Moral Framing of Military Artificial Intelligence: The Role of Trust and Attitudes in Defensive and Offensive Use

Oronzo Parlangeli
DISPOC
University of Siena
Siena, Italy
oronzo.parlangeli@unisi.it

Antonio Brancati
DISPOC
University of Siena
Siena, Italy
a.brancati@student.unisi.it

Giorgia Callori
DISPOC
University of Siena
Siena, Italy
g.callori@student.unisi.it

Matteo Chiasso
DISPOC
University of Siena
Siena, Italy
m.chiasso@student.unisi.it

Arianna Fiorito
DISPOC
University of Siena
Siena, Italy
a.fornito@student.unisi.it

Francesca Mambrini
DISPOC
University of Siena
Siena, Italy
f.mambrini1@student.unisi.it

Michele Pio Piscopo
DISPOC
University of Siena
Siena, Italy
m.piscopo1@student.unisi.it

Marta Veroni
DISPOC
University of Siena
Siena, Italy
m.veroni@student.unisi.it

Francesco Curro'
DISPOC
University of Siena
Siena, Italy
francesco.curro2@unisi.it

Abstract

Despite growing research on public perceptions of Artificial Intelligence (AI), limited attention has been paid to how moral framing shapes evaluations of military AI. Drawing on research on trust in AI and the halo effect, this study examines whether offensive versus defensive framing influences users' perceptions across multiple dimensions, including cognitive attitudes, affective attitudes, and perceived intelligence. The study employed an online questionnaire with a between-subjects design. Participants were randomly assigned to one of two conditions, in which they were presented with a scenario describing the development of an AI algorithm for either defensive or offensive military purposes. Results showed that, compared to defensive framing, offensive framing significantly reduced both cognitive and affective attitudes toward the AI system, as well as its perceived intelligence. Trust in AI emerged as a strong and independent predictor across outcomes. Structural equation modeling indicated that the effect of framing on perceived intelligence was largely mediated by affective and cognitive attitudes, consistent with a halo-like process in which moral evaluations shape competence judgments. No moderating effects of identification with all humanity were observed; personality traits were not significant predictors of attitudinal evaluations. However, identification with all humanity showed a significant negative association with affective attitudes. These findings suggest that the perceived intelligence of military AI does not reflect an independent assessment of system capability, but emerges from cognitive, affective and moral evaluations triggered by its intended use. More broadly, they highlight how relatively subtle differences in framing can systematically influence public perceptions, raising concerns about the potential use of defensive narratives as persuasive devices in the communication of military AI. This study contributes to HCI by showing that AI evaluations are not only context-dependent but also morally constructed, with important implications for the design, communication, and governance of AI in high-stakes domains.

CCS Concepts

• **Human-centered computing** → *User models*; • **Applied computing** → *Psychology*; • **Computing methodologies** → *Artificial intelligence.*

Keywords

AI, military AI, trust, attitudes, intelligence perception, halo effect, ethical implications, personality traits, deceptive patterns

ACM Reference Format:
Oronzo Parlangeli, Antonio Brancati, Giorgia Callori, Matteo Chiasso, Arianna Fiorito, Francesca Mambrini, Michele Pio Piscopo, Marta Veroni, and Francesco Curro'. 2026. Moral Framing of Military Artificial Intelligence: The Role of Trust and Attitudes in Defensive and Offensive Use. In *European Conference on Cognitive Ergonomics (ECCE 2026), September 15–18, 2026, Siena, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3822301.3822318

1 Introduction

Artificial Intelligence (AI) is increasingly deployed in high-stakes contexts where operational risks are significant. In such domains, perceptions of reliability, legitimacy, and controllability are as critical as technical performance. Research has consistently shown that perceptions and acceptance of AI systems are not static; rather, they depend on contextual factors such as task type, perceived risk, and system framing [17, 30, 42]. Among these factors, trust has been



This work is licensed under a Creative Commons Attribution 4.0 International License.
ECCE 2026, Siena, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2515-9/26/09
https://doi.org/10.1145/3822301.3822318

identified as a key determinant, shaping individuals' willingness to rely on and interact with AI systems [14, 29, 35].

These considerations are particularly salient in the military domain, where AI systems are associated with both potential benefits, such as enhanced operational efficiency and reduced human exposure to risk, and significant ethical concerns related to autonomy, accountability, and the delegation of potentially harmful decisions. In this context, public perception (closely linked to trust in AI) plays a pivotal role in determining the legitimacy and acceptance, or rejection, of such technologies, particularly within democratic frameworks [11, 24, 41].

Beyond societal-level perspectives, psychological research highlights systematic individual differences in how people perceive technology, risk, and morally sensitive domains. Personality traits, as conceptualized within the Big Five model, and identity-related constructs such as Identification With All Humanity (IWAH), influence how individuals evaluate situations involving harm, responsibility, and collective outcomes [36, 37]. Despite these advances, research on AI perception, particularly in military contexts, has predominantly focused on system characteristics and contextual factors, with limited attention to how individual differences interact with moral framing.

Moreover, a key distinction remains underexplored: whether individuals evaluate military AI differently depending on whether it is framed as offensive or defensive. This distinction goes beyond operational differences and reflects a deeper moral framing. Defensive deployments of AI are typically associated with protection and legitimacy, whereas offensive uses are linked to harm and escalation. Importantly, such framing may not only influence moral and affective evaluations but may also produce a halo effect [39], whereby negative evaluations extend to other system attributes, including perceived competence. Besides its theoretical relevance, the issue of moral framing also raises important questions regarding the role of communication in shaping public understanding of AI systems [46]. If evaluations of AI are sensitive to how its functions are described, then the way in which military AI is presented to the public may not simply reflect underlying technical characteristics, but may actively influence perceptions of legitimacy, acceptability, and competence. In this sense, framing can be understood not only as a cognitive biasing mechanism, but also as a potentially strategic component of communication, with implications for how information about AI systems is conveyed in public and institutional contexts. These considerations highlight the need to examine not only how individuals evaluate AI, but also how such evaluations may be shaped by the presentation of the technology itself.

2 Related Work

2.1 Trust and framing effects in AI perception

Trust in technology is a key determinant of its adoption and use [1, 16, 29, 35]. In the context of AI, trust depends not only on technical features such as accuracy and transparency but also on users' knowledge and beliefs, which evolve through experience [10].

