Front. Psychol.* **2021**, *12*, 604977. [[CrossRef](#)]
26. Merritt, S.M.; Unnerstall, J.L.; Lee, D.; Huber, K. Measuring individual differences in the perfect automation schema. *Hum. Factors* **2015**, *57*, 740–753. [[CrossRef](#)] [[PubMed](#)]
27. Bokros, S.E. A deference model of epistemic authority. *Synthese* **2021**, *198*, 12041–12069. [[CrossRef](#)]
28. Hauswald, R. Artificial epistemic authorities. *Soc. Epistemol.* **2025**, *39*, 716–725. [[CrossRef](#)]
29. Lange, B. Epistemic Deference to AI. In *Proceedings of the International Conference on Bridging the Gap Between AI and Reality*; Springer Nature: Cham, Switzerland, 2024; pp. 174–186.
30. Yang, S.; Ma, R. Classifying Epistemic Relationships in Human-AI Interaction: An Exploratory Approach. *arXiv* **2025**, arXiv:2508.03673. [[CrossRef](#)]
31. Renier, L.A.; Mast, M.S.; Bekbergenova, A. To err is human, not algorithmic—Robust reactions to erring algorithms. *Comput. Hum. Behav.* **2021**, *124*, 106879. [[CrossRef](#)]
32. Rubbi, I.; Lupo, R.; Lezzi, A.; Cremonini, V.; Carvello, M.; Caricato, M.; Conte, L.; Antonazzo, M.; Caldararo, C.; Botti, S.; et al. The social and professional image of the nurse: Results of an online snowball sampling survey among the general population in the post-pandemic period. *Nurs. Rep.* **2023**, *13*, 1291–1303. [[CrossRef](#)] [[PubMed](#)]
33. Kennedy-Shaffer, L.; Qiu, X.; Hanage, W. Snowball Sampling Study Design for Serosurveys Early in Disease Outbreaks. *Am. J. Epidemiol.* **2021**, *190*, 1918–1927. [[CrossRef](#)]
34. Zickar, M.; Keith, M. Innovations in Sampling: Improving the Appropriateness and Quality of Samples in Organizational Research. *Annu. Rev. Organ. Psychol. Organ. Behav.* **2022**, *10*, 315–337. [[CrossRef](#)]
35. Graham, J.; Haidt, J.; Nosek, B.A. Liberals and conservatives rely on different sets of moral foundations. *J. Personal. Soc. Psychol.* **2009**, *96*, 1029. [[CrossRef](#)] [[PubMed](#)]
36. Harper, C.A.; Rhodes, D. Reanalysing the factor structure of the moral foundations questionnaire. *Br. J. Soc. Psychol.* **2021**, *60*, 1303–1329. [[CrossRef](#)]
37. Körber, M. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VI: Transport Ergonomics and Human Factors (TEHF); Aerospace Human Factors and Ergonomics 20*; Springer: Cham, Switzerland, 2019; pp. 13–30.
38. Taherdoost, H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *Int. J. Acad. Res. Manag. (IJARM)* **2016**, *5*, 28–36. [[CrossRef](#)]
39. Ursachi, G.; Horodnic, I.A.; Zait, A. How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Econ. Financ.* **2015**, *20*, 679–686. [[CrossRef](#)]
40. Hair, J.F.; Ringle, C.M.; Gudergan, S.P.; Fischer, A.; Nitzl, C.; Menictas, C. Partial least squares structural equation modeling-based discrete choice modeling: An illustration in modeling retailer choice. *Bus. Res.* **2019**, *12*, 115–142. [[CrossRef](#)]
41. Okoye, K.; Hosseini, S. Correlation tests in R: Pearson cor, kendall's tau, and spearman's rho. In *R Programming: Statistical Data Analysis in Research*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 247–277.
42. McKight, P.E.; Najab, J. Kruskal-wallis test. In *The Corsini Encyclopedia of Psychology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; Volume 1. [[CrossRef](#)]
43. Fang, Z.; Du, R.; Cui, X. Uniform approximation is more appropriate for wilcoxon rank-sum test in gene set analysis. *PLoS ONE* **2012**, *7*, e31505. [[CrossRef](#)]
44. Tomczak, M.; Tomczak, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* **2014**, *1*, 19–25.
45. Keren, A. Expert Authority and Its Assessment. *Soc. Epistemol.* **2025**, *39*, 612–625. [[CrossRef](#)]
46. Croce, M.; Baghrarian, M. Experts—Part II: The Sources of Epistemic Authority. *Philos. Compass* **2024**, *19*, e70005. [[CrossRef](#)]

47. Anderson, A.A.; Scheufele, D.A.; Brossard, D.; Corley, E.A. The role of media and deference to scientific authority in cultivating trust in sources of information about emerging technologies. *Int. J. Public Opin. Res.* **2012**, *24*, 225–237. [[CrossRef](#)]
48. Schroder-Pfeifer, P.; Talia, A.; Volkert, J.; Taubner, S. Developing an assessment of epistemic trust: A research protocol. *Res. Psychother. Psychopathol. Process Outcome* **2018**, *21*, 330. [[CrossRef](#)]
49. Howell, E.L.; Wirz, C.D.; Scheufele, D.A.; Brossard, D.; Xenos, M.A. Deference and decision-making in science and society: How deference to scientific authority goes beyond confidence in science and scientists to become authoritarianism. *Public Underst. Sci.* **2020**, *29*, 800–818. [[CrossRef](#)]
50. Kapania, S.; Siy, O.; Clapper, G.; Sp, A.M.; Sambasivan, N. “Because AI is 100% right and safe”: User attitudes and sources of AI authority in India. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–18.
