

AI-assisted sales coaching framework: Empirically-derived models for B2B communication analysis

Joona Mäntyvaara ^a,* Paavo Nevalainen ^a, Kirill Glavatskiy ^b, Jukka Heikkonen ^a

^a University of Turku, Department of Computing, Turku, Finland

^b University of Newcastle, Newcastle, Australia

ARTICLE INFO

Dataset link: <https://seafle.utu.fi/f/1786a1fc71014bf8a25f/>

Keywords:

AI-assisted coaching frameworks
Network analysis
Temporal prediction
Sales performance optimization
Multimodal communication analysis
Conversation efficiency
Rejection prediction

ABSTRACT

This research addresses the design challenge of integrating multimodal communication analytics into AI-assisted coaching systems suitable for real-time deployment. Drawing on analysis of 5183 Finnish B2B sales calls, this study provides the first empirically-grounded design specifications for multimodal sales coaching by unifying graph-theoretic conversation analysis, temporal prediction, and rejection modeling. Network analysis reveals that successful conversations exhibit 4.3× lower structural density (0.0224 vs. 0.0960) while covering broader topic ranges, establishing conversational efficiency rather than complexity as a guiding design principle. Temporal prediction identifies a 60-second optimal intervention window, achieving 78.4% AUC for reliable real-time guidance. Rejection modeling achieves 95.7% AUC with interpretable early-warning signals validated through SHAP analysis. These findings are operationalized through evidence-based quality indicators spanning acoustic, semantic, and linguistic modalities, supported by computationally efficient formulations suitable for real-time processing. Duration-matched validation confirms threshold robustness independent of call length, and bias auditing demonstrates equitable performance across salesperson groups (FPR disparity = 0.027). The framework provides validated design specifications aligned with EU AI Act compliance provisions, demonstrating how multimodal communication analytics can be transformed into deployable coaching systems.

1. Introduction

Traditional sales coaching is typically conducted after conversations have ended, which limits the effectiveness of interventions and reduces opportunities for timely improvement [1–3]. In contrast, real-time coaching systems can identify performance degradation patterns, provide immediate corrective guidance, and reduce the likelihood of conversation failure [4]. These systems also enable simulation-based training for new salespeople prior to live customer interactions [5–7]. The demand for evidence-based coaching solutions has increased as organizations seek quantifiable performance improvements and measurable return on investment [8,9].

Despite this potential, current research lacks comprehensive design frameworks that integrate multimodal (acoustic, semantic, linguistic) communication analytics into operational coaching systems. Although prior studies have examined individual aspects such as acoustic features [10], semantic content [11], and conversation structure [12], effective real-time deployment requires validated performance thresholds, optimal intervention timing, computational efficiency [13,14], regulatory compliance, interpretability, and human oversight [15,16].

No existing framework systematically incorporates these requirements, resulting in a design challenge: *How can multimodal communication analytics be integrated into evidence-based coaching systems suitable for real-time deployment while maintaining interpretability, accuracy, and regulatory compliance?*

Recent work [17] provides an empirical basis for addressing this challenge. Through a large-scale multimodal analysis of 5183 Finnish B2B sales calls, the study validated a 41-feature framework comprising 15 acoustic, 20 semantic, and 6 linguistic variables. This study demonstrated the principle of acoustic–semantic equivalence, showing that sales success could be predicted equally well from acoustic delivery patterns and semantic content, with linguistic structure providing complementary value. Constrained optimization indicated that equal weighting (one-third for each modality) maximizes ensemble effectiveness, providing an understanding on how to allocate future sales communication training.

Based on these findings, the present research addresses this challenge by developing an AI-assisted coaching framework grounded in empirical analysis of core data and prior research. Employing a design

* Corresponding author.

E-mail addresses: joeman@utu.fi (J. Mäntyvaara), ptneva@utu.fi (P. Nevalainen), kirill.glavatskiy@newcastle.edu.au (K. Glavatskiy), jukhei@utu.fi (J. Heikkonen).

<https://doi.org/10.1016/j.array.2026.100755>

Received 4 January 2026; Received in revised form 12 March 2026; Accepted 12 March 2026

Available online 23 March 2026

2590-0056/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

science research methodology [18,19], the study constructs an IT artifact that integrates acoustic–semantic equivalence principles with validated specifications for system architecture, performance thresholds, intervention timing, and implementation requirements. This approach translates multimodal analytics into actionable coaching guidance [20].

