NextGen Scientific Publication
American International Journal of Computer Science and Technology

Volume 8 Issue 5, Pg. No. 41-52, AIJCST-V8I5P104, 2026
https://doi.org/10.63282/3117-5481/AIJCST-V8I5P104

Original Article

Calibrating Human Supervision in Generative AI Assisted Face to Face Direct Sales: Effects of Factual Accuracy and Representative Autonomy on Customer Trust and Acquisition Performance

***Krunal More**

Lawrence Technological University, Southfield, USA.

Abstract:

Generative artificial intelligence can assist interpersonal selling, but its acquisition value depends on the governance of factual information and decision authority. This study evaluated the Human-Supervised GenAI Acquisition Framework in a randomized, counterbalanced US direct-sales field experiment comprising 3,000 encounters across five conditions, with 600 encounters per condition. Factual audits and CRM outcomes covered all encounters; usable customer surveys and representative autonomy assessments numbered 2,850 and 2,940. C3 combined mandatory verification with retained representative authority, recording 98.1% factual accuracy, customer trust of 6.06/7, autonomy of 5.96/7, and 40.0% conversion. Human-only conversion was 31.0%, giving a 9.0-percentage-point difference and 29.0% relative increase, OR=1.48 [1.17, 1.88]. Accuracy and autonomy independently predicted trust, β=.350 and β=.230. Sequential and additional autonomy indirect effects were .086 [.055, .121] and .039 [.022, .061]; the reported combined effect was .125 [.091, .164]. The remaining direct effect was nonsignificant, p=.087. C3 exceeded verification-required C2 by 6.0 percentage points in conversion. These results support the combined C3 configuration and the specified statistical pathways. Incomplete implementation records prevent isolation of representative authority as the explanation for the C2–C3 difference; causal mediation additionally depends on identification assumptions.

Keywords:

Generative Artificial Intelligence, Human Supervision, Direct Sales, Factual Accuracy, Representative Autonomy, Customer Trust.

Article History:

Received: 12.08.2026

Revised: 11.09.2026

Accepted: 26.09.2026

Published: 02.10.2026

1. Introduction

Generative artificial intelligence can provide information during face-to-face selling, but its availability does not determine acquisition performance. Verification concerns factual support; decision authority concerns how recommendations are used. Human–AI performance varies across tasks and arrangements [1], while studies of customer support, writing, and knowledge work highlight deployment context [2–4]. These findings motivate examining how assistance is governed within customer encounters.

The automation–augmentation paradox makes human and machine contributions an organizational design question [5]. Marketing frameworks classify AI capabilities [6], collaborative intelligence links them to sales-funnel activities [7], and recent sales research reveals emerging applications [8]. Agentic systems intensify issues of authority [9]. Marketing meta-analytic evidence further supports examining when collaboration adds value [10]. HSGAI-AF evaluates supervision arrangements within this discussion.

Copyright by the Author(s). This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-NC-ND 4.0) License (https://creativecommons.org/licenses/by-sa/4.0/)

The study addresses three questions within live customer acquisition: how supervision configurations compare; whether factual accuracy and representative autonomy have distinct relationships with customer trust; and how these relationships connect with CRM behavior. HSGAI-AF brings audits, surveys, and acquisition outcomes into the same framework. Its contribution concerns the joint evaluation of verification and authority rather than a claim that human oversight is uniformly beneficial. The five-condition comparison examines specific arrangements; it neither spans all supervision levels nor identifies every feature responsible for their differences.

2. Literature Review and Research Framework

2.1. AI Assistance and Factual Reliability

Research on generative AI studies marketing and innovation [11, 12], and accountability issues [13]. Concerns about reliability in direct sales include product, pricing, policy and promotional statements. Mandatory verification introduces a checkpoint before generated advice enters customer communication. H1. Can checking factual support make communicated information more accurate? This procedural rationale does not imply any change in the capabilities of the language model. Thus, AI access and verification definitions are separated.

Human-autonomy teaming and trust reviews inform responses to AI-supported work [14–16]. Representative trust in AI concerns reliance on assistance; customer trust concerns evaluation of the representative and encounter. H3 examines whether independently audited accuracy predicts that interpersonal evaluation. It neither presumes that customers observe verification nor equates customer trust with trust in AI. Audits and customer responses assess distinct constructs.

2.2. Authority and Representative Autonomy

AI-mediated communication separates the production of messages from interpersonal communication [17]. Human-centered AI stresses reliable systems and human engagement [18]; mutual augmentation considers human and machine contributions [19]; research on algorithmic control explores workplace discretion [20]. These views separate informational assistance from the authority to use it. Autonomy refers to the representative's discretion in the interaction and factual accuracy to the information provided.

The research on intuition and AI motivates to study human judgment and assistance [21]. H2 Authority and autonomy issues of AI representatives H4 asks whether autonomy predicts customer trust after accounting for accuracy. Discretion concerns the representative's role; accuracy concerns communicated information. Their joint inclusion tests distinct statistical relationships. The study did not separately identify the behaviors through which representative discretion reached customers.

2.3. Trust and Acquisition Mechanisms

Hallucination surveys define the research problem of factual unreliability [22, 23], semantic-entropy research explores one detection method [24]. These studies motivate auditing, but do not validate this study's system or procedures. Consumer and service research studies value and experience enabled by AI [25–27]. Research on anthropomorphism separates trust from technical capability [28]. Reviews of service encounters and customer relationship management connect deployment with customer relationships [29, 30].