51. Ferrario, A.; Facchini, A.; Termine, A. Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems. *Minds Mach.* **2024**, *34*, 30. [[CrossRef](#)]
52. Grote, T.; Berens, P. On the ethics of algorithmic decision-making in healthcare. *J. Med. Ethics* **2020**, *46*, 205–211. [[CrossRef](#)] [[PubMed](#)]
53. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* **1997**, *39*, 230–253. [[CrossRef](#)]
54. Moes, M.; Knox, K.; Pierce, L.; Beck, H. Should I decide or let the machine decide for me. In Proceedings of the Poster Presented at the Meeting of the Southeastern Psychological Association, Savannah, GA, USA, 18–21 March 1999.
55. Schoenherr, J.R.; Thomson, R. When AI fails, who do we blame? Attributing responsibility in human–AI interactions. *IEEE Trans. Technol. Soc.* **2024**, *5*, 61–70. [[CrossRef](#)]
56. Wagner, A.R.; Borenstein, J.; Howard, A. Overtrust in the robotic age. *Commun. ACM* **2018**, *61*, 22–24. [[CrossRef](#)]
57. Siau, K.; Wang, W. Building trust in artificial intelligence, machine learning, and robotics. *Cut. Bus. Technol. J.* **2018**, *31*, 47.
58. Yeomans, M.; Shah, A.; Mullainathan, S.; Kleinberg, J. Making sense of recommendations. *J. Behav. Decis. Mak.* **2019**, *32*, 403–414. [[CrossRef](#)]
59. Logg, J.M.; Minson, J.A.; Moore, D.A. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Processes* **2019**, *151*, 90–103. [[CrossRef](#)]
60. Castelo, N.; Bos, M.W.; Lehmann, D.R. Task-dependent algorithm aversion. *J. Mark. Res.* **2019**, *56*, 809–825. [[CrossRef](#)]
61. Brailsford, J.; Vetere, F.; Velloso, E. Exploring the Association between Moral Foundations and Judgements of AI Behaviour. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024; pp. 1–15.
62. Maninger, T.; Shank, D.B. Perceptions of violations by artificial and human actors across moral foundations. *Comput. Hum. Behav. Rep.* **2022**, *5*, 100154. [[CrossRef](#)]
63. Hoff, K.A.; Bashir, M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust: Integrating Empirical Evidence on Factors That Influence Trust. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2014**, *57*, 407–434. [[CrossRef](#)]
64. Kaufmann, E.; Chacon, A.; Kausel, E.E.; Herrera, N.; Reyes, T. Task-specific algorithm advice acceptance: A review and directions for future research. *Data Inf. Manag.* **2023**, *7*, 100040. [[CrossRef](#)]
65. Milella, F.; Natali, C.; Scantamburlo, T.; Campagner, A.; Cabitza, F. The impact of gender and personality in human-AI teaming: The case of collaborative question answering. In Proceedings of the IFIP Conference on Human-Computer Interaction, York, UK, 28 August–1 September 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 329–349.
66. Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*; Harvard University Press: Cambridge, MA, USA, 2015.
67. Shin, D.; Park, Y.J. Role of fairness, accountability, and transparency in algorithmic affordance. *Comput. Hum. Behav.* **2019**, *98*, 277–284. [[CrossRef](#)]
68. Kim, T.; Molina, M.D.; Rheu, M.; Zhan, E.S.; Peng, W. One AI does not fit all: A cluster analysis of the laypeople’s perception of AI roles. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; pp. 1–20.
69. Ding, X.; Xing, C. Trust in AI vs. human doctors: The roles of subjective understanding, perceived epistemic authority and social proof. *Acta Psychol.* **2025**, *261*, 105945. [[CrossRef](#)]
70. Alvarado, R. AI as an epistemic technology. *Sci. Eng. Ethics* **2023**, *29*, 32. [[CrossRef](#)] [[PubMed](#)]
71. Facchini, A.; Fregosi, C.; Natali, C.; Termine, A.; Wilson, B. Algorithmic Authority & AI Influence in Decision Settings: Theories and Implications for Design. In Proceedings of the 12th International Conference on Human-Agent Interaction, Swansea, UK, 24–27 November 2024; pp. 472–474.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.