Trust in AI is a complex and multifaceted construct that has been examined from various theoretical perspectives, often leading to heterogeneous findings [17, 29]. According to Glikson and Woolley [2020], trust in AI is structured into a cognitive dimension, based on the rational evaluation of technical competence and reliability, and an affective dimension, rooted in emotional responses and social cues. The authors highlight that these components develop through distinct processes, suggesting that AI acceptance depends on the integration of both performance-based judgments and emotional reactions [17]. At the same time, perceptions of AI are not uniform but vary across different domains and levels of perceived risk. Comparative and cross-national studies show systematic differences across sectors such as healthcare, finance, and transportation, reflecting concerns related to safety, ethics, accountability, and data protection [3, 5]. Similarly, the perceived value of AI (whether as augmentative or substitutive) depends on context and task characteristics [20, 31], suggesting that AI is evaluated in relation to its specific application rather than as a homogeneous technology.

As trust develops through experience, it can serve as a relatively stable cognitive anchor over time [17, 29]. Consequently, it is primarily grounded in evaluations of system competence, including reliability, transparency, and performance [17, 35]. As such, trust is likely to inform outcome-related judgments that reflect the system's perceived capability, such as evaluations of its intelligence.

In contrast to competence-based evaluations, attitudinal and affective responses toward AI are more context-dependent and dynamically shaped by situational factors such as framing, exposure, and perceived risk [13, 40, 42]. These responses are highly sensitive to how AI is presented and experienced, and are therefore likely to fluctuate across contexts, even when underlying beliefs about system capability remain relatively stable. Based on this theoretical framework, trust is expected to play a key role in competence-related evaluations, predicting perceived intelligence both directly and indirectly through the cognitive and affective attitudes it shapes. These attitudes, in turn, are expected to predict perceived intelligence in their own right: the two hypotheses below therefore describe successive links in a single chain (trust → attitudes → perceived intelligence) rather than competing claims, and the attitude-to-intelligence link stated in H2 is the same pathway through which moral framing is later hypothesized to operate (see H7).

H1. Trust in AI will positively predict competence-based evaluations, particularly perceived intelligence, both directly and indirectly through cognitive and affective attitudes.

H2. Positive cognitive and affective attitudes will, in turn, positively predict perceived intelligence.

2.2 Military AI, moral framing, and ethical implications

The mere labeling of content as AI-generated can alter user judgments, regardless of the actual quality of the output [42]. This phenomenon aligns with broader research on algorithm aversion, where users lose trust in systems after observing errors [13, 41], as well as algorithm appreciation, which highlights a preference for algorithmic judgment in certain contexts [7, 33]. Collectively, this literature shows that AI evaluations are strongly influenced by contextual and interpretative factors.

These dynamics become particularly salient in the military domain, where public acceptance of AI varies depending on mission

type, with higher support for logistical or reconnaissance tasks than for combat operations [46]. Furthermore, perceptions differ depending on the nature of the task, with a distinction between skills perceived as "human" versus "mechanical" [27], and are shaped by expectations of reliability and human control [32].

Beyond contextual variation, military AI raises significant ethical dilemmas related to autonomy and the delegation of lethal decision-making, challenging the notions of meaningful human control and moral agency [23, 26]. The opacity of advanced systems further amplifies concerns about accountability [38], contributing to ambivalent public attitudes. Indeed, both quantitative and qualitative research reveal a persistent tension between perceived benefits, such as increased precision and reduced human casualties, and concerns about loss of human control and moral disengagement [21, 22].

Individuals' attitudes toward military AI may thus depend on whether it is perceived as defensive, protective, or harmful [25], highlighting the central role of moral framing. Taken together, this literature suggests that framing, particularly whether AI is presented as defensive or offensive, systematically shapes individuals' evaluations and attitudes. For the purpose of our study, "offensive" and "defensive" denote operational-technical functions (striking targets vs. intercepting threats), independently of the moral or legal legitimacy such actions may have in a specific context. These are the extremes of a continuum along which a multitude of intermediate or shifting functions could be identified, but which are not considered here in order to isolate the phenomenon under analysis more neatly. Crucially, no moral connotation was embedded in the two descriptions, one related to the production of offensive AI, the other to the production of defensive AI (see section 4.3), that were provided to participants as experimental material and stimuli. Rather, our aim was to test whether moral and affective evaluations would arise spontaneously from this minimal operational framing alone. For this reason, no information about actors or specific conflicts was included in the two scenarios. This was meant to prevent participants from associating them with ongoing real-world conflicts or a specific political orientation, and responding on that basis.

H3. AI framed as defensive will elicit more positive attitudinal evaluations than AI framed as offensive.

H4. AI framed as defensive will be perceived as more intelligent than AI framed as offensive.

2.3 Individual differences: personality and global identity

While HCI literature has predominantly focused on system-level and contextual factors, psychological research highlights the crucial role of individual differences. Personality traits within the Big Five Model have been associated with variations in trust and the acceptance of novel technologies [12, 18, 36]. For instance, openness to experience facilitates the adoption of innovation, whereas neuroticism heightens sensitivity to risk and uncertainty. Conscientiousness, linked to norm adherence, appears particularly relevant in evaluating technologies with ethical implications.

Beyond personality, Identification With All Humanity (IWAH) offers an additional lens for interpreting the social impact of AI. This construct reflects a sense of global connection toward all human beings [37], with higher levels of IWAH associated with greater concern for human rights and collective welfare. Recent studies suggest that interaction with AI may also reinforce this shared identity: by acting as a perceived "outgroup," AI can increase the salience of humanity as a common "ingroup" [51].

In morally sensitive domains such as the military, these individual differences become particularly relevant. Individuals with higher IWAH may react more negatively to AI applications perceived as threatening to human welfare, while showing greater support for uses perceived as protective. Taken together, these findings suggest that personality traits and identity orientations interact with moral framing (offensive vs. defensive) to shape systematically different perception patterns.

H5. Personality traits associated with prosocial orientation will be associated with less favourable evaluations of military AI.

H6. Identification with all humanity (IWAH) will be associated with less favourable affective evaluations of military AI.

2.4 Halo effects and moral evaluation of AI

Evaluations of AI systems are not purely analytical but are strongly influenced by global affective impressions. According to the halo effect, general evaluations of a target can bias judgments of its specific attributes, including competence and intelligence [39]. This implies that users' impressions of a technological system (whether positive or negative) can shape broader perceptions of system characteristics beyond objective performance.