HSGAI-AF links supervision, accuracy, trust, and acquisition, with an additional authority–autonomy–trust pathway. RO1 compares configurations; RO2 examines verification and accuracy; RO3 examines authority and autonomy; RO4 tests their independent trust relationships; RO5 connects the pathways with CRM outcomes. H5 integrates these relationships. Configuration comparisons evaluate arrangements; mediation decomposes relationships under model assumptions. Randomized assignment alone does not identify causal mediator effects [34, 35].

2.4. Research Hypotheses

- H1. GenAI assistance requiring mandatory human verification produces higher factual accuracy than assistance without mandatory verification.
- H2. Greater GenAI decision authority reduces representative autonomy relative to configurations retaining representative control.
- H3. Factual accuracy positively predicts customer trust after adjustment for the modeled covariates.
- H4. Representative autonomy positively predicts customer trust when factual accuracy is included in the same model.

- H5. Supervision relates to acquisition through the sequential accuracy–trust pathway, with an additional autonomy-related trust pathway.

3. Methodology

3.1. Study Setting and Analytical Sample

The study evaluated face-to-face customer acquisition in US direct sales using a randomized, counterbalanced field design. The analytical sample contained 3,000 encounters, with 600 in each condition, C0–C4. All encounters contributed factual-audit and CRM outcomes. Encounters were the reported observational unit; the number of encounters does not establish the number of independent representatives or territories. Table 2 summarizes sample composition and completeness.

3.2. Experimental Configurations and Intervention Delivery

C0 involved human-only selling. C1 supplied advisory GenAI without mandatory verification. C2 required factual verification before customer-facing use. C3 combined verification with full representative decision authority; C4 assigned greater authority to GenAI and reduced representative discretion. Table 1 preserves these five configurations. Both C2 and C3 required verification, so the comparison between C2 and C3 was central to the question of authority. However, the descriptions do not set out C2's accept, modify, reject or override permissions, or if other delivery features matched C3. Their difference is consequently interpreted at the configuration level.

Table 1. Experimental configurations

Condition	Genai Assistance	Verification	Representative Authority
C0	None	Not applicable	Human-only selling
C1	Advisory	Not mandatory	May use recommendations
C2	Available	Mandatory	Not further specified
C3	Human-supervised	Mandatory	Full final authority
C4	Higher AI autonomy	Not specified	Reduced discretion

Table 2. Sample composition and completeness

Dimension	Categories	Encounter N	Percent
Condition	C0, C1, C2, C3, C4	600 each	20.0 each
Complexity	Low / moderate / high	1,050 / 1,200 / 750	35.0 / 40.0 / 25.0
Familiarity	Low / moderate / high	1,260 / 1,050 / 690	42.0 / 35.0 / 23.0
Experience	<2 / 2–5 / >5 years	990 / 1,260 / 750	33.0 / 42.0 / 25.0
Offer	Standard / promotional	2,130 / 870	71.0 / 29.0
CRM and audit	Available	3,000 each	100.0
Customer survey	Usable	2,850	95.0
Representative survey	Usable	2,940	98.0

3.3. Assignment and Repeated Exposure

The design incorporated randomized allocation and counterbalancing. Maximum baseline standardized mean difference was .07, indicating limited imbalance on the baseline measures. Multilevel analyses addressed representative and territory clustering. These features cannot establish allocation implementation or demonstrate that carryover and contamination were excluded. The reporting requirements are the assignment unit, condition sequences, training, adherence, deviations and handling of repeated exposure as outlined in Section 5.3.

3.4. Measures and Factual Auditing

Independent factual audits assessed product, pricing, policy, and promotional information. Accuracy was expressed as a proportion; factual-error rate and error severity were reported separately. Customer trust and representative autonomy used seven-point scales, as did competence, transparency, interaction quality, and personalization. CRM measures comprised conversion, qualified

interest, and follow-up acceptance. These behavioral outcomes were analyzed separately. The given descriptions do not refer to their observation windows nor create a sequential acquisition funnel. Covariates included representative experience, familiarity, complexity, territory, period, and offer type. Item wording, scoring, audit denominators, and operational definitions remain unreported.

3.5. Sample Completeness and Analysis Population

Usable customer surveys numbered 2,850 (95.0%), and representative autonomy assessments numbered 2,940 (98.0%). These are assessment totals rather than counts of unique representatives. Audit and CRM coverage remained 3,000. Structural estimation used full-information maximum likelihood; the complete-case sensitivity analysis contained 2,796 encounters. Little's χ^2 (42) =49.1, p=.211 did not reject the tested missingness null. This result does not establish unbiased estimation. Marginal response totals do not reconstruct the joint missingness pattern or identify exclusions, condition-specific response rates, and model-specific populations.

3.6. Statistical Models and Sensitivity Analyses

Condition means and CRM counts described observed performance. Mixed-effects models estimated H1 using C2+C3 versus C1 and H2 using C4 versus C3. H3 and H4 examined accuracy and autonomy simultaneously as trust predictors, adjusting for condition, experience, complexity, familiarity, territory, period, and offer type. Logistic conversion modeling estimated the trust–acquisition relationship; multilevel structural modeling and bootstrapped mediation evaluated H5. The four principal C3 conversion comparisons formed the reported Holm correction family. Count-based odds ratios, adjusted contrast coefficients, standardized trust relationships, and indirect effects are reported separately. Measurement and structural fit indices describe the fitted models [31, 32].

