American Journal of Human Psychology (AJHP)

ISSN: 2994-8878 (ONLINE)

VOLUME 4 ISSUE 3 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Human Psychology (AJHP)

Volume 4 Issue 3, Year 2026
ISSN: 2994-8878 (Online)
DOI: https://doi.org/10.54536/ajhp.v4i3.8098
https://journals.e-palli.com/home/index.php/ajhp

Perceptions of AI-Assisted Case Assessment Among Social Work Professionals: A Quantitative Survey on Trust, Efficiency, and Ethical Concerns

Sora Pazer[1*]

Article Information

Received: August 09, 2026

Accepted: June 27, 2026

Published: September 14, 2026

Keywords

AI-Assisted Assessment, Algorithmic Bias, Digital Transformation, Ethical Governance, Large Language Models, Professional Trust, Social Work

ABSTRACT

The progressive digitalization of human services administration has generated urgent questions about professional trust, algorithmic governance, and the appropriate institutional boundaries of automated decision support. Social work constitutes a particularly consequential arena for examining how practitioners evaluate AI-assisted tools, given the high-stakes relational and ethical demands that characterize the profession. This study examined the predictors of professional trust in AI-assisted case assessment among social work professionals, with a specific focus on the roles of prior AI experience, perceived efficiency, perceived quality, and ethical and bias concerns. A cross-sectional quantitative survey was administered to N = 127 social work professionals using validated Likert-format scales. An independent-samples t-test revealed that professionals with high prior AI experience demonstrated substantially greater trust than those with limited experience, $t(125) = 8.24$, $p < .001$, Cohen's $d = 1.45$. Multiple regression analysis identified perceived efficiency ($\beta = .31$, $p < .001$, 95% CI [.17, .45]), AI experience ($\beta = .29$, $p < .001$, 95% CI [.15, .43]), and perceived assessment quality ($\beta = .27$, $p < .001$, 95% CI [.14, .40]) as significant positive predictors of trust, while ethical concerns exerted a significant negative effect ($\beta = -.18$, $p = .014$, 95% CI [−.32, −.04]). Perceived bias risk did not reach significance in the multivariate model ($\beta = -.12$, $p = .081$). The overall model explained 58% of the variance in trust ($R^2 = .58$, $F(5,121) = 33.44$, $p < .001$). Social work professionals conditionally endorse Large Language Models as decision-support instruments, contingent on transparency, ethical governance, and structured professional training. The findings offer an empirically grounded basis for designing responsible AI integration frameworks in social services.

INTRODUCTION

It is 4:47 on a Tuesday afternoon when a child protection caseworker in Frankfurt opens the eighty-third case file of the month. The client is a single mother of three; the allegations are serious; the deadline for submitting an assessment recommendation to the family court is seven hours away. The practitioner has seventeen years of professional experience, has completed every mandated continuing education module offered by her agency, and still feels the familiar weight of decisional uncertainty descend as she begins to read. What if she misses a pattern? What if the risk configuration embedded in this file resembles one she has never encountered before? Into this constellation of professional vulnerability, temporal pressure, and structural under-resourcing, the question of artificial intelligence enters - not as a utopian promise, but as a practical proposition: could a Large Language Model assist in organizing the information, identifying contradictory data points, or flagging risk indicators that human fatigue might allow to pass unnoticed?

The digitalization of welfare state administration has accelerated dramatically over the past decade, driven by intersecting pressures of efficiency, documentation demand, and the expanding sophistication of machine learning systems. In Germany alone, social service agencies employ over 430,000 full-time equivalent professionals, many of whom report chronic work overload and systemic documentation burdens that divert attentional resources from direct client contact (Harlow, 2003; Munro, 2011). European and North American social work associations have begun to engage formally with questions of AI governance, issuing position statements that acknowledge both the transformative potential and the profound ethical risks of automated decision support in social services (International Federation of Social Workers [IFSW], 2014). The structural conditions of social work - high caseloads, inadequate staffing ratios, and legally consequential decisions made under conditions of epistemic uncertainty - render the field not merely an intellectually interesting test case for AI integration, but a morally urgent one.

Social work professionals occupy a structurally unique position in the welfare apparatus. Unlike administrative officers, they are required by both training and professional ethics to exercise relational judgment, holistic assessment, and advocacy that cannot be reduced to rule-based procedures. Unlike physicians, they frequently lack the institutional protection, peer consultation structures, and systematic documentation supports that buffer individual decision-making under high-stakes conditions (Webb, 2006). This dual vulnerability - to organizational pressure from above and relational demand from clients - creates a distinctive configuration of risks and needs when AI enters the professional environment. Understanding how social workers perceive, trust, and negotiate AI-assisted tools is therefore not merely a question of technology

[1] IU Internationale Hochschule, Germany
[*] Corresponding author's e-mail: Sorapazer@gmail.com

adoption dynamics but a question of professional identity, epistemic autonomy, and the institutional conditions of ethical practice.