Recent research has shown that the halo effect plays a significant role in shaping trust in AI and related systems [28]. Importantly, halo-based judgments can operate in both positive and negative directions. General affective impressions, whether favorable or unfavorable, can bias evaluations of specific attributes, leading to inflated or deflated judgments [39]. This process is consistent with the affect heuristic, whereby positive or negative evaluations extend across multiple dimensions of judgment [45], and with affect infusion processes, through which affective states influence cognitive evaluations [15]. Accordingly, negative framing or associations may produce generalized negative evaluations, affecting not only perceived trustworthiness but also other attributes such as intelligence.

Moreover, when technologies are less familiar or poorly understood, individuals tend to rely more on heuristic processing and global impressions rather than detailed knowledge [8, 49]. Under these conditions, halo effects are more likely to emerge, leading general impressions to bias evaluations of specific attributes [39]. This suggests that in emerging or complex domains, such as military AI, users may be particularly prone to such biases.

These findings suggest that affective evaluations may influence competence-based judgments through halo-like processes.

H7. The effect of AI framing on perceived intelligence will be mediated by attitudinal evaluations.

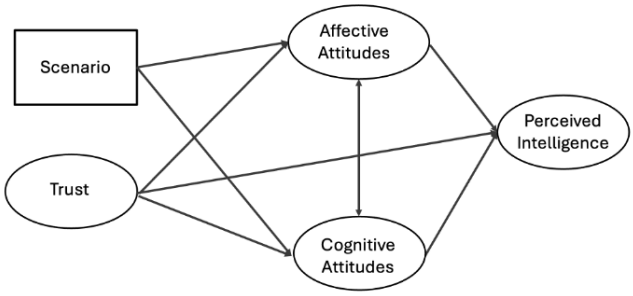


Figure 1: Hypothesized structural model illustrating the relationships between Scenario, Trust in AI, Cognitive and Affective Attitudes, and Perceived Intelligence. The possible role of personality factors is not represented in this model for the sake of parsimony.

Together, these hypotheses reflect a model in which AI framing shapes attitudinal evaluations, which in turn influence competence-based judgments. Trust further modulates these effects. Individuals' differences in personality (and IWAH) covariate with attitudes.

3 The Study

Based on the hypotheses formulated above, the relationships between the variables under consideration are represented in the model shown in Figure 1. The independent variable, Scenario, is manipulated through the description of an AI system designed for either offensive or defensive military applications. This manipulation is expected to influence both cognitive and affective attitudes toward the AI system. In line with the proposed model, Trust in AI is also expected to act as a dispositional predictor of both attitudinal components. These attitudinal responses are in turn hypothesized to predict perceived intelligence, capturing competence-related evaluations of the system. Finally, cognitive and affective attitudes are expected to be positively associated, reflecting their shared evaluative basis.

4 Method

4.1 Participants

A total of 401 Italian-speaking adults participated in the study (52.6% females, 47.4% males). Participants were recruited through convenience sampling via university networks and social media platforms. Eligibility criteria required participants to be native or proficient Italian speakers aged between 18 and 65 years (M = 29.77 years; SD = 12.30). No financial compensation was provided for participation.

4.2 Procedure

The study employed a between-subjects experimental design in which participants were randomly assigned to one of two conditions. Each condition presented a scenario describing the development of a military artificial intelligence (AI) system framed either as defensive (i.e., designed to protect against threats) or offensive (i.e., designed to conduct attacks).

Data were collected via an online self-report questionnaire administered through a survey platform (Google Forms). After providing informed consent, participants completed a brief sociodemographic section including age, gender, educational level, and occupational status. Participants then completed two scales to assess baseline differences in AI familiarity and Trust prior to the experimental manipulation. Following exposure to one of the two experimental scenarios, participants completed measures assessing cognitive and affective attitudes (likeability) toward the AI system, and perceived intelligence. The questionnaire concluded with measures assessing personality traits and identification with all humanity.

The study was approved by the Ethical Research Committee of the University of Siena (CAREUS 3/2026/0093244).

4.3 Scenarios

The two scenarios were equivalent in structure, length, and wording, differing only in elements conveying the offensive or defensive nature of the AI system. The scenarios are reported below:

"According to the latest issue of *Advanced Robotics Journal*, a leading technology research agency will launch a program next year for the development of advanced autonomous weapons. The project involves hiring 120 experts to create an Artificial Intelligence (AI) system specifically designed to coordinate [*defensive / offensive*] weapons. In the military domain, [*defensive / offensive*] weapons refer to armament systems designed to [*protect critical infrastructure and neutralize hostile attacks before they reach their target / destroy critical infrastructure and conduct hostile attacks that reach their target*].

The system analyzes real-time data and, through a sensor network, autonomously coordinates drone swarms to [*intercept and neutralize incoming threats / destroy identified targets*] with millimeter precision. The agency estimates that the deployment of this AI, with minimal human intervention, will increase the [*effectiveness of territorial protection / destructive capacity of operations*] by 35% compared to current systems."

The scenarios were framed as a news report announcing a system still under development, rather than as an account of a system already in use. This choice was deliberate, for three reasons. First, presenting the technology prospectively, before any operational track record, ensured that evaluations, including perceived intelligence, could not be anchored to observed performance or outcomes, thereby isolating the effect of moral framing on otherwise outcome-neutral information. Second, the announcement format, introduced as an item from a technology journal, reflects the way emerging military AI is typically encountered by the public, that is, through media coverage of research programs rather than through direct experience. Third, a development scenario kept the two conditions fully equivalent and free of situational outcomes (e.g., casualties or mission results), consistent with the deliberately decontextualized framing described above.

4.4 Measures and materials

4.4.1 Artificial Intelligence Literacy. AI literacy was assessed using the Artificial Intelligence Literacy Scale (AILS) [2], a 16-item tool comprising three subscales: AI Knowledge and Usage Competence

(items 1–6), Societal Impacts of AI (items 7–11), and Ethical Awareness in AI (items 12–16). Responses were recorded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). No reverse-coded items are included. This measure was administered prior to the experimental manipulation and included as a control variable in the analysis.

4.4.2 Trust in AI. Trust in the AI system was assessed using a 6-item scale adapted from Choudhury and Shamszare [9]. The scale measures the extent to which participants believe that the AI system can provide reliable information and operate without causing harm. Responses were collected on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater perceived trustworthiness.