1.1. Contributions of this research

This study advances AI-assisted sales coaching systems by presenting theoretical, methodological, and artifact contributions, each organized into distinct categories.

Theoretical Contributions:

A graph-theoretic communication network analysis approach for B2B sales quantifies communication efficiency and distinguishes structural characteristics of successful versus unsuccessful interactions using network density and semantic transition analysis.

A multimodal computational framework integrates acoustic, semantic, and linguistic quality indicators using empirically optimized equal weighting for each modality. This framework employs computationally efficient operations suitable for real-time processing and maintains interpretability. Complete mathematical formulations are detailed in Supplementary Materials Section 2.

Methodological Contributions:

Progressive temporal analysis is used to identify the optimal intervention window, enabling coaching systems to achieve sufficient predictive reliability for real-time deployment.

Rejection prediction involves an early-warning system for identifying conversational failure risk. This system is validated through SHAP interpretability analysis to ensure that predictions reflect genuine communication dynamics rather than statistical artifacts.

Artifact Contribution:

An evidence-based alert system incorporates empirically derived performance thresholds, structured feedback mechanisms, and EU AI Act compliance provisions. This system provides operational specifications for translating multimodal analytics into actionable real-time coaching guidance (complete system architecture in Supplementary Materials Section 3). Future work will focus on system validation and field deployment, thereby completing the research cycle from theoretical foundations through framework design to operational assessment.

1.2. Framework architecture overview

The AI-assisted coaching framework comprises three integrated analytical components that transform multimodal conversation data into evidence-based coaching guidance: (1) *Communication Network Analysis*—models conversations as directed graphs with utterances as nodes and sequential transitions as edges, quantifying structural efficiency through network density and semantic transition patterns across eight operational clusters; (2) *Temporal Prediction Modeling*—establishes optimal intervention windows through progressive accuracy analysis, enabling timely coaching deployment; (3) *Rejection Risk Detection*—identifies early-warning signals of conversation failure through interpretable feature analysis validated via SHAP attribution.

These analytical components feed into an evidence-based alert system translating quantitative indicators into coaching guidance through 15 calibrated quality indicators spanning acoustic delivery (5 indicators), semantic content (4 indicators), linguistic structure (3 indicators), and integrated multimodal metrics (3 indicators). Each indicator provides color-coded performance zones (Green/Yellow/Red) with specific coaching priorities derived from empirical analysis of 5183 Finnish B2B conversations. The framework's architectural specifications – mathematical formulations, threshold calibrations, processing algorithms – provide reusable design knowledge that organizations can instantiate for context-specific implementations following systematic validation procedures demonstrated in this research (detailed specifications in Supplementary Materials Sections 2–3).

Section 2 reviews the sales communication literature and the use of graph theory in communication. Section 3 presents the methodology, including network analysis approaches and temporal prediction frameworks. Section 4 details the empirical results from conversation efficiency analysis, real-time prediction performance, and rejection pattern identification. Section 5 interprets these findings, compares them with prior work, and presents a comprehensive bias audit. Section 6 presents the AI coaching system design, including the evidence-based alert system. Section 7 concludes with future directions and limitations.

2. Background and context

2.1. B2B sales conversation structure

B2B sales conversations may be conceptualized as progressing through four canonical stages, each associated with distinct communicative functions [21,22]. *Opening* establishes rapport and conversational footing, with early interactional turns potentially constraining subsequent exchange patterns [11,12]. *Discovery* focuses on needs diagnosis through investigative questioning, which empirical research suggests constitutes a foundation for agreement in complex B2B contexts [23,24]. *Proposal* involves articulating value propositions that link solution attributes to customer-specific requirements [25–27]. *Confirmation and Closing* secures commitment through incremental agreement rather than discrete closing techniques [23,28].

Stage transitions may not follow strictly sequential paths; B2B conversations can cycle dynamically among stages in response to emerging conversational contingencies [29,30]. This stage-based framework provides a conceptual foundation for the analytical approach employed in this research. *Progression efficiency* – operationalized as the systematic traversal of canonical stages – offers a metric for assessing conversational structure in the dataset examined here. Extended theoretical background and detailed literature review are provided in Supplementary Materials Section S1.

2.2. Coaching intervention framework

The integration of adaptive feedback mechanisms into sales contexts is based on established principles in organizational psychology and computational intervention research. Prior work suggests that effective coaching requires temporally aligned feedback cycles that mirror task dynamics and cognitive load [31,32]. Three archetypal intervention modes have been identified in the coaching and human–computer interaction literature: real-time alerts, strategic redirection suggestions, and post-call performance feedback.