The empirical literature on AI in social services has expanded rapidly since approximately 2015, with particular attention to predictive risk modeling in child welfare (Keddell, 2019; Gillingham, 2019), algorithmic case management tools in unemployment administration (Eubanks, 2018), and, most recently, the deployment of Large Language Models as documentation and assessment support systems (Brown *et al.*, 2020; OpenAI, 2023). A consistent finding across this body of literature is that practitioner trust in AI systems is neither uniform nor static: it varies systematically with prior technology experience, perceived transparency, and the degree to which AI is framed as augmenting rather than replacing professional judgment (Broadhurst *et al.*, 2010; Poudel & Maharjan, 2025; Zerilli *et al.*, 2019). Surveys of social work and allied health professionals reveal persistent ambivalence — strong appreciation for efficiency gains coupled with deep concern about algorithmic bias, accountability structures, and the potential erosion of professional discretion (Asif *et al.*, 2025).

Despite this accumulating empirical base, significant theoretical and methodological gaps persist. First, the majority of existing studies have examined general attitudes toward AI rather than investigating the specific predictors of trust in AI-assisted case assessment as a structurally distinct cognitive and professional act. The mechanisms through which AI experience, perceived quality, and ethical concern interact to shape professional trust remain underspecified in the extant literature. Second, social work research has paid insufficient attention to the differential contributions of perceived efficiency versus perceived assessment quality as theoretically separable constructs within technology acceptance models - conflating these dimensions in ways that obscure their independent predictive paths. Third, empirical investigations employing validated quantitative instruments with social work professionals exclusively are rare; most large-scale surveys have recruited heterogeneous professional samples that limit disciplinary specificity and contextual interpretation. This study directly addresses all three gaps.

The central research question guiding this investigation is: What factors predict professional trust in AI-assisted case assessment among social work professionals, and what are the respective roles of perceived efficiency, perceived quality, prior AI experience, and ethical and bias concerns in shaping this trust? The practical stakes of answering this question are immediate and substantial. As organizations increasingly pilot AI tools in social service contexts, and as commercial vendors market Large Language Model-based documentation systems with growing aggressiveness, empirically grounded frameworks for understanding and supporting appropriate professional trust calibration are urgently needed. Without such frameworks, implementation decisions will continue to be driven by vendor narratives rather than practitioner knowledge, with potentially serious consequences for the populations that social work is mandated to protect.

LITERATURE REVIEW

Digital Transformation in the Social Work Profession

The integration of digital technologies into social work practice has followed a trajectory markedly different from that observed in medicine, law, or financial services. Where those adjacent professions encountered digitalization primarily through electronic record-keeping and diagnostic decision-support tools deployed within stable institutional structures, social work has confronted it simultaneously through administrative management platforms, case management software, predictive risk algorithms, and now conversational AI systems (Gillingham, 2019; Webb, 2006). This heterogeneity of digital penetration reflects the structural heterogeneity of social work itself: a profession spanning child protection, homelessness services, addiction counseling, disability support, and numerous adjacent practice fields, each with its own legislative frameworks, organizational cultures, and relational obligations that mediate how technology is received and embedded.

Theoretically, the digitalization of social work can be analyzed through the lens of institutional logics (Benbya *et al.*, 2020). When algorithmic tools enter organizational environments that have historically valued relational discretion, holistic assessment, and client-centered advocacy, they introduce competing logics of efficiency, standardization, and quantifiable accountability. The tension between these logics is not merely administrative; it is constitutive of professional identity under transformation. A social worker who is required to input client data into a predictive risk scoring system is simultaneously a professional exercising judgment and an operator executing a computational procedure - and these roles are not always reconcilable within a single professional self-understanding (Munro, 2011). The normative dimension of this tension is precisely what elevates AI integration in social work above the level of a technical administrative problem.

Recent developments in Large Language Models have introduced a qualitatively new dimension to this longstanding tension between professional judgment and algorithmic process. Unlike earlier rule-based decision support systems that were structurally legible as tools, LLMs can process unstructured natural language, generate contextually sensitive recommendations, and produce outputs that superficially resemble the nuanced prose of an experienced practitioner (Brown *et al.*, 2020; OpenAI, 2023). This mimetic capacity creates a specific psychological and professional challenge: it becomes progressively more difficult to maintain the experiential boundary between tool-use and professional judgment precisely when that boundary is most normatively significant. The epistemological implications of this blurring for professional training, supervision, and accountability are only beginning to receive systematic theoretical attention.

Large Language Models as Decision-Support Instruments

Large Language Models represent a fundamental architectural shift in artificial intelligence, moving from narrow task-specific systems to general-purpose language engines capable of synthesizing, generating, and contextualizing information across domains with remarkable flexibility (LeCun *et al.*, 2015; Brown *et al.*, 2020). Built on transformer architectures trained on vast corpora of human-generated text, LLMs such as GPT-4, Gemini, and Claude can be prompted to summarize case documentation, identify thematic patterns in unstructured data, generate risk hypotheses from complex case histories, and produce structured assessment outputs in seconds. This technical capability has generated substantial enthusiasm among health and social services innovators, with pilot applications already implemented in psychiatric triage, social needs screening, and automated report generation (Topol, 2019).

The decision-support potential of LLMs in social work is theoretically significant across several dimensions. First, LLMs can reduce the cognitive load associated with documentation by synthesizing complex case histories into structured summaries, potentially freeing practitioner attention for relational engagement and direct client contact (Kleinberg *et al.*, 2018). Second, LLMs can function as a second-reviewer operating without the fatigue, confirmation bias, or availability heuristics that compromise human assessment under temporal pressure - systematically flagging contradictory or incomplete information in case materials that a fatigued practitioner might overlook. Third, by processing larger quantities of comparative case data than any individual practitioner could realistically accumulate through personal practice, LLMs can potentially identify risk configurations that would not be salient to a professional with a limited or highly specialized caseload. These capabilities position LLMs not as replacements for professional judgment but as amplifiers of it, operating in a human-AI collaborative frame.