4.4.3 Cognitive Attitudes Toward Military AI. Cognitive attitudes toward military AI were assessed using five items developed by the research team, as existing standardised measures were not suitable for capturing this specific domain. The items assessed perceptions of safety, the ethical acceptability of delegating decisions to AI, willingness to support the national adoption of the system under development, trust in the system's autonomous decision-making, and overall moral acceptability.

Responses were collected on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), with higher scores indicating more favourable attitudes. To evaluate the internal structure of the scale, we performed an Exploratory Factor Analysis (EFA) using the ML extraction method and the Promax rotation. Results support a monofactorial solution: the extracted single factor has an eigenvalue of 3.21 and explains 64.3% of the total variance. All items show appropriate loadings on the factor (range: .755 – .841). Fit indices indicate a good model fit to the data (RMSEA = .055 (CI 90%: .005–.100); SRMR = .016; CFI = .995; TLI = .989; BIC = −18.806) supporting the adequacy of the single-factor solution. Given the exploratory nature of the scale, further validation in independent samples is warranted.

4.4.4 Godspeed Questionnaire. Perceptions of the AI system were assessed using a subset of the Godspeed Questionnaire [4]. Perceived Intelligence (GSQ_Perc_Int) was assessed using the 5-item subscale. Affective attitudes (GSQ_Likeability) were evaluated using 3 items of the Likeability subscale. Items were presented as semantic differential scales on a 5-point continuum between opposing adjective pairs.

4.4.5 Personality Traits. Personality traits were assessed using the BFI-10 [43], a 10-item short version of the Big Five Inventory measuring Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (two items per dimension). Responses were recorded on a 5-point Likert scale. Reverse-scored items were coded according to the original scoring procedure. Personality traits were examined as individual-difference predictors of cognitive and affective attitudes.

4.4.6 Identification With All Humanity. Global human identification was assessed using a 9-item version of the Identification With All Humanity (IWAH) scale [37]. The first eight items were rated on 5-point Likert scales with construct-specific anchors (e.g., perceived closeness, sense of shared identity, concern for others, global responsibility). The final item required participants to select one of five diagrams depicting varying degrees of overlap between the self and humanity as a whole. This measure evaluates a person's inclination to adopt an inclusive social identity that encompasses humanity at large rather than being confined to local, national, or cultural affiliations. By combining affective and cognitive aspects of identification, it reflects the extent to which individuals perceive commonality and connectedness with people worldwide. In the analysis, this variable was treated as an individual characteristic potentially influencing both cognitive judgments and affective responses.

4.5 Data analysis

All analysis were conducted using Positron (Posit Software, PBC; Version: 2026.03) on a MacBook Pro. Descriptive statistics were computed to summarize sample characteristics (gender, age) and to assess group equivalence (gender, age, AI literacy, and trust in AI). Means, standard deviations, and frequency distributions were calculated for all demographic and psychological variables. Before conducting the main analysis, we assessed the equivalence of the experimental groups (defensive vs. offensive condition). A chi-square test was used for categorical variables (gender), and independent-samples Welch's t-tests were conducted for continuous variables (age, AI literacy, and trust in AI). Descriptive analysis were used to compute means, standard deviations, and frequency distributions for all demographic and psychological variables. Main hypotheses were tested using Welch's t-tests, mediation analysis (estimated within the structural equation model using the delta method), and structural equation modelling (SEM). Given non-normal distributions of the study variables, as indicated by the Shapiro–Wilk test, robust maximum likelihood estimation (MLR) was used in the SEM and moderation analyses.

Cronbach's alpha (α) was calculated to assess scale reliability. The AILS scale demonstrated very good reliability (total α = .91; AI Knowledge and Usage Competence α = .91; Societal Impacts of AI α = .79; and Ethical Awareness in AI α = .85). The Trust in AI scale also showed very good reliability (α = .82). The Cognitive Attitude towards military AI scale showed very good reliability (α = .87). The GSQ indicated very good reliability for Affective attitude (Likeability) (α = .87) and Perceived Intelligence (α = .89). The IWAH scale had very good reliability (α = .84). The reliability of the BFI-10 subscales was assessed using inter-item correlations, which are considered more appropriate for brief instruments than Cronbach's alpha, as recommended by the original authors [43]. The following scores were obtained: Extraversion (r = .26), Agreeableness (r = .11), Conscientiousness (r = .04), Neuroticism (r = .20), Openness (r = .22). These results were consistent with expectations for brief instruments, since the BFI-10 subscales are mostly designed for brief assessment rather than internal consistency [43]. The low score for Conscientiousness (r = .04) was examined further; however, item distribution and a scatterplot ruled out encoding issues: the two items (3 reverse and 8) appear to capture different dimensions, which limits the homogeneity of Conscientiousness, a point already noted by Rammstedt and John [43].

Table 1: Descriptive Statistics and Between-Condition Comparisons.

	Defence (n = 197)		Offence (n = 204)				
Variable	M	SD	M	SD	t	df	p
Cognitive Attitude towards AI	2.28	0.86	1.96	0.83	3.82	399	<.001
Affective Attitude	2.52	0.89	2.12	0.92	4.41	400	<.001
Perceived Intelligence	2.86	0.77	2.58	0.82	3.44	400	<.001
Trust in AI	2.79	0.71	2.69	0.70	1.39	399	0.166
IWAH	3.41	0.80	3.39	0.71	0.34	391	0.734
AILS total	61.75	12.20	62.65	11.68	−0.76	398	0.448
Age	29.23	11.33	30.26	13.16	−0.85	394	0.397

Affective Attitude refers to the GSQ Likeability variable, and Perceived Intelligence refers to the GSQ Perceived Intelligence variable.

5 Results

5.1 Preliminary analysis

Prior to the main analysis, we assessed the equivalence of the two experimental conditions (Defence vs. Offence) on all relevant background variables, using Welch's t tests for continuous variables and a chi-square test for gender. No significant difference emerged in gender distribution (Defence: 52.8% female; Offence: 52.5% female; $\chi^2(1) = 0.005$, p = .946). Age did not differ significantly between conditions (Defence: M = 29.23, SD = 11.33; Offence: M = 30.26, SD = 13.16; t(394) = −0.85, p = .397). Likewise, no significant between-condition differences were found for AI literacy (AILS total: t(398) = −0.76, p = .448) or Trust in AI (t(399) = 1.39, p = .166), suggesting successful randomisation. Because Welch's correction was applied and some measures contained missing responses, degrees of freedom vary across tests. Descriptive statistics for all variables by condition are reported in Table 1.