Real-time alerts provide immediate feedback during ongoing performance when indicators signal potential degradation. These systems parallel the concept of just-in-time adaptive interventions (JITAs) that dynamically adjust support based on contextual cues [13,33]. Such real-time mechanisms have been shown to prevent error escalation and sustain engagement under time-constrained decision conditions.

Strategic redirection suggestions represent mid-task guidance designed to guide ongoing behavior toward optimal trajectories. Studies in human–AI collaboration demonstrate that adaptive and context-sensitive guidance can enhance task performance without diminishing perceived autonomy [34]. These findings provide theoretical grounding for conversational AI systems capable of offering in-situ strategic prompts.

Post-call performance feedback reflects the traditional model of reflective learning through structured evaluation. The feedback intervention theory highlights that evidence-based retrospective feedback fosters metacognitive learning and long-term performance improvement when grounded in concrete behavioral data [35,36]. Such feedback mechanisms remain essential for consolidating learning after task execution.

Together, these frameworks position real-time, strategic, and retrospective feedback as dimensions of a coherent coaching system. Their theoretical integration enables the design of human-centered AI systems that support continuous improvement while maintaining human agency and interpretability.

2.2.1. Optimal intervention timing definition

Determining the timing of an effective intervention has been widely discussed in adaptive learning and real-time decision-support research. The balance between predictive confidence and actionable opportunity mirrors the trade-off identified in just-in-time adaptive systems [13, 14]. Early interventions may lack reliability due to incomplete contextual data, whereas late interventions, although accurate, reduce opportunities for behavioral change.

Research in adaptive tutoring, behavioral guidance, and real-time feedback systems suggests that mid-sequence interventions often achieve the greatest effectiveness by combining an adequate situational context with the remaining time for corrective action [14,33]. In sales coaching, this principle implies that intervention windows should be empirically calibrated based on conversational phase progression, response dynamics, and performance variability.

Consistent with human-centered AI design principles [16], coaching systems should maintain advisory rather than prescriptive authority. This ensures compliance with autonomy and transparency requirements while allowing dynamic adaptation to user context, conversation state, and individual performance history.

2.3. Conversation analytics tools

Graph theory applications in communication research have established quantitative methods for examining conversation structure, as pioneered by Wasserman and later developed by Scott [37,38]. Social network analysis has proven to be valuable for understanding communication phenomena and structural patterns, as demonstrated by the work of Borgatti et al. [39].

The field of temporal conversation analysis has witnessed considerable advancement, particularly through the integration of computational linguistics [40] and research on dialog systems [41,42]. Studies on turn-taking dynamics [12] have established the temporal foundations of human communication. Furthermore, recent developments in conversational AI [43] and context-aware dialog act classification [44] exemplify increasingly sophisticated approaches to modeling conversation dynamics. Meeting summarization research [45] has also contributed by examining temporal patterns in recorded conversations, thereby advancing real-time processing capabilities.

Despite these methodological advances, systematic application of network and temporal analysis to sales conversations has received limited empirical attention. Current sales training approaches often rely substantially on subjective evaluation and post-hoc analysis. The limited application of rigorous analytical frameworks to sales conversation optimization represents both a theoretical gap and a practical limitation in leveraging network and temporal modeling capabilities for AI-assisted sales coaching systems.

Research on streaming data analytics demonstrates how sliding window approaches and incremental learning maintain predictive performance with evolving data [46–48], providing methodological foundations for temporal prediction frameworks. Interpretable machine learning emphasizes transparent explanations and reliable confidence measures for high-stakes business applications [49,50]. These principles inform advisory system design that supports human decision-making through explainable predictions rather than automated decisions. Human-centered system design research emphasizes transparency and oversight in automated systems [51], establishing principles for advisory rather than determinative guidance systems in business applications. These methodological approaches provide the foundation for developing interpretable, real-time coaching frameworks that maintain transparency and human oversight.

3. Methodology

The analytical objective is to identify the structural and temporal patterns that distinguish successful from unsuccessful conversations. This comparison enables the development of evidence-based coaching interventions that can guide salespeople toward more effective communication strategies during ongoing conversations.