However, the very properties that generate LLMs' decision-support potential simultaneously generate risks that have been carefully documented in adjacent literatures. The tendency of LLMs to produce fluent, plausible-sounding outputs regardless of their factual accuracy - the so-called hallucination problem - creates a specific danger in consequential assessment contexts where erroneous outputs may be mistaken for expert insight by practitioners lacking the technical literacy to recognize their limitations (Mittelstadt *et al.*, 2016). More structurally, LLMs trained on historical case data encode the systemic biases present in that data: if prior child protection decisions were shaped by racial, class-based, or gender-related bias, an LLM trained on those decisions will reproduce and potentially amplify such bias at scale (Noble, 2018; Barocas *et al.*, 2019). The practical consequence is a technology that could simultaneously improve efficiency and institutionalize structural injustice - a combination that poses acute ethical dilemmas for a profession whose core mandate is the pursuit of social justice.

Professional Trust and Technology Acceptance

The construct of professional trust in technology has been theorized from multiple disciplinary perspectives, most influentially through the Technology Acceptance Model (TAM; Davis, 1989) and its subsequent extensions. TAM posits that perceived usefulness and perceived ease of use are the proximate determinants of technology adoption intention, with external variables - including training, prior experience, and organizational context - exerting their influence on adoption through these two mediating constructs. In the context of AI-assisted professional tools, TAM's perceived usefulness dimension maps directly onto the constructs of efficiency improvement and quality enhancement examined in the present study, providing a theoretically grounded rationale for treating these perceptions as central predictors of professional trust and acceptance (Venkatesh *et al.*, 2003).

However, TAM was originally developed in organizational computing contexts and has been criticized for its limited attention to the normative and relational dimensions of technology acceptance in value-laden professional environments. Social work practice is governed not only by technical rationality but by ethical frameworks, professional standards, relational commitments, and institutional accountability structures that create distinctive criteria for evaluating technology beyond mere efficiency (Webb, 2006). A social worker does not simply ask whether a tool helps her complete tasks more quickly; she asks whether it supports or undermines the relational, ethical, and epistemological foundations of professional practice. This normative layer of acceptance - which might be termed professional legitimation - represents a dimension of technology acceptance that standard TAM does not formally model and that the present study was specifically designed to examine through the inclusion of ethical concern and bias perception as predictor variables. Recent empirical work in adjacent service domains reinforces this multidimensional conception of acceptance: in AI-driven service settings, trust has been shown to be strongly context-dependent, with users extending confidence to automated systems for routine, efficiency-critical tasks while reserving reliance on human judgment for emotionally complex or ethically sensitive interactions (Asif *et al.*, 2025). Complementary adoption research indicates that acceptance is conditioned less by technical capability alone than by prior user experience, equitable access, and the perceived ethical soundness of implementation (Poudel & Maharjan, 2025).

Automation complacency - the well-documented tendency to accept automated outputs as authoritative without applying systematic independent verification (Cummings, 2004) - constitutes one of the most practically significant risks associated with AI integration in consequential professional practice. When practitioners place excessive

trust in AI-generated assessments without applying their own critical scrutiny, errors in the system's outputs can be amplified rather than corrected by professional oversight, with potentially serious consequences for clients. This risk is heightened in social work precisely because client populations are frequently vulnerable, underrepresented in historical training data, and least able to contest algorithmic conclusions that negatively affect their access to services or their legal standing. Managing appropriate trust calibration - neither dismissive rejection nor uncritical acceptance - thus becomes a central practical and ethical challenge for professional education, clinical supervision, and organizational governance in any social service agency that deploys AI tools.

Research Hypotheses

Drawing on the Technology Acceptance Model, social learning theory, automation complacency research, and the normative literature on professional AI ethics, five hypotheses were derived to guide the present investigation. H1 predicted that social work professionals with greater prior experience using AI tools would demonstrate significantly higher trust in AI-assisted case assessment than those with limited prior AI experience, consistent with social learning and habituation mechanisms. H2 predicted that perceived efficiency of AI-assisted assessment would be a significant positive predictor of professional trust, consistent with TAM's perceived usefulness pathway. H3 predicted that perceived quality of AI-assisted case assessment would be a significant positive predictor of professional trust, operationalizing a distinct dimension of usefulness not captured by efficiency alone. H4 predicted that perceived ethical concerns about AI-assisted assessment would be a significant negative predictor of professional trust, reflecting the normative legitimation process theorized above. H5 predicted that perceived risk of algorithmic bias would be a significant negative predictor of professional trust, reflecting practitioners' sensitivity to the justice implications of AI use in social services contexts.

MATERIALS AND METHODS

Study Design

This study employed a cross-sectional quantitative survey design, implemented as a fully anonymous self-administered online questionnaire. The cross-sectional design was selected for its pragmatic suitability to capturing professional attitude distributions across a geographically dispersed target population within a constrained research timeline, while the quantitative survey format enabled the systematic collection of interval-level data appropriate for the correlational and regression analyses specified by the hypotheses (Podsakoff *et al.*, 2003). The study is descriptive-correlational in scope: while it identifies predictive relationships between AI experience, professional perceptions, and trust outcomes, the cross-sectional design precludes causal inference, a limitation explicitly acknowledged in the Discussion. All design decisions were guided by principles of voluntary participation, informational autonomy, and data minimization consistent with current ethical standards for anonymous behavioral survey research.