Ablations removed framework pathways and reported AIC, BIC, and explanatory measures. Sensitivity specifications varied missing-data treatment, clustering, territory adjustment, complexity, and factual-error definition. Common-method influences [33], mediator identification [34], and covariate roles [35] informed interpretation. Reproduction requires the original equations, distributions, links, estimators, random-effects structures, software, model-specific samples, and bootstrap settings.

3.7. Ethics and Research Transparency

The supplied study records did not establish the ethics approval or exemption position, consent procedures, privacy safeguards, funding, conflicts of interest, registration status and timing, or data/code access conditions. These disclosures remain incomplete. The manuscript therefore makes no claim of approval, exemption, consent, prospective registration, unrestricted data availability, or absence of conflicts. Verified author disclosures are required to complete this section before submission.

4. Results and Findings

4.1. Descriptive Profile

Overall factual accuracy was 94.7%, with a 5.3% factual-error rate. Mean autonomy was 5.64 and mean trust was 5.48. CRM conversion occurred in 966 encounters (32.2%); qualified interest occurred in 1,668, and follow-up acceptance in 1,290. Table 3 presents descriptive statistics. Behavioral and audit outcomes used the complete encounter sample, while survey constructs used their stated usable-response denominators.

Table 3. Descriptive Statistics

Variable	N	Mean	SD	Median	Range	IQR
Factual accuracy	3,000	0.947	0.109	0.982	0.420–1.000	0.070
Factual-error rate	3,000	0.053	0.109	0.018	0.000–0.580	0.070
Error severity, 0–3	3,000	0.18	0.43	0.00	0.00–3.00	0.00
Autonomy, 1–7	2,940	5.64	1.08	5.80	1.40–7.00	1.50
Customer trust, 1–7	2,850	5.48	1.03	5.60	1.00–7.00	1.50
Competence, 1–7	2,850	5.72	0.92	5.80	1.40–7.00	1.20
Transparency, 1–7	2,850	5.58	0.96	5.70	1.20–7.00	1.30
Interaction quality, 1–7	2,850	5.66	0.94	5.80	1.30–7.00	1.30
Personalization, 1–7	2,850	5.49	1.00	5.60	1.00–7.00	1.40
CRM conversion	3,000	0.322	0.467	0	0–1	1

Qualified customer interest	3,000	0.556	0.497	1	0–1	1
Follow-up acceptance	3,000	0.430	0.495	0	0–1	1

Accuracy had a median of .982, IQR=.070, and range .420–1.000; the median error rate was .018. Error severity averaged .18 on the zero-to-three scale. Secondary customer evaluations averaged 5.72 for competence, 5.58 for transparency, 5.66 for interaction quality, and 5.49 for personalization; corresponding standard deviations were .92, .96, .94, and 1.00. These descriptive constructs were not assigned additional structural effects.

4.2. Configuration Performance

C3 recorded the highest observed accuracy, trust, conversion, qualified interest, and follow-up acceptance (Table 4). Its autonomy mean was 5.96, compared with 6.10 in Co and 4.82 in C4. C1 recorded 91.1% accuracy and 29.0% conversion, compared with 94.4% and 31.0% in Co. The descriptive pattern therefore varied by configuration: AI availability alone did not correspond to uniformly higher observed performance. Statistical significance was not established for every pairwise difference.

Table 4. Performance by Condition

Measure	Co	C1	C2	C3	C4
Encounters	600	600	600	600	600
Accuracy %	94.4	91.1	96.8	98.1	93.2
Error rate %	5.6	8.9	3.2	1.9	6.8
Autonomy mean	6.10	5.72	5.58	5.96	4.82
Trust mean	5.47	5.18	5.81	6.06	4.89
Conversion % (n)	31 (186)	29 (174)	34 (204)	40 (240)	27 (162)
Interest % (n)	54 (324)	51 (306)	59 (354)	66 (396)	48 (288)
Follow-up % (n)	41 (246)	39 (234)	46 (276)	53 (318)	36 (216)

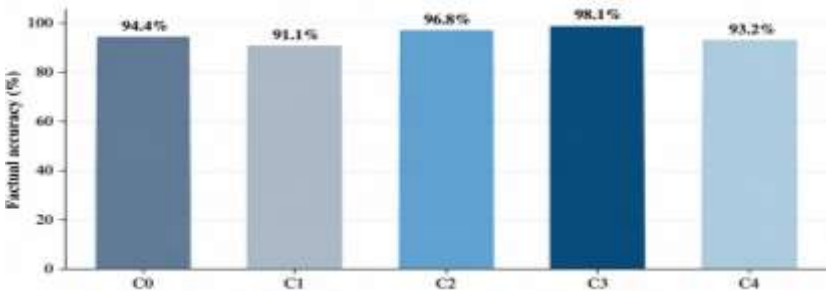

Figure 1. Factual Accuracy across Configurations. Bars Show Reported Rates; Uncertainty Intervals Were Not Supplied