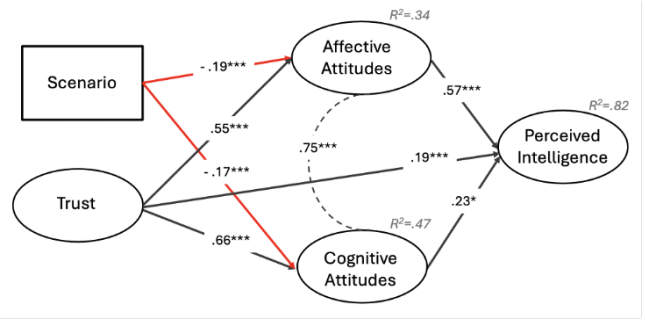


Figure 2: The SEM representation shows the influence of Trust in AI and Scenario manipulation on Affective Attitudes and Cognitive Attitudes, with these attitudes serving as mediators, and on Perceived Intelligence. Significance levels at .05 are indicated by "*", and those at .001 by "*".**

5.2 Structural Equation Modelling

A structural equation model was estimated using maximum likelihood with robust standard errors (MLR estimator) in lavaan [44]. The model included four latent constructs: Affective Attitudes (GSQ_Likeability), Cognitive Attitudes (Attitude_AI), Perceived Intelligence (GSQ_Perc_Int), and Trust in AI (Trust_AI). The scenario was included as an observed exogenous variable (0 = Defence, 1 = Offence). Trust in AI, measured prior to the experimental manipulation, was treated as a stable individual-level predictor. Three residual covariances were freed based on theoretical overlap between item pairs: GSQ_1 ~~ GSQ_2, GSQ_2 ~~ GSQ_4, and AI_TRUST_1 ~~ AI_TRUST_3. The model demonstrated good fit to the data: robust CFI = .976, robust TLI = .971, robust RMSEA = .040 [90% CI: .030, .050]; SRMR = .039; AIC = 17723.34; BIC = 17995.10. All factor loadings were statistically significant (all p < .001), with standardised values ranging from .49 to .91, indicating adequate to strong indicator reliability. Explained variance was substantial for Perceived Intelligence (R^2 = .82) and moderate for Cognitive Attitudes (R^2 = .47) and Affective Attitudes (R^2 = .34). The model is shown in Figure 2.

Exposure to the offensive scenario significantly reduced both affective attitudes toward the AI system (β = −.19, p < .001) and cognitive attitudes (β = −.17, p < .001) compared to the defensive scenario. Trust in AI was a strong positive predictor of Affective attitudes (β = .55, p < .001), Cognitive attitudes (β = .66, p < .001) and Perceived intelligence (β = .19, p< .001). Both Affective attitudes (β = .57, p < .001) and Cognitive attitudes (β = .23, p < .05) significantly predicted perceived intelligence, which showed the highest proportion of explained variance in the model (R^2 = .82). The residual covariance between affective and cognitive attitudes was significant (r = .75, p < .001), indicating shared variance beyond their common predictors. R^2 values for the remaining endogenous variables were .34 for Affective Attitudes and .47 for Cognitive Attitudes. Indirect effects were estimated within the SEM framework using the delta method and are reported as standardised coefficients (β). The offensive Scenario exerted a significant negative indirect effect on Perceived intelligence via Affective attitudes (β = −.11, SE = .042, z = −3.57, p = .001) and a marginally significant negative effect via Cognitive attitudes (β = −.04, SE = .028, z = −1.89, p = .058). The total indirect effect of Scenario on perceived intelligence was significant (β = −.15, SE = .048, z = −4.31, p < .001), indicating that the impact of Scenario type on Perceived intelligence was primarily mediated by the two attitudinal pathways. Trust in AI exerted significant positive indirect effects on Perceived Intelligence via Affective Attitudes (β = .31, SE = .086, z = 5.07, p < .001) and via Cognitive Attitudes (β = .15, SE = .098, z = 2.18, p < .05), with a total indirect effect (β = .47, SE = .101, z = 6.45, p < .001) and a direct effect (β = .19, SE = .08, z = 3.38, p = .001). These findings are

consistent with an affective halo mechanism, whereby the moral-operational framing of an AI system shapes affective responses that in turn bias judgments of its intelligence, independently of its actual capabilities.

5.3 Moderation analysis

To explore whether Trust in AI or IWAH moderated the effect of the experimental scenario on attitudinal responses, four additional moderation models were estimated for cognitive and affective attitudes with mean-centered continuous predictors. None of the interaction terms reached significance: Scenario × Trust in AI on Cognitive Attitudes ($\beta = -.042$, p = .463), Scenario × IWAH on Cognitive Attitudes ($\beta = -.015$, p = .819), Scenario × Trust in AI on Affective Attitudes ($\beta = .040$, p = .512), and Scenario × IWAH on Affective Attitudes ($\beta = -.058$, p = .374). Simple slopes analysis confirmed that the effects of Trust in AI and IWAH on attitudinal responses were consistent across both experimental conditions. These results indicate that the impact of scenario type on AI evaluations did not differ across individuals differing in prior trust and tendencies to identify with all humanity, suggesting that contextual framing affects attitudes independently of stable individual dispositions.

5.4 Supplementary Regression analysis

To explore whether personality traits and identification with all humanity contributed to attitudinal responses beyond the primary predictors established in the SEM, three supplementary regression models were estimated. Big Five personality traits and IWAH were not included in the structural model to preserve parsimony, given the exploratory nature of the hypothesised mediation structure. Trust in AI and Scenario were retained as covariates to control for their established effects. Cognitive Attitudes, Affective Attitudes, and Perceived Intelligence were entered as dependent variables, whereas IWAH and the Big Five personality traits were included as predictors. These models were estimated to explore direct associations between individual-level predictors and key outcome variables, complementing the SEM analysis.

Across all three models, Trust in AI emerged as the strongest predictor (Cognitive Attitudes: $\beta = .554$, p < .001; Affective Attitudes: $\beta = .444$, p < .001; Perceived Intelligence: $\beta = .545$, p < .001), while Scenario retained a significant negative effect in all models (Cognitive Attitudes: $\beta = -.150$, p < .001; Affective Attitudes: $\beta = -.188$, p < .001; Perceived Intelligence: $\beta = -.129$, p = .002). IWAH was a significant negative predictor only of Affective Attitudes ($\beta = -.124$, p = .023), but not of Cognitive Attitudes (p = .123) or Perceived Intelligence (p = .834), suggesting that identification with all humanity is more associated with affective than cognitive evaluations. None of the Big Five personality traits was a significant predictor in any model (all p-values > .24). The three models explained 35%, 26%, and 33% of variance in Cognitive Attitudes, Affective Attitudes, and Perceived Intelligence, respectively.