In the original data set, each conversation is classified as either “successful” (WON) or “unsuccessful” (LOST) based on specific business outcomes. Successful conversations (254 conversations) result in confirmed purchases, signed contracts, scheduled follow-up meetings, or explicit commitment to the next steps. Unsuccessful conversations (4929 conversations) end without commitment, with explicit rejection, or with an indefinite postponement, indicating a lack of genuine interest.

In accordance with established literature [12,40], sales conversations are segmented into utterances, defined as continuous speech segments bounded by natural pauses exceeding 500 ms. Utterances are extracted from audio recordings using the OpenAI Whisper (base variant) [52] speech recognition software, followed by voice activity detection to segment continuous speech at natural pause boundaries greater than 500 ms. This approach ensures that utterance boundaries correspond to speaker intent rather than arbitrary time intervals.

The analytical framework begins with stage progression analysis to validate classical sales stage models, measuring systematic progression through the four canonical sales stages and identifying optimal allocation patterns that characterize successful conversations. Building on this foundational understanding, three complementary methodologies establish the empirical basis for the real-time alert system. First, communication network analysis employs graph-theoretic modeling to quantify communication structure and efficiency patterns, revealing how successful calls organize the flow of communication. Second, temporal prediction modeling establishes the optimal intervention windows in which coaching systems can achieve sufficient accuracy for deployment. Third, rejection pattern analysis identifies behavioral patterns associated with unsuccessful calls. These three analytical components converge in the design of the alert system, translating empirical findings into actionable coaching guidance through evidence-based performance thresholds and multimodal quality indicators.

3.1. Sales stage progression efficiency

This analysis measures the systematic progression of conversations through the four canonical sales stages. The progression analysis evaluates conversation dynamics by systematically assessing temporal allocation patterns and transition efficiency metrics across Opening, Discovery, Proposal (which combines value proposition and objection handling), and Confirmation (which combines confirmation and closing).

Stage progression efficiency quantifies the breadth of stage coverage during conversation flow, operationalizing the canonical sales stage framework [21,22]:

$$\text{Progression Efficiency} = \frac{U}{4} \quad (1)$$

where U represents the number of unique sales stages traversed in chronological sequence (Opening, Discovery, Proposal, Confirmation). Values range from 0.25 (single stage only) to 1.0 (complete progression through all four canonical stages), with higher scores indicating more comprehensive sales process execution.

Additionally, we measure stage transitions as the number of switches between the four sales stages during a conversation. A stage transition occurs when consecutive utterances are assigned to different sales stages (e.g., Opening → Discovery, Discovery → Proposal). The total number of stage transitions quantifies the frequency of stage switching throughout the conversation.

This metric provides empirical evidence for sales stage allocation patterns that distinguish successful from unsuccessful conversations. The results section demonstrates how systematic progression through canonical stages correlates with conversation outcomes, revealing optimal temporal allocation strategies that characterize effective sales interactions.

3.2. Communication network analysis

While classical sales theory organizes conversations around four canonical stages (Opening, Discovery, Proposal, Confirmation), real-time coaching systems require finer-grained classification to capture the dynamic communication behaviors that distinguish successful from unsuccessful interactions. We therefore extend the four-stage framework into seven operational phases that enable more precise detection of critical communication dynamics:

- **Opening:** Rapport building and establishing conversational footing (corresponds to classical Opening stage)
- **Discovery:** Needs assessment and diagnostic questioning (corresponds to classical Discovery stage)
- **Value Proposition:** Solution presentation aligned with identified needs (operational refinement of classical Proposal stage)
- **Objection Handling:** Addressing customer concerns and resistance (critical transition behavior often embedded within Proposal stage)
- **Confirmation:** Securing agreement and commitment signals (corresponds to classical Confirmation stage)
- **Closing:** Explicit commitment requests and next-step finalization (operational refinement separating soft confirmation from explicit closing)
- **Control:** Conversation direction management and topic steering (meta-conversational behavior cutting across all stages)

This seven-phase operational taxonomy preserves theoretical alignment with canonical sales stages while providing the behavioral granularity necessary for real-time coaching intervention. The distinction between *value proposition* and *objection handling* within the classical Proposal stage enables coaching systems to differentiate between proactive value articulation and reactive concern management. Similarly, separating *confirmation* from *closing* distinguishes between securing intermediate agreement and requesting final commitment—a critical distinction for optimal intervention timing. The *control* phase captures meta-conversational behaviors (redirecting discussion, managing time, handling interruptions) that occur throughout the conversation but require distinct coaching guidance. An eighth category, *general*, captures utterances that do not align with specific sales phases, such as off-topic discussions or social exchanges; these are excluded from phase progression metrics but retained in network analysis as they reflect conversation structure patterns that distinguish successful from unsuccessful calls.