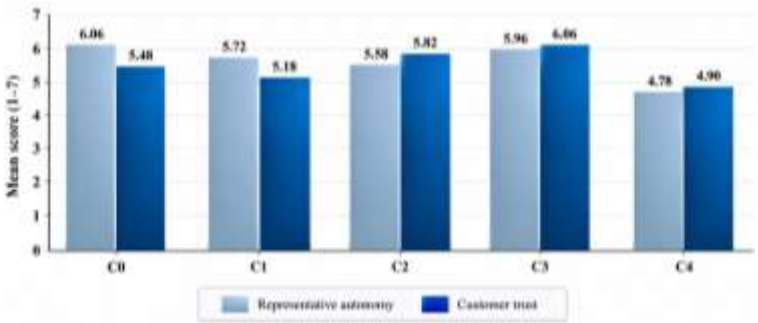

Figure 2. Representative Autonomy and Customer Trust. Raw Means Retain the Reported Seven-Point Scales

Conversion differed across conditions, $\chi^2(4)=28.25$, $p<.001$, Cramér's $V=.097$. Qualified interest also differed, $\chi^2(4)=48.90$, $p<.001$, $V=.128$, as did follow-up acceptance, $\chi^2(4)=43.57$, $p<.001$, $V=.121$. Figure 3 presents these three behavioral outcomes together. C3 recorded 240 conversions, 396 qualified-interest outcomes, and 318 follow-up acceptances among its 600 encounters.

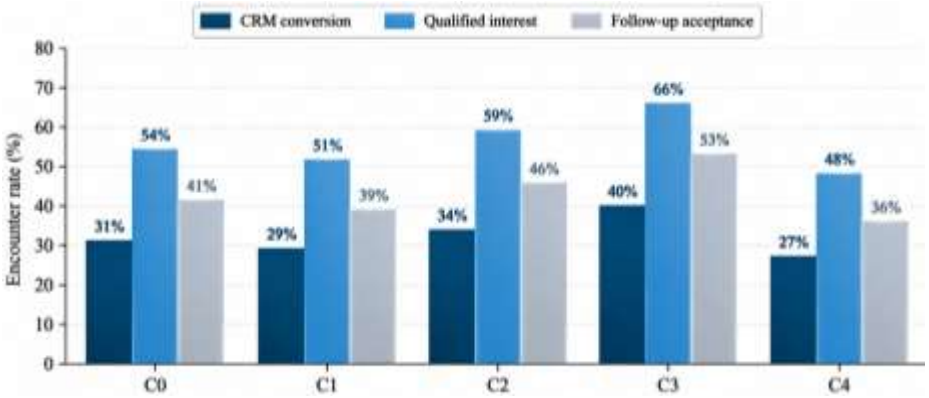

Figure 3. Customer Acquisition Outcomes by Condition. All Behavioral Denominators Are 600 Encounters per Condition

C3 qualified interest and follow-up acceptance reached 66.0% and 53.0%, compared with 54.0% and 41.0% under Co and 48.0% and 36.0% under C4. Overall rates were 55.6% and 43.0%. These outcomes describe distinct acquisition-related behaviors; the analysis did not establish a participant-level sequence from interest through follow-up to conversion.

4.3. Condition Comparisons

C3 conversion exceeded Co by 9.0 percentage points, corresponding to a 29.0% relative increase. The count-based OR was 1.48, 95% CI [1.17, 1.88], with reported Holm-adjusted p=.002. Conversion differences relative to C1, C2, and C4 were 11.0, 6.0, and 13.0 percentage points (Table 5). The count-based odds ratio is distinguished from the covariate-adjusted coefficients reported for the hypothesis models.

Table 5. Planned C3 Comparisons

Comparator	Accuracy Δ pp	Trust Δ / d	Conversion Δ pp / relative %	OR [95% CI]	Holm p
Co	+3.7	+.59 / .62	+9.0 / +29.0	1.48 [1.17–1.88]	.002
C1	+7.0	+.88 / .90	+11.0 / +37.9	1.63 [1.28–2.08]	<.001
C2	+1.3	+.25 / .28	+6.0 / +17.6	1.29 [1.02–1.64]	.032
C4	+4.9	+1.17 / NR	+13.0 / +48.1	1.80 [1.41–2.30]	<.001

Note: *pp = percentage points; NR = not reported. Trust Δ is on the seven-point scale.*

C3 exceeded C2 by 6.0 percentage points in conversion and .25 points in trust; the trust effect size was d=.28. The C3–C4 trust difference was 1.17 points; its numerical Cohen's d was not reported. These comparisons support C3's combined configuration advantage. They do not isolate representative authority, because equivalence of prompts, information, interface, training, timing, and adherence was not established.

Relative conversion improvements over C1, C2, and C4 were 37.9%, 17.6%, and 48.1%, alongside accuracy advantages of 7.0, 1.3, and 4.9 percentage points. Trust effect sizes against Co, C1, and C2 were .62, .90, and .28.

4.4. Hypotheses and Mediation

All five hypotheses were supported by the reported analyses (Table 6). H1 yielded an adjusted accuracy coefficient of .064, SE=.010, z=6.40, p<.001, 95% CI [.044, .083], and d=.61. Verified configurations achieved approximately 97.5% accuracy compared with C1's 91.1%. H2 yielded b=−1.14, SE=.087, t=−13.10, p<.001, 95% CI [−1.31, −.97], and d=−1.06 for the C4–C3 autonomy contrast.