6 Discussion

The present study examined how offensive versus defensive framing shapes the evaluation of military AI systems, focusing on the roles of prior trust, attitudinal responses, and individual differences. Overall, the findings were largely consistent with the proposed hypotheses and provide a coherent account of how moral framing influences perceived AI competence through attitudinal pathways.

Consistent with H1, Trust in AI, measured prior to the experimental manipulation, emerged as the strongest individual-level predictor across models, influencing competence-related evaluations through both direct and indirect pathways. This effect operated independently of scenario framing: higher trust increased evaluations regardless of condition, while offensive framing reduced them irrespective of prior trust. This pattern supports the characterization of trust as a relatively stable cognitive anchor [17, 29] that shapes AI evaluations independently of situational factors. In line with H2, perceived intelligence was jointly predicted by affective and cognitive attitudes, with a stronger contribution from the affective component, highlighting the central role of emotional responses in competence judgments. Trust exerted both direct and indirect effects on Perceived Intelligence, with the latter mediated by attitudinal pathways, indicating that perceived intelligence reflects the integration of dispositional trust and context-dependent evaluations.

Consistent with H3 and H4, participants exposed to the offensive scenario reported more negative affective and cognitive attitudes, as well as lower perceived intelligence. These findings extend prior work on military AI evaluation [32, 46] by showing that the offensive–defensive distinction constitutes a meaningful moral boundary. Defensive framing, associated with protection and legitimacy, appears to enhance perceived moral acceptability and, in turn, overall evaluations. Notably, framing and trust operated independently and additively, with no evidence of moderation.

Contrary to H5, Big Five personality traits did not predict attitudinal responses, suggesting that in high-stakes, morally sensitive contexts, evaluations are driven more by situational and domain-specific factors than by general personality dispositions. Because personality traits were examined only as direct predictors and not as moderators of the experimental manipulation in this study, whether they shape sensitivity to moral framing specifically remains an open question for future research. Similarly, H6 was only partially supported: IWAH negatively predicted affective attitudes, but not cognitive attitudes or perceived intelligence, and did not moderate framing effects. This pattern suggests that global human identification primarily influences the emotional dimension of evaluation without affecting competence judgments. The dissociation between affective and cognitive pathways is consistent with dual-process accounts of moral judgment [19], indicating that IWAH shapes baseline affective responses rather than amplifying reactions to offensive framing.

The mediation analysis were consistent with H7, indicating that the effect of framing on perceived intelligence was primarily mediated via affective attitudes, with weaker evidence for cognitive mediation. This pattern aligns with a halo-like process [39], whereby the moral valence of the use case shapes a general evaluative impression that extends to competence judgments. The findings extend halo effect research to military AI, showing that both affective and cognitive components contribute to this process. They are also consistent with the affect infusion model [15] and the affect heuristic [45], whereby negative affect associated with offensive use generalizes to judgments of system capability.

7 Theoretical and Practical Implications

From a practical standpoint, the findings highlight the role of communication strategies in shaping public perceptions of military AI systems. Framing an AI system as defensive rather than offensive significantly improves its perceived likeability, moral acceptability, and competence, with these effects operating through attitudinal pathways. This suggests that public acceptance may be shaped less by objective system characteristics than by their communicative framing. These findings raise concerns about the strategic use of moral framing in public communication. Emphasizing defensive purposes may not simply reflect functional differences, but may systematically bias evaluation. This interpretation is consistent with research on framing effects, which shows that even minimal changes in presentation can alter judgments and preferences [47, 50].

From this perspective, defensive narratives may resemble mechanisms associated with deceptive patterns, design strategies that steer decisions by exploiting cognitive biases rather than supporting fully informed choice [6, 34]. The term "deceptive patterns" [6] captures the core mechanism more precisely: deliberately leading users toward choices they would not otherwise have made. It should be noted, however, that deceptive patterns are typically defined in relation to behavioral outcomes, getting users to do, or refrain from doing, something, whereas the dynamics observed here operate at the level of opinion and attitude formation rather than immediate action. In this sense, our findings extend the logic of deceptive patterns from interface-level behavioral steering to the framing of high-stakes technologies in public communication, where morally acceptable aspects are emphasized and ethically problematic ones are downplayed.

Although communicative intent was not directly assessed, the findings show that minimal differences in description produce systematic shifts in perceived acceptability and competence, suggesting that framing can function as a powerful persuasive device. This underscores the need for greater transparency and ethical scrutiny in how military AI systems are presented to the public. Communication strategies that emphasize protective functions may increase acceptance, but may also obscure morally relevant aspects of the technology. Framing should therefore not be considered a neutral process, but a key factor shaping societal attitudes and perceived legitimacy.

8 Limitations and Future Direction

Several limitations should be acknowledged.

First, the sample consisted predominantly of young Italian adults recruited through convenience sampling. This limits generalizability along several dimensions that are particularly consequential for the present topic. Italy is a country without compulsory military service and without recent direct experience of armed conflict on its territory, so participants' baseline familiarity with, and exposure to, military matters is likely to be low. Attitudes toward military AI may differ markedly in populations with higher military familiarity, for instance, in countries with conscription or near-universal military service, or in those currently or recently exposed to war, where such technologies may be perceived as more necessary and normalized or, conversely, as more immediately threatening. Moreover, because the salience of armed conflict in public discourse fluctuates over time, evaluations of military AI may themselves be sensitive to the geopolitical context at the moment of data collection. Cross-national and cross-cultural replication, including populations differing in military exposure and conscription regimes and ideally repeated over time, would help establish the robustness and boundary conditions of the present findings.

Second, the cross-sectional single-exposure design restricts causal inference beyond the experimental manipulation; longitudinal research is needed to examine how trust and attitudes evolve over time.

Third, although the SEM showed excellent fit, the substantial residual covariance between affective and cognitive attitudes (r = .75) suggests shared variance not fully captured by the model, potentially reflecting a higher-order evaluative factor.