We model sales conversations as directed graphs $G = (V, E)$, where nodes V represent individual utterances and edges E represent sequential temporal transitions between consecutive utterances [37,38,53,54]. Each utterance is classified into one of the seven operational phases using rule-based semantic analysis that incorporates Finnish-specific linguistic markers. The analysis is limited to conversations containing at least five utterances (4467 conversations), thereby ensuring sufficient material for meaningful network structure analysis and excluding brief exchanges such as immediate rejections.

Quantitative network metrics can be translated into visually distinct topological patterns that illustrate the fundamental architectural differences between successful and unsuccessful sales conversations. Fig. 1 illustrates distinct topological patterns that demonstrate the principle of “efficiency over complexity” in practice.

The graph structure follows the temporal utterance sequence, with each edge connecting consecutive utterances in chronological order.

Network density emerges from the relationship between conversation length (number of utterances) and the sequential edge structure. Each edge possesses three quantified attributes that capture conversation dynamics: (1) temporal gap τ (duration in milliseconds between consecutive utterances), quantifying response timing and conversation rhythm, (2) semantic transition (ordered pair of conversation phases assigned to consecutive utterances, e.g., “discovery→value”), capturing phase progression patterns, (3) position jump $\Delta p = (t_{\text{current}} - t_{\text{previous}})/(t_{\text{total}})$ representing the proportional advancement through the conversation timeline.

We calculate the network density using the standard graph-theoretic formulation [37]:

$$\rho(G) = \frac{|E|}{|V|(|V| - 1)} \quad (2)$$

where $|E|$ represents the number of sequential utterance transitions and $|V|$ represents the number of utterances. In sequential networks where $|E| \approx |V| - 1$, this metric primarily reflects conversation length, with longer conversations yielding lower density values due to the quadratic growth of the denominator. Lower density thus indicates sustained engagement with more utterances, while higher density indicates shorter conversations with proportionally fewer total utterances.

Furthermore, we measure conversation phase diversity:

$$\text{Coverage}(G) = \frac{C}{7} \times |V| \quad (3)$$

where C represents the number of distinct conversation phases that actually occur in the conversation (ranging from 1 to 7 phases). This metric combines phase diversity with conversation length, reflecting both topical breadth and conversational comprehensiveness.

Additionally, we calculate centrality variance to quantify the distribution of node importance within each conversation network. For each node $v \in V$, we compute its degree centrality $d(v)$, then calculate the variance across all nodes. In sequential networks, centrality variance reflects the distribution of utterance connectivity at conversation boundaries and phase transitions, with higher variance indicating concentrated connectivity patterns and lower variance suggesting more evenly distributed flow.

To identify conversation flow patterns within the seven operational phases, we analyze semantic phase transition patterns. For each conversation, we count the occurrence of each phase-to-phase transition type based on the operational taxonomy (e.g., discovery→value_proposition, confirmation→confirmation). Transition frequency for a specific transition type is calculated as the average number of occurrences per call within each outcome group. Following standard practice in comparative analysis [39], the enrichment factor for each transition is defined as:

$$\text{Enrichment} = \frac{\text{Frequency}_{\text{successful}}}{\text{Frequency}_{\text{unsuccessful}}} \quad (4)$$

where higher enrichment factors indicate transitions that occur more frequently in successful calls relative to unsuccessful calls.

The semantic phase transitions captured in edge attributes reveal conversation management patterns beyond the sequential topology. Successful salespeople demonstrate systematic phase progression patterns (e.g., opening→discovery→value→confirmation), while unsuccessful calls show more scattered phase switching despite following the same sequential utterance structure. These phase-level patterns, combined with conversation length differences, create the observed network density distinctions between successful and unsuccessful calls.

We calculate the above metrics for each of 4467 conversations. To establish whether the observed differences in network metrics represent genuine patterns rather than random variation, we compare successful ($n = 254$) and unsuccessful ($n = 4213$) conversations using appropriate statistical tests. Given the non-normal distribution of network metrics, the Mann–Whitney U tests provide a robust comparison of distribution rankings between outcome groups [55]. Effect sizes are quantified using Cohen’s d to assess practical significance beyond statistical significance [56].