Sample and Recruitment

The final analytical sample comprised N = 127 social work professionals. Participants were recruited through professional networks, social work associations, and academic institutions in German-speaking countries, supplemented by targeted outreach through social work practitioner communities on social media platforms. The mean professional experience of the sample was M = 11.2 years (SD = 7.8), indicating a relatively experienced professional cohort with substantial practice exposure across diverse service contexts. Prior experience with AI tools was moderate on average (M = 3.42, SD = 1.08 on the five-point response scale), indicating that the sample encompassed practitioners with widely varying levels of AI familiarity and thus offered adequate attitudinal variance for the comparative and predictive analyses specified in the hypotheses. Participation was voluntary, anonymous, and uncompensated. Participants were informed prior to beginning the survey that their responses would be used for research purposes, that they could withdraw at any time without consequence, and that submission of a completed survey constituted informed consent. The estimated completion time was approximately twelve to fifteen minutes.

Measures

Trust in AI-assisted case assessment (TRUST) was measured using a four-item scale reflecting participants' confidence in the accuracy, reliability, contextual appropriateness, and professional applicability of AI-generated case assessment outputs. All items were rated on a five-point Likert scale anchored at 1 (Strongly disagree) and 5 (Strongly agree). An acceptance scale was formed from TRUST, perceived efficiency (EFF), perceived assessment quality (QUALITY), and future adoption intention (FUTURE), yielding four four-item components; this combined scale demonstrated excellent internal consistency (α = .88). Perceived efficiency (EFF) captured the degree to which participants believed AI-assisted assessment could reduce their workload, accelerate documentation, and improve time management in practice. Perceived quality (QUALITY) assessed participants' beliefs about whether AI tools could improve the accuracy, comprehensiveness, and consistency of assessment outputs relative to unaided professional judgment.

Risk perception was operationalized through a parallel four-item scale comprising concern about algorithmic bias (BIAS), ethical issues associated with AI use in social services (ETHICS), requirements for transparency in AI decision processes (TRANSP), and commitment to the primacy of human professional judgment (HUMAN). This risk perception scale demonstrated equally strong

internal consistency (α = .91). Future adoption intention (FUTURE) assessed the degree to which participants anticipated AI-assisted assessment becoming a standard feature of social work practice and whether they would be willing to use such tools in their own professional context. Prior AI experience (AI_USE) was measured as a single self-report item asking participants to rate their overall experience with AI tools on the same five-point scale. Professional experience (EXP) was measured in full years of employment in social work or closely adjacent professional roles. Training need (TRAIN) assessed participants' perceived necessity of specialized professional training prior to implementing AI tools in practice.

Statistical Analysis

Analyses were conducted in three sequential stages following the structure of the research hypotheses. In the first stage, descriptive statistics - means, standard deviations, and Cronbach's α coefficients - were computed for all scale variables. In the second stage, Pearson product-moment correlations were calculated among the acceptance-relevant variables (TRUST, EFF, QUALITY, FUTURE) and among the risk-relevant variables (BIAS, ETHICS, TRANSP), with two-tailed significance testing applied throughout. In the third stage, an independent-samples t-test was conducted to examine whether trust differed significantly between professionals with high prior AI experience (AI_USE ≥ 4) and those with low prior AI experience (AI_USE ≤ 2), with effect magnitude estimated using Cohen's d. Subsequently, a simultaneous multiple regression analysis was performed with TRUST as the dependent variable and AI_USE, EFF, QUALITY, ETHICS, and BIAS as the five simultaneous predictors, generating standardized regression coefficients (β), significance values, and 95% confidence intervals for each predictor. All analyses were performed in SPSS Version 28. The significance threshold was set at α = .05 for all tests.

RESULTS AND DISCUSSION

Descriptive Statistics

Descriptive analyses revealed a distinctive bifurcation in the attitudinal profile of the sample: participants simultaneously endorsed the beneficial potential of AI-assisted assessment and expressed pronounced concern about its risks. Within the acceptance domain, perceived efficiency received the second highest mean rating (M = 4.38, SD = 0.61), indicating near-consensus agreement that AI could improve efficiency in case assessment workflows. Perceived assessment quality was rated somewhat more moderately (M = 3.91, SD = 0.72), reflecting greater uncertainty about the quality implications of AI involvement. Trust in AI-assisted assessment was moderate overall (M = 3.46, SD = 0.86), exhibiting the most substantial variability of the acceptance variables - a pattern consistent with the heterogeneity of AI experience in the sample. Future adoption intention was similarly moderate (M = 3.74, SD = 0.81).

By contrast, risk-related items received uniformly high ratings that reflected a strong collective professional concern. Concern about algorithmic bias (M = 4.42, SD = 0.58) and ethical concerns more broadly (M = 4.35, SD = 0.66) were both endorsed well above the scale midpoint. The necessity of AI transparency (M = 4.79, SD = 0.42) and the primacy of human professional judgment (M = 4.83, SD = 0.39) received the two highest mean ratings of any variable in the study, with strikingly low variability indicating near-universal consensus among participants that human practitioners must retain final decision-making authority in case assessment. The training need variable also reflected strong consensus (M = 4.71, SD = 0.48), underscoring practitioners' recognition that responsible AI use requires structured professional preparation. Descriptive statistics for all variables and reliability coefficients for the composite scales are presented in Table 1.