In addition, the study focused exclusively on the offensive or defensive distinction; future research should consider other morally relevant dimensions, such as levels of human oversight, target type, or international legality, to develop a more comprehensive model of military AI evaluation. The deliberately decontextualized framing, while intended to isolate the offensive–defensive contrast, also limits generalizability to real geopolitical settings, where moral evaluation is shaped by additional cues such as the identity of the actors involved. Relatedly, future research could move beyond the binary offensive–defensive contrast by introducing conditions in which a single system is described as serving both functions, or as switching between them depending on the situation. Such dual-use framings would more closely reflect the ambivalent nature of many real military AI systems and could reveal whether more nuanced or conflicting evaluations emerge when protective and harmful potentials are simultaneously salient.

A further limitation concerns the experimental manipulation. No manipulation check was included, preventing direct verification that participants interpreted the AI system as intended. The wording of the scenarios may itself have carried evaluative connotations independent of the manipulation. Although the expression "hostile attacks" appeared in both conditions, its referent differed: in the defensive scenario it denoted incoming threats neutralized by the system, whereas in the offensive scenario it denoted actions carried out by the system itself. This asymmetry in attribution may have contributed to a more negative evaluation of the offensive system beyond its operational purpose. Future studies should systematically vary such connotative cues and include manipulation checks, so as to disentangle the effect of the offensive–defensive distinction from that of specific lexical framing. Moreover, although framing effects were statistically significant, their magnitude was modest relative to the influence of prior trust, suggesting that dispositional factors may play a more prominent role than situational framing alone. Finally, the use of a brief written vignette may have reduced ecological validity and emotional engagement. In particular, although the scenario was framed as a journalistic report, a single text-based vignette only partially reproduces the way public perceptions are actually shaped through repeated and multimodal media exposure. Future studies should employ more immersive and multimodal stimuli (e.g., videos or simulations) to strengthen the

realism and psychological impact of the manipulation, in line with construal level theory [48].

References

[1] Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar, and Hananeh Alambeigi. 2024. Trust in AI: Progress, Challenges, and Future Directions. *Humanities and Social Sciences Communications* 11, 1 (Nov. 2024), 1568. doi:10.1057/s41599-024-04044-8

[2] Sayım Aktay, Seçkin Gök, and Özgenur Kanata. 2026. Artificial Intelligence Literacy Scale. *ESA Pub* (Feb. 2026), 10. doi:10.71410/ESApub.ITEJ.2026.10-01.02

[3] Sulfikar Amir, Sabrina Ching Yuen Luk, Shrestha Saha, Iuna Tsyrulneva, and Marcus T. L. Teo. 2025. Measuring Social Trust in AI: How Institutions Shape the Usage Intention of AI-Based Technologies. *Human Behavior and Emerging Technologies* 2025, 1 (Jan. 2025), 4084384. doi:10.1155/hbe2/4084384

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. doi:10.1007/s12369-008-0001-3

[5] Filip Bialy, Mark Elliot, and Robert Meckin. 2025. Perceptions of AI Across Sectors: A Comparative Review of Public Attitudes. doi:10.48550/ARXIV.2509.18233

[6] Harry Brignull. 2023. *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You* (first edition, updated march 27th, 2024 ed.). Testimonium Ltd, Eastbourne.

[7] Cindy Candrian and Anne Scherer. 2022. Rise of the Machines: Delegating Decisions to Autonomous AI. *Computers in Human Behavior* 134 (Sept. 2022), 107308. doi:10.1016/j.chb.2022.107308

[8] Shelly Chaiken and Yaacov Trope. 1999. *Process Theories in Social Psychology*. Guilford Press.

[9] Avishek Choudhury and Hamid Shamszare. 2023. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *Journal of Medical Internet Research* 25 (June 2023), e47184. doi:10.2196/47184

[10] Sarah J. Daly, Anna Wiewiora, and Greg Hearn. 2025. Shifting Attitudes and Trust in AI: Influences on Organizational AI Adoption. *Technological Forecasting and Social Change* 215 (June 2025), 124108. doi:10.1016/j.techfore.2025.124108

[11] Karl De Fine Licht and Jenny De Fine Licht. 2020. Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy. *AI & SOCIETY* 35, 4 (Dec. 2020), 917–926. doi:10.1007/s00146-020-00960-w

[12] Sarv Devaraj, Robert F. Easley, and J. Michael Crant. 2008. Research Note — How Does Personality Matter? Relating the Five-Factor Model to Technology Acceptance and Use. *Information Systems Research* 19, 1 (March 2008), 93–105. doi:10.1287/isre.1070.0153

[13] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114–126. doi:10.1037/xge0000033

[14] Debora Firmino De Souza, Sonia Sousa, Kadri Kristjuhan-Ling, Olga Dunajeva, Mare Roosileht, Avar Pentel, Mati Mõttus, Mustafa Can Özdemir, and Žanna Gratšjova. 2025. Trust and Trustworthiness from Human-Centered Perspective in Human–Robot Interaction (HRI)—A Systematic Literature Review. *Electronics* 14, 8 (April 2025), 1557. doi:10.3390/electronics14081557

[15] Joseph P. Forgas. 1995. Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin* 117, 1 (1995), 39–66. doi:10.1037/0033-2909.117.1.39

[16] David Gefen, Elena Karahanna, and Detmar W. Straub. 2003. Trust and TAM in Online Shopping: An Integrated Model1. *MIS Quarterly* 27, 1 (March 2003), 51–90. doi:10.2307/30036519

[17] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals* 14, 2 (July 2020), 627–660. doi:10.5465/annals.2018.0057

[18] Simone Grassini, Sebastian Oltedal Thorp, Alexander Sævild Ree, Aleksandra Sevic, and Enrico Cipriani. 2025. Distinct Predictors of Positive Attitudes toward Artificial Intelligence and General Technology: Big Five Traits, Gender, and Age. *Behaviour & Information Technology* (Dec. 2025), 1–14. doi:10.1080/0144929X.2025.2598623

[19] Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293, 5537 (Sept. 2001), 2105–2108. doi:10.1126/science.1062872

[20] Stefano Guidi, Lorenzo Reina, Francesco Currò, Enrico Cipriani, Simone Grassini, and Oronzo Parlangeli. 2026. Ecce Histor_IA. The Role of Narratives, Usage Contexts, and Tasks in Influencing Attitudes towards Artificial Intelligence. *Behaviour & Information Technology* (Jan. 2026), 1–13. doi:10.1080/0144929X.2025.2606832

[21] Lee Hadlington, Jens Binder, Sarah Gardner, Maria Karanika-Murray, and Sarah Knight. 2023. The Use of Artificial Intelligence in a Military Context: Development of the Attitudes toward AI in Defense (AAID) Scale. *Frontiers in Psychology* 14 (May 2023), 1164810. doi:10.3389/fpsyg.2023.1164810