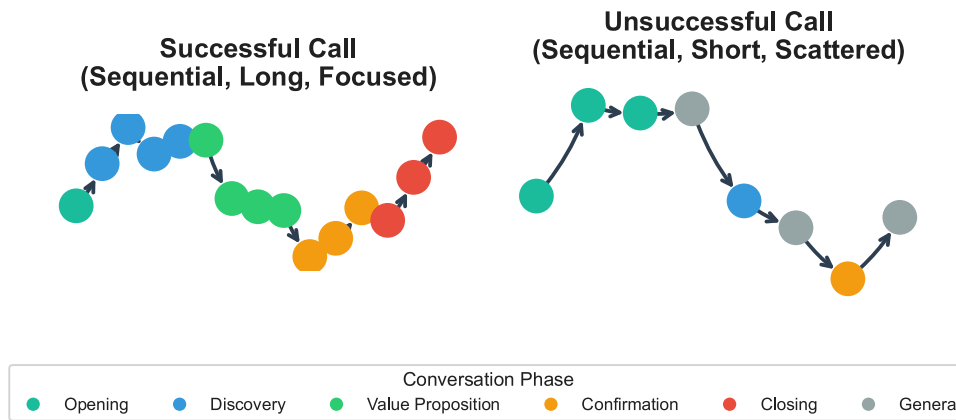


Fig. 1. Representative sequential conversation networks. Successful calls (left) demonstrate longer utterance chains with focused phase progression—brief opening, extensive discovery and value proposition, strong confirmation, and clear closing. Unsuccessful calls (right) show shorter chains with extended opening phase, scattered general discussion, and absence of closing commitment. Nodes represent individual utterances colored by conversation phase; edges represent sequential temporal transitions. Lower density in successful calls results from increased conversation length within the sequential network structure.

3.3. Temporal prediction

We implement temporal analysis to identify conversation outcomes from sales calls. The prediction target is binary using our data pre-classification: successful calls (confirmed purchases, signed contracts, committed next steps) versus unsuccessful calls (rejections, no commitment, no follow-up).

Features are extracted at fixed 30-second intervals using cumulative conversation data from the start of each call up to each time point. Three modalities are analyzed: **acoustic features** (pitch stability, volume consistency, speaking rate, pause patterns, voice energy), **semantic features** (FinBERT embeddings [57,58] with PCA dimensionality reduction from 768D to 20D, topic content analysis, conversation phase identification), and **linguistic features** (conversation structure metrics, question pattern analysis). Rate-of-change indicators are calculated by comparing current window features to the previous 30-second interval.

To quantify differential modality contributions to prediction performance, feature importance is extracted from trained Random Forest, Gradient Boosting, and Multi-Layer Perceptron models at each temporal interval. The relative importance of each modality is calculated as the proportion of total feature importance attributed to that modality:

$$\text{Contribution}_{\text{modality}} = \frac{\sum \text{Importance}_{\text{features in modality}}}{\sum \text{Importance}_{\text{all features}}} \quad (5)$$

This proportional calculation enables analysis of how acoustic, semantic, and linguistic information contribute to prediction accuracy across different temporal windows, supporting identification of optimal coaching intervention timing and modality-specific guidance strategies.

3.3.1. Intervention window analysis

The intervention window is the set of time intervals during a conversation when prediction models achieve sufficient accuracy for coaching deployment. We evaluate prediction performance at discrete time points (30 s, 60 s, 120 s, 180 s, 300 s from conversation start) to identify optimal intervention timing.

Three machine learning algorithms are employed to ensure robust prediction performance: Random Forest [59] (ensemble of decision trees), Gradient Boosting [60,61] (sequential error correction), and Multi-Layer Perceptron [62,63] (neural network). Each algorithm is trained to distinguish successful from unsuccessful conversations using the extracted features.

At each time interval, we generate binary outcome predictions using conversation data accumulated from the start of the call up to that time point. Prediction accuracy is measured using Area Under the Receiver Operating Characteristic Curve (AUC) [64].

Prediction reliability is modeled using a Temporal Prediction Confidence Index (TPCI):

$$\text{TPCI}_t = \text{AUC}_{\text{base}} \cdot \exp(-\alpha \cdot |t - t_{\text{optimal}}|) \quad (6)$$

where AUC_{base} represents maximum achieved accuracy, t_{optimal} indicates the empirically determined peak performance time point, and α controls accuracy decay rate away from the optimal window.