Table 6. Hypothesis Tests

Hypothesis and Path	Estimate	SE	Test	95% CI	p
H1 Verification → accuracy	b=.064	.010	z=6.40	[.044,.083]	<.001
H2 AI authority → autonomy	b=−1.14	.087	t=−13.10	[−1.31,−.97]	<.001
H3 Accuracy → trust	β=.350	.040	z=8.75	[.272,.428]	<.001
H4 Autonomy → trust	β=.230	.036	z=6.39	[.159,.301]	<.001
H5 Sequential indirect	βind=.086	.017	z=5.06	[.055,.121]	<.001

Note: *All hypotheses were supported. Indirect confidence intervals are bootstrapped.*

Accuracy independently predicted trust, β=.350, SE=.040, z=8.75, p<.001, 95% CI [.272, .428]. Autonomy also predicted trust when entered with accuracy, β=.230, SE=.036, z=6.39, p<.001, 95% CI [.159, .301]. Trust predicted conversion with b=.420 log-odds, SE=.069, z=6.09, p<.001, OR=1.52 [1.33, 1.74]. These relationships supported H3 and H4 and the trust-to-acquisition component of H5. Figure 4 distinguishes unstandardized coefficients from standardized relationships.

The sequential indirect effect through accuracy and trust was .086, with bootstrapped 95% CI [.055, .121], p<.001. The additional autonomy pathway was .039, 95% CI [.022, .061], and the reported combined indirect effect was .125, 95% CI [.091, .164], both p<.001. The remaining direct coefficient was .048, SE=.028, z=1.71, p=.087. These estimates supported H5 within the specified model. The nonsignificant direct path does not establish a zero effect or complete mediation. Interpretation of the combined indirect effect also requires confirmation of common exposure coding and an additive effect scale.

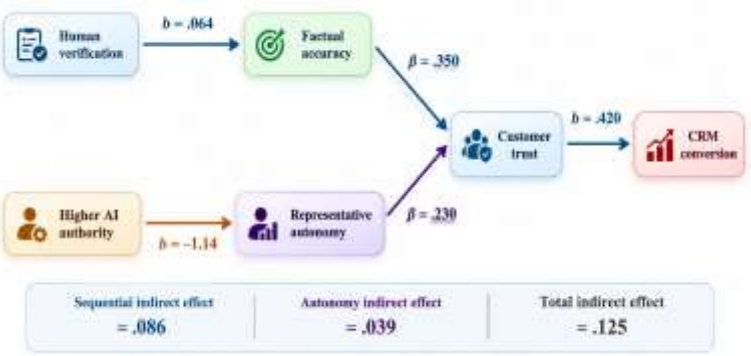

Figure 4. Estimated HSGAI-AF Pathways. B Denotes An Unstandardized Coefficient; B Denotes A Standardized Coefficient. Arrows Represent the Specified Model

4.5. Subgroup Comparisons

C3's reported advantage was positive in every experience and complexity subgroup (Table 7 and Figure 6). Experience-stratum differences ranged from 8.7 to 9.4 percentage points. Complexity-stratum differences ranged from 7.5 to 11.0 points, with the largest numerical difference in high-complexity encounters. Interaction tests were nonsignificant for experience, p=.84, and complexity, p=.27. Accordingly, the results did not establish differential effects across these strata. Subgroup N values describe whole strata rather than condition-specific cell sizes.

Table 7. C3 versus Co Subgroup Results

Subgroup	Overall N	Co / C3 %	Δ pp [95% CI]	Approx OR	Interaction p
Experience <2 years	990	28.5 / 37.2	8.7 [3.1,14.3]	1.49	.84
Experience 2–5 years	1,260	31.4 / 40.6	9.2 [4.2,14.2]	1.49	.84
Experience >5 years	750	33.4 / 42.8	9.4 [2.8,16.0]	1.49	.84
Low complexity	1,050	34.0 / 41.5	7.5 [2.0,13.0]	1.38	.27
Moderate complexity	1,200	31.2 / 40.3	9.1 [3.9,14.3]	1.49	.27

High complexity	750	25.8 / 36.8	11.0 [4.4,17.6]	1.68	.27

4.6. Structural Ablation and Robustness

The full framework yielded trust R^2=.49 and conversion pseudo-R^2=.18 (Table 8). Removing accuracy increased AIC by 88; removing autonomy increased it by 54. Removing trust's conversion pathway increased AIC by 121 and reduced conversion pseudo-R^2 to .10. The AI/no-AI specification had ΔAIC=196 and conversion pseudo-R^2=.09. Figure 5 presents the reported differences. Their use for model ranking requires common observations, compatible likelihoods, and matching observed outcomes. Those conditions were not documented. Pseudo-R^2 reductions describe the reported fit measure rather than losses in sales or conventional explained variance.

Table 8. Structural Ablation

Specification	Trust R^2 / Conversion Pseudo-R^2	AIC / BIC	ΔAIC	Indirect	Pseudo-R^2 loss %
Full framework	.49 / .18	16,420 / 16,678	0	.125	0
Without accuracy	.40 / .15	16,508 / 16,752	88	.063	16.7
Without autonomy	.44 / .16	16,474 / 16,724	54	.086	11.1
Without trust → conversion	.49 / .10	16,541 / 16,783	121	.000	44.4
Simple AI/no-AI	.24 / .09	16,616 / 16,824	196	.000	50.0