[22] Lee Hadlington, Maria Karanika-Murray, Jane Slater, Jens Binder, Sarah Gardner, and Sarah Knight. 2025. Public Perceptions of the Use of Artificial Intelligence in Defence: A Qualitative Exploration. *AI & SOCIETY* 40, 2 (Feb. 2025), 277–290. doi:10.1007/s00146-024-01871-w

[23] Kirsi Marjaana Helkala, George Lucas, Edward Barrett, and Henrik Syse. 2023. Editorial: Ethical Challenges in AI-enhanced Military Operations. *Frontiers in Big Data* 6 (June 2023), 1229252. doi:10.3389/fdata.2023.1229252

[24] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, 3 (May 2015), 407–434. doi:10.1177/0018720814547570

[25] Michael C. Horowitz and Sarah Maxey. 2026. Morally Opposed? A Theory of Public Attitudes and Emerging Military Technologies. *Conflict Management and Peace Science* 43, 1 (Jan. 2026), 58–80. doi:10.1177/07388942251320027

[26] James Johnson. 2026. Can AI Behave Ethically during Military Crises? Preserving Human Moral Agency. *International Affairs* 102, 1 (Jan. 2026), 63–83. doi:10.1093/ia/iiaf191

[27] Uwe Klein, Jana Depping, Laura Wohlfahrt, and Pantaleon Fassbender. 2024. Application of Artificial Intelligence: Risk Perception and Trust in the Work Context with Different Impact Levels and Task Types. *AI & SOCIETY* 39, 5 (Oct. 2024), 2445–2456. doi:10.1007/s00146-023-01699-w

[28] Jakub Kraciuk, Elżbieta Małgorzata Kacperska, and Marcin Idzik. 2026. The Halo Effect as a Factor Influencing Consumer Trust in Innovative Technological Solutions. *Sustainability* 18, 6 (March 2026), 2984. doi:10.3390/su18062984

[29] J. D. Lee and K. A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (Jan. 2004), 50–80. doi:10.1518/hfes.46.1.50_30392

[30] Min Kyung Lee. 2018. Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data & Society* 5, 1 (Jan. 2018), 2053951718756684. doi:10.1177/2053951718756684

[31] Fan Li, Xiaotian Wei, Chenyang Wang, Chenyue Zhang, Guoming Yu, Ya Yang, and Yuhan Liu. 2025. How Do Users Perceive AI? A Dual-Process Perspective on Enhancement and Replacement. *Telematics and Informatics Reports* 19 (Sept. 2025), 100227. doi:10.1016/j.teler.2025.100227

[32] Eleri Lillemäe, Kairi Talves, and Wolfgang Wagner. 2025. Public Perception of Military AI in the Context of Techno-Optimistic Society. *AI & SOCIETY* 40, 2 (Feb. 2025), 929–943. doi:10.1007/s00146-023-01785-z

[33] 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

[34] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–32. doi:10.1145/3359183

[35] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (July 1995), 709. jstor:258792 doi:10.2307/258792

[36] Robert R. McCrae and Paul T. Costa. 1997. Personality Trait Structure as a Human Universal. *American Psychologist* 52, 5 (May 1997), 509–516. doi:10.1037/0003-066X.52.5.509

[37] Sam McFarland, Matthew Webb, and Derek Brown. 2012. All Humanity Is My Ingroup: A Measure and Studies of Identification with All Humanity. *Journal of Personality and Social Psychology* 103, 5 (2012), 830–853. doi:10.1037/a0028724

[38] Forrest E Morgan, Benjamin Boudreaux, Andrew J Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman. 2020. Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World. (2020).

[39] Richard E. Nisbett and Timothy D. Wilson. 1977. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology* 35, 4 (April 1977), 250–256. doi:10.1037/0022-3514.35.4.250

[40] Oronzo Parlangeli, Maria Cristina Caratozzolo, and Stefano Guidi. 2014. Multitasking and Mentalizing Machines: How the Workload Can Have Influence on the System Comprehension. In *Engineering Psychology and Cognitive Ergonomics*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, and Don Harris (Eds.). Vol. 8532. Springer International Publishing, Cham, 50–58. doi:10.1007/978-3-319-07515-0_6

[41] Oronzo Parlangeli, Francesco Curro', Paola Palmitesta, and Stefano Guidi. 2024. Moral Judgements of Errors by AI Systems and Humans in Civil and Criminal Law. *Behaviour & Information Technology* 43, 9 (July 2024), 1718–1728. doi:10.1080/0144929X.2023.2283622

[42] Irene Rae. 2024. The Effects of Perceived AI Use On Content Perceptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. doi:10.1145/3613904.3642076

[43] Beatrice Rammstedt and Oliver P. John. 2007. Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41, 1 (Feb. 2007), 203–212. doi:10.

1016/j.jrp.2006.02.001
[44] Yves Rosseel. 2012. Lavaan : An *R* Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 2 (2012). doi:10.18637/jss.v048.i02
[45] Paul Slovic, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor. 2007. The Affect Heuristic. *European Journal of Operational Research* 177, 3 (March 2007), 1333–1352. doi:10.1016/j.ejor.2005.04.006
[46] Peter Szikora. 2025. Public Perceptions and Trust Dimensions of Military Autonomous Vehicles in the Age of Artificial Intelligence. *The Eurasia Proceedings of Educational and Social Sciences* 44 (Sept. 2025), 70–81. doi:10.55549/epess.937
[47] Richard H. Thaler and Cass R. Sunstein. 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. *Yale University Press* 8, 1 (2008). doi:10.1016/S1477-3880(15)30073-6
[48] Yaacov Trope and Nira Liberman. 2010. "Construal-level Theory of Psychological Distance": Correction to Trope and Liberman (2010). *Psychological Review* 117, 3 (July 2010), 1024–1024. doi:10.1037/a0020319
[49] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty. *Science* 185, 4157 (Sept. 1974), 1124–1131. doi:10.1126/science.185.4157.1124
[50] Amos Tversky and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211, 4481 (Jan. 1981), 453–458. doi:10.1126/science.7455683
[51] Congyu Wang and Kaiping Peng. 2023. AI Experience Predicts Identification with Humankind. *Behavioral Sciences* 13, 2 (Jan. 2023), 89. doi:10.3390/bs13020089