3.4. Rejection analysis

Rejection analysis examines unsuccessful calls ($n = 4,929$) to identify failure patterns indicating salesperson communication breakdown. This analysis categorizes unsuccessful interactions into four types based on engagement duration and intensity: instant refusals (0–15 s with minimal engagement), quick dismissals (16–60 s with brief interaction), engaged conversations (61–300 s with substantial discussion ending unsuccessfully), and deep rejections (over 300 s with extended interaction before ultimate failure).

To capture the multifaceted nature of rejection patterns, composite features aggregate related indicators into interpretable risk components. The *Communication Risk Component* combines weighted scores for aggressive language (pushy directives: “buy”, “must”, “absolutely necessary”), unclear value proposition (hedging language: “maybe”, “somehow”, “kind of”), poor timing pressure (urgency signals: “immediately”, “now”, “today”), and desperation signals (“last chance”, “special offer”, “now or never”). The *Acoustic Risk Component* aggregates multiple acoustic problems including low voice energy (volume mean < 0.3), monotone delivery (low pitch range), inconsistent pace, and excessive pausing (>2 s). The *Engagement Decline Rate* quantifies temporal patterns by measuring the linear regression slope of prospect engagement (utterance word count) over conversation time, capturing how quickly interest deteriorates.

These composite features enable the model to capture complex behavioral patterns (e.g., a salesperson being simultaneously aggressive and unclear) rather than treating indicators as independent. Individual component features (pitch variation, voice energy, speech rate, question density) supplement the composite measures to provide granular behavioral insights.

3.4.1. Classification approach

Rejection risk prediction employs Random Forest [59] classification using ten features spanning acoustic indicators (pitch variation, voice energy, pitch level, speech rate), semantic-linguistic measures (question density, unclear value score, too aggressive score), and composite risk components (communication risk, acoustic risk, engagement decline

Table 1
Stage duration allocation and progression efficiency results.

Metric	Successful	Unsuccessful	Difference	p-value	Cohen's d
Stage duration allocation (%)					
Opening	18.9	30.8	-11.9	<0.0001	-0.75
Discovery	23.4	21.8	+1.6	0.0028	+0.11
Proposal	10.5	8.0	+2.5	<0.0001	+0.31
Confirmation	27.1	22.8	+4.4	<0.0001	+0.29
Progression efficiency					
Average score	7.220	1.873	+5.348	-	-
Efficiency ratio	3.85:1 (Successful:Unsuccessful)				

rate). The prediction target divides conversations into high-risk and low-risk categories based on the median split of the composite rejection score, enabling identification of salesperson behavioral patterns most strongly associated with conversation failure.

Model evaluation uses stratified 5-fold cross-validation with performance measured through AUC. Model interpretability is ensured through SHAP (SHapley Additive exPlanations) [65] analysis, which provides feature importance attributions at both individual conversation and aggregate levels. This validates that the prediction accuracy reflects genuine communication patterns rather than statistical artifacts.

3.5. Algorithm validation framework

Stratified 5-fold cross-validation [66] ensures balanced representation of both outcome classes (successful and unsuccessful calls) across training and testing sets. The dataset is divided into five parts, with each part serving as a test set while the remaining four parts train the model. This process repeats five times to ensure reliable performance estimates.

AUC quantifies how well the model separates successful from unsuccessful calls across all possible classification thresholds. F1-scores provide additional evaluation by balancing precision (accuracy of success predictions) and recall (percentage of successful calls correctly identified).

3.6. Performance threshold derivation

To translate analytical findings into actionable coaching guidance, we derive empirical performance thresholds by contrasting the distributions of measured indicators between successful and unsuccessful calls. This threshold derivation methodology enables systematic identification of performance patterns that distinguish effective from ineffective sales communication.

Performance ranges are segmented into three zones based on empirical distribution analysis: *green* zones represent indicator values typical of top-quartile successful calls, *yellow* zones indicate borderline performance requiring attention, and *red* zones identify patterns strongly associated with conversation failure.

4. Analysis results

4.1. Stage progression analysis results

Stage progression analysis examines conversation dynamics through systematic evaluation of temporal allocation patterns and transition efficiency metrics across the four canonical sales stages: Opening, Discovery, Proposal, and Confirmation.