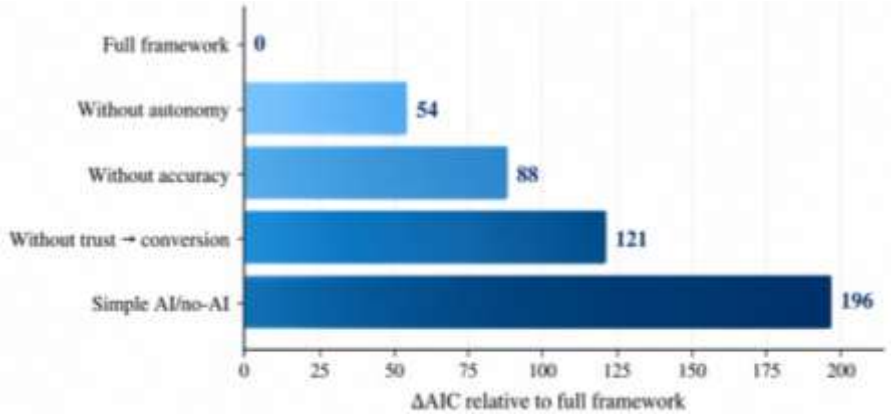

Figure 5. Model Fit Deterioration after Structural Ablation. Larger ΔAIC Indicates Weaker Fit Relative to the Full Framework

Across sensitivity specifications, accuracy-to-trust coefficients ranged from .310 to .370, autonomy-to-trust coefficients from .210 to .240, and sequential indirect effects from .074 to .091. C3–Co conversion ORs ranged from 1.45 to 1.68 (Table 9). The high-complexity specification contained 750 encounters; other models retained 3,000 except the complete-case analysis. Directional stability across these documented specifications supported the central pattern without establishing robustness to untested assumptions.

Table 9. Sensitivity Specifications

Specification	N	Accuracy β	Autonomy β	Sequential Indirect	C3/Co OR [95% CI]
Primary FIML	3,000	.350	.230	.086	1.48 [1.17–1.88]
Complete cases	2,796	.340	.240	.083	1.46 [1.14–1.87]
Representative clustering	3,000	.350	.230	.087	1.49 [1.18–1.89]
Territory fixed effects	3,000	.340	.220	.082	1.47 [1.16–1.87]
High complexity	750	.370	.210	.091	1.68 [1.18–2.39]
Material errors only	3,000	.310	.240	.074	1.45 [1.14–1.85]

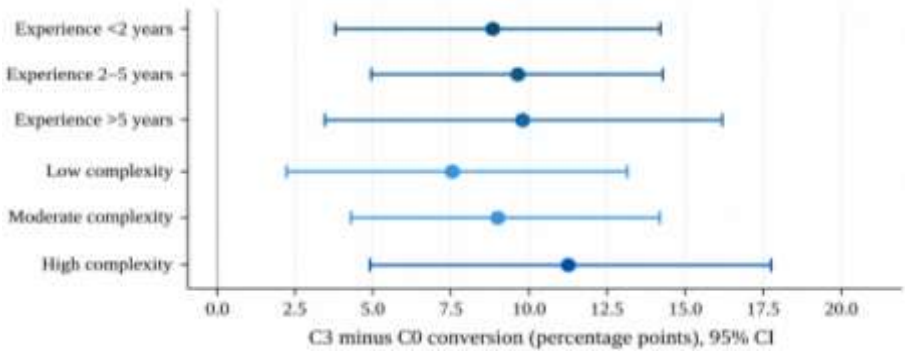

Figure 6. C3 versus Co Subgroup Conversion Differences. Points and Intervals Reproduce the Reported Estimates and 95% Confidence Intervals

4.7. Measurement and Model Diagnostics

Reliability estimates ranged from α=.86 to .92, composite reliability from .87 to .93, and average variance extracted from .63 to .74 across the six reported constructs. Maximum HTMT was .82. CFA indices were CFI=.963, TLI=.957, RMSEA=.041, and SRMR=.038; structural indices were .958, .952, .044, and .041, respectively. Audit agreement was κ=.91 [.89, .93]. Representative and territory ICCs were .08 and .03 for trust and .06 and .02 for conversion (Table 10).

Table 10. Measurement and Model Diagnostics

Construct Or Diagnostic	Reported Statistics
Autonomy	α=.89; CR=.90; AVE=.69
Customer trust	α=.92; CR=.93; AVE=.74
Competence	α=.88; CR=.89; AVE=.67
Transparency	α=.86; CR=.87; AVE=.63
Interaction quality	α=.90; CR=.91; AVE=.72
Personalization	α=.87; CR=.88; AVE=.65
Discriminant validity	Maximum HTMT=.82
CFA	CFI=.963; TLI=.957; RMSEA=.041; SRMR=.038
CFA RMSEA 90% CI	[.037,.045]
Structural model	CFI=.958; TLI=.952; RMSEA=.044; SRMR=.041
Structural RMSEA 90% CI	[.040,.048]
Audit agreement	κ=.91; 95% CI [.89,.93]
Trust ICC	Representative=.08; territory=.03
Conversion ICC	Representative=.06; territory=.02

Note: *CR = composite reliability; AVE = average variance extracted; ICC = intraclass correlation*.

5. Discussion

5.1. Accuracy and Autonomy as Complementary Mechanisms

The study's contribution is to evaluate factual verification and representative discretion jointly within customer acquisition. These dimensions answer different governance questions: whether information is supported and how authority over its use is distributed. Verification was associated with accuracy, greater AI authority with lower autonomy, and both accuracy and autonomy with customer trust. The framework connects those relationships with CRM behavior without treating accuracy and autonomy as interchangeable measures. The empirical contribution lies in their joint examination across the tested configurations; it does not establish that either dimension alone explains every acquisition difference.