Table 1: Descriptive Statistics and Internal Consistency for All Study Variables

Scale	n	M	SD	α
AI Experience (AI_USE)	1	3.42	1.08	—
Professional Experience (EXP)	1	11.20	7.80	—
Trust in AI (TRUST)	4	3.46	0.86	.88a
Perceived Efficiency (EFF)	4	4.38	0.61	—
Assessment Quality (QUALITY)	4	3.91	0.72	—
Future Adoption (FUTURE)	4	3.74	0.81	—
Algorithmic Bias Risk (BIAS)	4	4.42	0.58	.91b
Ethical Concerns (ETHICS)	4	4.35	0.66	—
Transparency (TRANSP)	4	4.79	0.42	—
Human Judgment (HUMAN)	4	4.83	0.39	—
Training Need (TRAIN)	4	4.71	0.48	—

Note. N = 127. n = number of items per scale or subscale. M = arithmetic mean; SD = standard deviation; a = Cronbach's alpha coefficient. a Cronbach's a = .88 for the Acceptance Scale comprising Trust (TRUST), Efficiency (EFF), Quality (QUALITY), and Future Adoption (FUTURE). b Cronbach's a = .91 for the Risk Perception Scale comprising Bias Risk (BIAS), Ethical Concerns (ETHICS), Transparency (TRANSP), and Human Judgment (HUMAN). AI_USE and EXP are single-item measures; TRAIN is a single-item measure reported for descriptive completeness.

Correlational Findings

Pearson correlation analyses revealed strong and statistically significant positive intercorrelations among all four acceptance-domain variables. Trust in AI was significantly associated with perceived efficiency ($r = .69$, $p < .001$), perceived quality ($r = .71$, $p < .001$), and future adoption intention ($r = .65$, $p < .001$). The correlations between efficiency and quality ($r = .63$, $p < .001$), efficiency and future adoption ($r = .58$, $p < .001$), and quality and future adoption ($r = .66$, $p < .001$) were likewise substantial and significant. These results confirm that the acceptance-domain variables constitute an internally coherent attitudinal cluster while also demonstrating that trust, as the primary dependent variable, shares meaningful unique variance with each of its theorized predictors.

Within the risk perception cluster, bias concerns, ethical concerns, and transparency requirements were significantly and positively intercorrelated at comparable magnitudes: bias–ethics ($r = .72$, $p < .001$), bias–transparency ($r = .59$, $p < .001$), and ethics–transparency ($r = .68$, $p < .001$). This pattern confirms that the three risk variables tap a shared underlying risk perception disposition while remaining distinguishable as separate construct facets - a finding consistent with the theoretical framing advanced in Section 2 and with the two-scale measurement structure (acceptance and risk) employed in the present study. Full bivariate correlation matrices for both domains are presented in Table 2.

Table 2: Pearson Correlation Matrices for Acceptance and Risk Perception Variables

Panel A: Acceptance Variables				
Variable	**1**	**2**	**3**	**4**
1. Trust (TRUST)	—			
2. Efficiency (EFF)	.69***	—		
3. Quality (QUALITY)	.71***	.63***	—	
4. Future Adoption (FUTURE)	.65***	.58***	.66***	—
Panel B: Risk Perception Variables				
Variable	**1**	**2**	**3**	
1. Bias Risk (BIAS)	—			
2. Ethical Concerns (ETHICS)	.72***	—		
3. Transparency (TRANSP)	.59***	.68***	—	

*Note. N = 127. ** p* < .001. Correlation coefficients are presented for lower off-diagonal elements only. Diagonal elements represent perfect self-correlation.*

Group Comparison and Multiple Regression Analysis

To test H1, an independent-samples t-test was conducted comparing mean trust levels between social work professionals with high prior AI experience (AI_USE ≥ 4; n = 58, M = 3.94, SD = 0.58) and those with low prior AI experience (AI_USE ≤ 2; n = 45, M = 2.91, SD = 0.77). The analysis yielded a statistically significant group difference, $t(125) = 8.24$, $p < .001$, with a large effect size, Cohen's $d = 1.45$. H1 was thus strongly supported: social work professionals with greater prior AI experience demonstrated substantially and significantly higher trust in AI-assisted case assessment, with the between-group difference exceeding conventional benchmarks for a large effect by a substantial margin.

A simultaneous multiple regression analysis was then conducted to examine the independent predictive contributions of all five theorized predictors to trust in AI-assisted assessment (H2 through H5). The overall regression model was statistically significant, $F(5, 121) = 33.44$, $p < .001$, and explained 58% of the variance in trust ($R^2 = .58$), indicating excellent model fit for a social science survey study. Examining individual predictors, perceived efficiency emerged as the strongest positive predictor of trust ($\beta = .31$, $p < .001$, 95% CI [.17, .45]), supporting H2. AI experience constituted the second strongest positive predictor ($\beta = .29$, $p < .001$, 95% CI [.15, .43]), replicating and extending the group comparison finding within a multivariate framework that controls for the remaining predictors. Perceived assessment quality was the third significant positive predictor ($\beta = .27$, $p < .001$, 95% CI [.14, .40]), supporting H3. Ethical concerns exerted a significant and negative effect on trust ($\beta = -.18$, $p = .014$, 95% CI [−.32, −.04]), supporting H4. Perceived bias risk did not attain statistical significance in the multivariate model ($\beta = -.12$, $p = .081$, 95% CI [−.26, .02]), resulting in the non-confirmation of H5. Regression results are presented in full in Table 3.