This account complements evidence of heterogeneous human–AI performance [1] and the automation–augmentation tension [5]. Marketing and sales frameworks examine capability and task allocation [6, 7, and 9]. HSGAI-AF brings information verification, representative discretion, customer trust, and recorded acquisition into one explanatory account. Its distinctive question concerns how the governance of assistance relates to customer-facing performance. The C2–C3 comparison is informative because both required verification but differed in observed outcomes. However, incomplete operational definitions leave alternative implementation explanations open. The contribution is therefore comparative and explanatory, with the authority mechanism requiring further isolation.

The accuracy pathway connects factuality concerns [22–24] with information communicated during selling. The autonomy pathway concerns representative discretion, distinct from customer control and trust in AI. Their simultaneous trust relationships support retaining both constructs in the framework. Human-only Co nevertheless recorded higher autonomy than C3. The pattern therefore does not support maximizing autonomy independently of information reliability. C3 combined comparatively high autonomy with the strongest observed accuracy and acquisition profile. Calibration describes that tested combination; a continuous trade-off, universally optimal supervision level, and isolated authority effect were not established.

5.2. Implications for Direct Sales Governance

The findings support assessing verification and retained judgment together when evaluating direct-sales assistance. C3's favorable acquisition profile provides a basis for examining that arrangement in comparable settings. C1's lower observed accuracy and conversion caution against treating access to unverified advice as sufficient evidence of effective assistance. Similarly, lower autonomy and trust of C4 distinguishes greater AI authority from stronger observed augmentation. These implications concern the configurations studied. Implementation cost, profitability, verification workload, deployment feasibility, long-term customer retention, and performance in other commercial settings were not measured and cannot be inferred from conversion alone.

Customer trust linked the specified accuracy and autonomy pathways with conversion, complementing research on autonomous-system reliance and service encounters [14–16, 28, 29]. Trust here concerns the customer-facing interaction; it is not evidence of representative trust in AI. The pathway ablations favored the full framework on the reported criteria, conditional on a comparable estimation basis. They do not establish that the fitted account excludes other explanations. Temporal ordering, mediator–outcome confounding, and differences between configurations remain consequential for causal interpretation [34, 35]. The practical implication is to evaluate the combined arrangement while keeping its mechanisms open to further testing.

5.3. Limitations and Reporting Boundaries

Implementation records limit replication and generalizability. Dates, products, locations, recruitment, eligibility, representative and territory count, refusals, and repeat participation remain insufficiently specified. Allocation, concealment, counterbalancing, training, adherence, deviations, exclusions, carryover, and contamination procedures are incomplete. Model/version, model/version, prompts, settings, evidence sources, interface, verification, escalation, logging, and updates are not available. Unresolved are the C2-C3 permission difference and the equivalence of other features. Verified records are a must for access and ethics disclosures.

Measurement limitations include missing scale items, scoring, audit denominators, aggregation, auditor training/blinding, double-coding, adjudication, outcome windows, and complexity definitions. Reliability does not establish measurement invariance [31, 32]; mixed sources do not eliminate survey common-method influences [33]. Analytical gaps concern software, equations, estimators, random effects, cluster counts, uncertainty procedures, bootstrap settings, diagnostics, subgroup cells, mediation scales, pseudo-R^2, and comparable likelihoods. Missingness tests do not establish unbiasedness. Causal mediation requires temporal and confounding assumptions [34, 35].

6. Conclusion

Across 3,000 encounters, C3 combined mandatory verification with retained representative authority and achieved the strongest observed acquisition profile. Conversion reached 40.0%, compared with 31.0% under human-only selling: a 9.0-percentage-point difference, OR=1.48 [1.17, 1.88]. Accuracy and autonomy independently predicted customer trust, and the specified indirect pathways linked supervision with conversion. The results support the combined C3 configuration and distinguish informational

reliability from representative agency. Their interpretation remains bound by incomplete implementation documentation and mediation assumptions. C3's advantage over C2 cannot be attributed exclusively to representative authority.

References

[1] Vaccaro, M.; Almaatouq, A.; Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8(12):2293-2303. https://doi.org/10.1038/s41562-024-02024-1
[2] Brynjolfsson, E.; Li, D.; Raymond, L. (2025). Generative AI at Work. The Quarterly Journal of Economics, 140(2):889-942. https://doi.org/10.1093/qje/qjae044
[3] Noy, S.; Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654):187-192. https://doi.org/10.1126/science.adh2586
[4] Dell'Acqua, F.; McFowland, E.; Mollick, E.; Lifshitz, H.; Kellogg, K.C.; Rajendran, S.; et al. (2026). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. Organization Science, 37(2):403-423. https://doi.org/10.1287/orsc.2025.21838
[5] Raisch, S.; Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. Academy of Management Review, 46(1):192-210. https://doi.org/10.5465/amr.2018.0072
[6] Huang, M.H.; Rust, R.T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49(1):30-50. https://doi.org/10.1007/s11747-020-00749-9
[7] Paschen, J.; Wilson, M.; Ferreira, J.J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. Business Horizons, 63(3):403-414. https://doi.org/10.1016/j.bushor.2020.01.003
[8] Jarotschkin, V.; Soykoth, M.W.; Chaker, N.N. (2025). Artificial intelligence in sales research: Identifying emergent themes and looking forward. Journal of Business Research, 198:115383. https://doi.org/10.1016/j.jbusres.2025.115383
[9] Gonzalez, G.R.; Habel, J.; Hunter, G.K. (2026). AI agents, agentic AI, and the future of sales. Journal of Business Research, 202:115799. https://doi.org/10.1016/j.jbusres.2025.115799
[10] Ngo, V.M. (2025). Human–AI collaboration for marketing capabilities: a meta-analysis. Marketing Letters, 36(4):839-855. https://doi.org/10.1007/s11002-025-09783-5
[11] Grewal, D.; Satornino, C.B.; Davenport, T.; Guha, A. (2025). How generative AI Is shaping the future of marketing. Journal of the Academy of Marketing Science, 53(3):702-722. https://doi.org/10.1007/s11747-024-01064-3
[12] Cillo, P.; Rubera, G. (2025). Generative AI in innovation and marketing processes: A roadmap of research opportunities. Journal of the Academy of Marketing Science, 53(3):684-701. https://doi.org/10.1007/s11747-024-01044-7
[13] Lund, B.D.; Wang, T.; Mannuru, N.R.; Nie, B.; Shimray, S.; Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5):570-581. https://doi.org/10.1002/asi.24750
[14] O'Neill, T.; McNeese, N.; Barron, A.; Schelble, B. (2022). Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. Human Factors: The Journal of the Human Factors and Ergonomics Society, 64(5):904-938. https://doi.org/10.1177/0018720820960865
[15] Lockey, S.; Gillespie, N.; Holm, D.; Someh, I.A. (2021). A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2021.664
[16] Kaplan, A.D.; Kessler, T.T.; Brill, J.C.; Hancock, P.A. (2023). Trust in Artificial Intelligence: Meta-Analytic Findings. Human Factors: The Journal of the Human Factors and Ergonomics Society, 65(2):337-359. https://doi.org/10.1177/00187208211013988
[17] Hancock, J.T.; Naaman, M.; Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. Journal of Computer-Mediated Communication, 25(1):89-100. https://doi.org/10.1093/jcmc/zmz022
[18] Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human–Computer Interaction, 36(6):495-504. https://doi.org/10.1080/10447318.2020.1741118
[19] Jarrahi, M.H.; Lutz, C.; Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. Big Data & Society, 9(2):20539517221142824. https://doi.org/10.1177/20539517221142824
[20] Kellogg, K.C.; Valentine, M.A.; Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. Academy of Management Annals, 14(1):366-410. https://doi.org/10.5465/annals.2018.0174
[21] Vincent, V.U. (2021). Integrating intuition and artificial intelligence in organizational decision-making. Business Horizons, 64(4):425-438. https://doi.org/10.1016/j.bushor.2021.02.008
[22] Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; et al. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12):1-38. https://doi.org/10.1145/3571730
[23] Huang, L.; Yu, W.; Ma, W.; Zhong, W.; Feng, Z.; Wang, H.; et al. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43(2):1-55. https://doi.org/10.1145/3703155
[24] Farquhar, S.; Kossen, J.; Kuhn, L.; Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 630(8017):625-630. https://doi.org/10.1038/s41586-024-07421-0
[25] Liu, S.Q.; Vakeel, K.A.; Smith, N.A.; Alavipour, R.S.; Wei, C.V.; Wirtz, J. (2024). AI concierge in the customer journey: what is it and how can it add value to the customer?. Journal of Service Management, 35(6):136-158. https://doi.org/10.1108/josm-12-2023-0523

[26] Huang, M.H.; Rust, R.T. (2024). The Caring Machine: Feeling AI for Customer Care. Journal of Marketing, 88(5):1-23. https://doi.org/10.1177/00222429231224748
[27] Puntoni, S.; Reczek, R.W.; Giesler, M.; Botti, S. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. Journal of Marketing, 85(1):131-151. https://doi.org/10.1177/0022242920953847
[28] Troshani, I.; Rao Hill, S.; Sherman, C.; Arthur, D. (2021). Do We Trust in AI? Role of Anthropomorphism and Intelligence. Journal of Computer Information Systems, 61(5):481-491. https://doi.org/10.1080/08874417.2020.1788473
[29] George, S.R.; C, M.; Edward, M. (2025). Artificial Intelligence in Frontline Service Encounters: A Systematic Review and Research Agenda. International Journal of Consumer Studies, 49(3):e70048. https://doi.org/10.1111/ijcs.70048
[30] Ledro, C.; Nosella, A.; Vinelli, A. (2022). Artificial intelligence in customer relationship management: literature review and future research directions. Journal of Business & Industrial Marketing, 37(13):48-63. https://doi.org/10.1108/jbim-07-2021-0332
[31] Goretzko, D.; Siemund, K.; Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. Educational and Psychological Measurement, 84(1):123-144. https://doi.org/10.1177/00131644231163813
[32] McNeish, D.; Wolf, M.G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. Psychological Methods, 28(1):61-88. https://doi.org/10.1037/met0000425
[33] Podsakoff, P.M.; Podsakoff, N.P.; Williams, L.J.; Huang, C.; Yang, J. (2024). Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix. Annual Review of Organizational Psychology and Organizational Behavior, 11(1):17-61. https://doi.org/10.1146/annurev-orgpsych-110721-040030
[34] Nguyen, T.Q.; Schmid, I.; Stuart, E.A. (2021). Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. Psychological Methods, 26(2):255-271. https://doi.org/10.1037/met0000299
[35] Cinelli, C.; Forney, A.; Pearl, J. (2024). A Crash Course in Good and Bad Controls. Sociological Methods & Research, 53(3):1071-1104. https://doi.org/10.1177/00491241221099552