In summary: H1 (AI experience → trust) was strongly confirmed both in the group comparison ($d = 1.45$) and in the multivariate regression ($\beta = .29$). H2 (efficiency → trust) was confirmed, with efficiency emerging as the single strongest regression predictor. H3 (quality → trust) was confirmed, with quality contributing significant unique variance beyond efficiency and experience. H4 (ethical concerns → reduced trust) was confirmed, with ethical concerns exerting a significant negative effect. H5 (bias risk → reduced trust) was not confirmed in the multivariate model; bias risk's effect on trust appears to operate primarily through its shared variance with

ethical concerns rather than as an independent predictor pathway.

Discussion

Summary and Theoretical Integration

The present study set out to identify the predictors of professional trust in AI-assisted case assessment among social work practitioners. The results yield a theoretically coherent and practically instructive picture. Trust in AI-assisted assessment is shaped by a configuration of facilitating factors - prior AI experience, perceived efficiency, and perceived quality - and counterweighted by ethical concerns that exert a significant independent negative effect. The overall model's explanatory power ($R^2 = .58$) is impressive by social science survey standards and suggests that the five-predictor framework captures a substantial share of the attitudinal variance governing professional trust in this domain.

Table 3: Multiple Regression Analysis Predicting Trust in AI-Assisted Case Assessment

Predictor	β	p	95% CI
Perceived Efficiency (EFF)	.31	< .001	[.17, .45]
AI Experience (AI_USE)	.29	< .001	[.15, .43]
Assessment Quality (QUALITY)	.27	< .001	[.14, .40]
Ethical Concerns (ETHICS)	−.18	.014	[−.32, −.04]
Bias Risk (BIAS)	−.12	.081	[−.26, .02]

Note. N = 127. β = standardized regression coefficient. CI = confidence interval. The overall model: $R^2 = .58$, $F(5, 121) = 33.44$, $p < .001$. Dependent variable: Trust in AI-Assisted Assessment (TRUST). All predictors entered simultaneously. Confidence intervals based on standard errors from ordinary least squares regression.

Mapped onto the Technology Acceptance Model, the present findings both validate and complicate Davis's (1989) original framework. The significant positive effects of perceived efficiency (β = .31) and perceived quality (β = .27) are consistent with TAM's foundational emphasis on perceived usefulness as the primary driver of adoption intention, and they validate the extension of this model to AI-specific professional assessment contexts (Venkatesh *et al.*, 2003). Crucially, however, the significant negative effect of ethical concerns (β = −.18) points to the inadequacy of standard TAM in capturing the normative dimension of technology acceptance in value-laden professional environments. Social work practitioners do not evaluate AI solely on technical utility grounds; they evaluate it against the ethical commitments of their profession, and this normative evaluation carries independent predictive weight that standard TAM does not formally represent. The present findings thus argue for extending TAM with a professional legitimation construct that captures the distinctively normative dimension of technology acceptance in regulated helping professions.

The large effect size in the experience comparison (d = 1.45) warrants specific theoretical attention. The trust difference between high-experience and low-experience AI users substantially exceeds conventional benchmarks for large effects and suggests that AI experience is not a background demographic variable to be controlled but a substantive attitudinal determinant with direct policy implications. This pattern is theoretically consistent with two complementary mechanisms: habituation, through which repeated exposure to AI tools reduces the novelty-induced anxiety and uncertainty that suppress trust; and competence acquisition, through which actual engagement with AI systems develops practitioners' ability to critically evaluate outputs and thus calibrate trust more accurately. The non-significance of bias risk in the multivariate model (β = −.12, p = .081) is theoretically interpretable rather than anomalous: the variance in trust attributable to bias concerns appears to be substantially absorbed by the ethical concerns variable, suggesting that practitioners do not experience algorithmic bias as a conceptually distinct risk category but rather as a concrete instantiation of the broader ethical concern about AI in practice.

Practical Implications

The practical implications of these findings extend across individual, organizational, and systemic levels of social service practice. At the individual level, the finding that AI experience is both the most robust group-level differentiator and the second strongest regression predictor of trust strongly recommends systematic AI training as a prerequisite - not merely a supplement - for responsible AI implementation in social work agencies. This training should be specifically designed for social work practice contexts, addressing not only technical operation but also the recognition of hallucinated outputs, the identification of potential bias in specific client subpopulations, and the development of critical rather than complacent engagement with AI-generated case assessments. The automation complacency risk documented by Cummings (2004) should be explicitly addressed in professional development curricula and clinical supervision protocols as a distinct learning objective.

At the organizational level, the significant negative predictive weight of ethical concerns (β = −.18)

constitutes a clear empirical signal that implementation strategies that fail to directly address practitioners' ethical concerns will encounter structural professional resistance that may undermine the effectiveness of AI tools even when those tools are technically superior. Organizations seeking to introduce AI-assisted assessment tools would benefit substantially from co-design processes that involve frontline practitioners in the selection, customization, governance, and ongoing evaluation of AI systems. The near-universal endorsement of transparency (M = 4.79) and human primacy (M = 4.83) should be translated into concrete design and procurement specifications: social work practitioners require AI systems that explain their recommendations in auditable, human-readable terms, clearly identifying the informational basis and confidence level of each assessment output. Black-box systems that generate recommendations without legible reasoning chains are likely to encounter principled professional resistance from socially conscious practitioners, regardless of their objective predictive performance.

At the systemic and policy level, the findings reinforce calls for regulatory frameworks that specifically address the deployment of AI in high-stakes social services contexts (Shneiderman, 2022; Mittelstadt *et al.*, 2016). The combination of strong support for transparency and near-universal commitment to the primacy of human judgment - with human judgment receiving the highest endorsement in the entire study (M = 4.83) - signals that the social work professional community is not categorically resistant to AI but is deeply and articulately committed to meaningful human oversight. Policy frameworks should build on this commitment rather than circumventing it, by mandating human review of all AI-generated case assessment outputs, establishing clear liability frameworks, and creating institutional mechanisms that locate final professional accountability unambiguously with the human practitioner.

Limitations

The interpretation of the present findings must be qualified by several methodological limitations that bear directly on the validity and generalizability of the conclusions. First and most fundamentally, the cross-sectional survey design precludes any causal inference about the identified relationships. While it is theoretically plausible that accumulated AI experience increases trust through habituation and competence mechanisms, it is equally conceivable that practitioners who are dispositionally more open to new technology are both more likely to seek out AI experience and more likely to express trust in AI tools - a reverse-causal interpretation that cannot be excluded on the basis of the present data. Longitudinal studies tracking practitioners before and after systematic AI training programs are necessary to establish directional causality and to assess the stability of professional trust as an attitudinal construct over time.

Second, the sample was recruited through professional networks, academic institutions, and online practitioner communities, a strategy that almost certainly over-represents social work professionals with above-average digital engagement and familiarity and underrepresents practitioners working in highly traditional, under-resourced, or rural contexts who may hold markedly different views on AI. The generalizability of the present findings to the full population of social work professionals in German-speaking countries, and particularly to international professional communities operating within different welfare-state traditions and regulatory frameworks, is accordingly limited and should not be assumed without replication.

Third, all constructs in the present study were assessed through self-report measures administered within a single questionnaire, creating the structural preconditions for common method bias (Podsakoff *et al.*, 2003). The observed correlations between trust and its predictors may be partially inflated by shared method variance arising from common informant, common measurement context, or social desirability effects. Future research should incorporate behavioral measures of AI engagement - such as actual rates of AI tool consultation, rate of AI output modification, and frequency of supervisory consultation about AI recommendations - to obtain more ecologically valid assessments of the trust construct that triangulate the self-report findings presented here.

Fourth, the present study did not assess potentially important individual-difference moderators of the AI experience–trust relationship. Personality characteristics such as openness to experience, need for cognitive closure, and general technology anxiety are theoretically plausible moderators that may substantially alter the magnitude or direction of the experience–trust relationship in subpopulations not captured by the present sample. Organizational climate variables - including supervisory support for AI experimentation, organizational resource allocation for technology training, and peer norms regarding AI adoption - constitute additional contextual moderators whose omission limits the explanatory completeness and practical applicability of the current model.

Fifth, the measurement instruments used in this study were developed specifically for the present investigation and, while demonstrating excellent internal consistency (α = .88 and α = .91 respectively), have not been subjected to independent confirmatory factor analysis or cross-sample validation. The establishment of factorial validity, measurement invariance across professional subgroups, and test–retest reliability represents a necessary step before these instruments can be recommended for routine use in AI acceptance research or applied organizational assessment contexts.

Future Research Directions

The present findings point toward a productive and urgently needed agenda for future empirical inquiry. Priority should be accorded to longitudinal intervention studies that experimentally manipulate AI training content

and intensity among social work practitioners, examining the downstream effects on trust calibration, critical AI engagement, and downstream assessment quality as measured through case outcome data. Qualitative methodologies - including ethnographic observation of AI tool use in active field practice and phenomenological interviewing of practitioners who categorically resist or negotiate AI integration - would provide the micro-level process insights into professional–AI interaction dynamics that quantitative survey research is structurally unable to access. Cross-national comparative research would clarify whether the attitudinal configuration identified here reflects universal features of professional AI acceptance or whether trust, ethics, and legitimation concerns are inflected by national welfare-state traditions, professional education systems, regulatory environments, and organizational cultures in ways that require differentiated theoretical and policy responses.

Research attention must also turn urgently toward the client side of the professional–AI–client triad. The present study, like the overwhelming majority of existing empirical work in this domain, focuses exclusively on practitioner perceptions. The perspectives of service users - the individuals whose interests constitute the ultimate moral referent of any professional ethical claim in social work - on AI-assisted case assessment remain almost entirely unexamined in the literature. Understanding whether and under what conditions clients experience algorithmic assessment as trustworthy, just, and respectful of their dignity and agency is both ethically necessary and practically essential for designing AI integration frameworks that are genuinely accountable to the populations they are intended to serve.

CONCLUSION

Social work professionals are neither categorically opposed to AI-assisted case assessment nor uncritically receptive to it. They apply a sophisticated, normatively structured set of evaluative criteria - centered on efficiency, quality, ethics, and the primacy of human professional judgment - that mirror the foundational values of a profession mandated to protect and empower society's most vulnerable members. The present findings show that professional trust is not granted automatically but earned conditionally: it rises with prior AI experience, perceived efficiency, and perceived assessment quality, and declines when ethical concerns are salient. When AI tools demonstrably improve efficiency and quality, when their operations are transparent and auditable, and when professional training enables practitioners to engage with them critically rather than compliantly, professional trust follows. The path to responsible AI integration in social services therefore does not run around practitioner concerns but directly through them - through investment in training, participatory co-design, robust ethical governance, and an institutional commitment to the proposition that algorithms serve human beings, and never the reverse.

Statements and Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author had no financial support for the conception, conduct, statistical analysis, or write-up of the study.

Conflict of Interest

The author declares that there are no competing financial or non-financial interests that could be perceived to have influenced the work reported in this article.

Ethics Approval

This study was conducted in accordance with the ethical principles set out in the Declaration of Helsinki (World Medical Association, 2013) and the professional guidelines of the German Psychological Society (DGPs). The study employed a fully anonymous online survey design in which no personal data, sensitive health information, or identifying information were collected at any point. Formal institutional review board (IRB) approval was not required under applicable guidelines, as the study involved fully anonymous behavioral survey data with no clinical intervention, no collection of biometric or sensitive personal data, and no involvement of protected populations or deceptive procedures.

Informed Consent

All participants received complete written information about the purpose of the study, the anonymous and voluntary nature of their participation, their right to withdraw at any point without explanation or consequence, and the planned use of aggregate data for research publication. Participants were informed that submitting a completed survey constituted their informed consent to participation. No participant received compensation for participation.

Data Availability Statement

The anonymized dataset analyzed in this study is available from the corresponding author upon reasonable request. The data are not deposited in a publicly accessible repository due to institutional data protection guidelines; however, the author is committed to sharing aggregate summary data and analysis code with qualified researchers upon direct request.

Use of Generative AI

Generative AI tools were used in the preparation of this manuscript for language editing, structural organization, and reference formatting. All scientific content, theoretical arguments, analytical decisions, result interpretations, and conclusions are the sole intellectual responsibility of the author. The author assumes full accountability for the accuracy and integrity of the work.

REFERENCES

Asif, N. H., Mahmud, M., Rahman, M. W., & Islam, M. B. (2025). The trust paradox in AI-driven customer support: A mixed-methods analysis of human vs. AI trust. *American Journal of Data Science and Artificial Intelligence, 1*(1), 36–45. https://doi.org/10.54536/ajdsai.v1i1.4779

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. https://fairmlbook.org

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). *Artificial intelligence in organizations: Current state and future opportunities*. MIS Quarterly Executive, 19(4), vi–xxi.

Broadhurst, K., Wastell, D., White, S., Hall, C., Peckover, S., Thompson, K., Pithouse, A., & Davey, D. (2010). Performing 'initial assessment': Identifying the latent conditions for error at the front-door of local authority children's services. *British Journal of Social Work, 40*(2), 352–370. https://doi.org/10.1093/bjsw/bcn162

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Agarwal, S. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Cummings, M. L. (2004). Automation bias in intelligent time critical decision support systems. In *AIAA 1st Intelligent Systems Technical Conference* (p. 6313). https://doi.org/10.2514/6.2004-6313

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Statistisches Bundesamt [Destatis]. (2022). *Statistisches Jahrbuch Deutschland 2022*.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Gillingham, P. (2019). Decision-making tools and the development of expertise in child protection practitioners: Are we 'skilling up' or 'deskilling' the workforce? *Child & Family Social Work, 24*(1), 2–10. https://doi.org/10.1111/cfs.12432

Harlow, E. (2003). New managerialism, social service departments and social work practice today. *Practice: Social Work in Action, 15*(2), 29–44. https://doi.org/10.1080/09503150308416917

International Federation of Social Workers. (2014). *Global definition of the social work profession*. https://www.ifsw.org/what-is-social-work/global-definition-of-social-work/

Keddell, E. (2019). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Sciences, 8*(10), 281. https://doi.org/10.3390/socsci8100281

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2007). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Munro, E. (2011). *The Munro review of child protection: Final report - A child-centred system*. Department for Education.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

OpenAI. (2023). *GPT-4 technical report*. arXiv. https://doi.org/10.48550/arXiv.2303.08774

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Poudel, S., & Maharjan, S. (2025). Artificial intelligence and education in Nepal: A mixed-methods study on student adoption and learning outcomes. *American Journal of Data Science and Artificial Intelligence, 1*(2), 18–25. https://doi.org/10.54536/ajdsai.v1i2.4763

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Steyvers, M., & Kumar, A. (2023). Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science, 19*(4), 732–742. https://doi.org/10.1177/17456916231181102

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision-making. *Small Group Research, 40*(3), 271–300. https://doi.org/10.1177/1046496409332928

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified theory. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Wang, Y., & Yang, Y. (2019). Artificial intelligence in educational assessment: Promises and challenges. *Journal of Computer Assisted Learning, 35*(5), 471–475. https://doi.org/10.1111/jcal.12360

Webb, S. A. (2006). *Social work in a risk society: Social and political perspectives*. Palgrave Macmillan.

World Medical Association. (2013). Declaration of

Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology, 32*(4), 661–683. https://doi.org/10.1007/s13347-018-0330-6