Table 1 presents stage duration allocation with statistical significance testing. Successful calls demonstrate shorter Opening stage duration (18.9% vs. 30.8%, difference = -11.9%, $p < 0.0001$, Cohen's $d = -0.75$), indicating more rapid transition to substantive conversation elements (Fig. 2). In the Discovery, Proposal, and Confirmation stage,

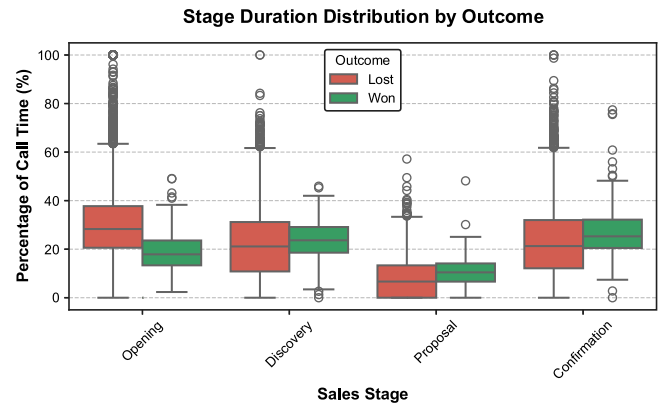


Fig. 2. Stage duration distribution by call outcome. Successful calls allocate significantly less time to Opening (18.9% vs. 30.8%, $p < 0.0001$) and proportionally more time to Discovery, Proposal, and Confirmation stages (all $p < 0.01$).

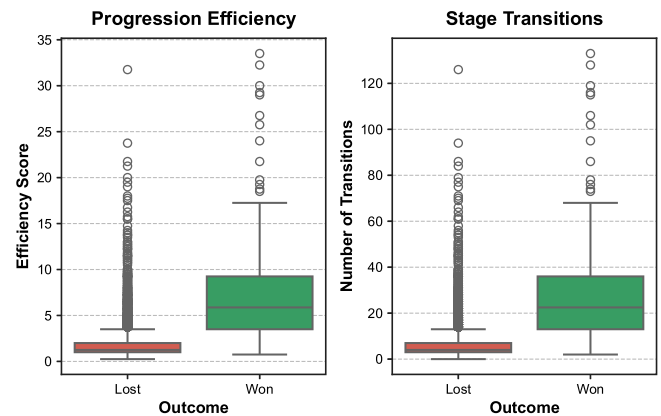


Fig. 3. Stage progression efficiency and transition metrics by call outcome. Successful calls demonstrate 3.85× higher progression efficiency (7.220 vs. 1.873) and more systematic stage transitions compared to unsuccessful calls.

successful calls allocate proportionally more time than unsuccessful calls, with all differences reaching statistical significance ($p < 0.01$).

Successful calls achieved 3.85× higher progression efficiency scores (7.220 vs. 1.873, difference of 5.348), as shown in Fig. 3. This difference indicates that successful calls not only allocate time differently across stages, but also demonstrate more complete progression through the canonical stage sequence with fewer redundant transitions.

Stage progression analysis provides complementary insights into conversation structure and informs training program design by identifying optimal time allocation patterns across sales stages. However, these metrics are not directly incorporated into the real-time alert system due to their retrospective nature requiring complete conversation analysis. The alert system instead focuses on real-time computable indicators from network analysis, temporal prediction, and rejection risk modeling that enable intervention during ongoing conversations.

4.2. Communication network analysis results

We successfully performed network analysis on 4467 of the 5183 total calls (86.2%), excluding 716 calls (13.8%) due to insufficient utterances for meaningful graph construction. The excluded calls typically represented immediate rejections or technical failures, providing insights into the distribution of interaction types in authentic B2B sales environments.

Table 2
Conversation network structure comparison results.

Network metric	Successful	Unsuccessful	p-value
Number of calls	254	4213	
Average network density	0.0224	0.0960	<0.0001
Median number of nodes	63.0	11.0	<0.0001
Average centrality variance	0.0001	0.0034	<0.0001
Median coverage	39.2	3.9	<0.0001

Table 3
Semantic transition enrichment factors in successful vs. unsuccessful calls.

Transition pattern	Enrichment	Won Freq.	Lost Freq.
value_proposition → value_proposition	33.17	0.016	0.000
control → value_proposition	16.59	0.008	0.000
confirmation → value_proposition	16.59	0.004	0.000
confirmation → confirmation	14.15	0.594	0.042
value_proposition → control	11.06	0.008	0.001
discovery → value_proposition	11.06	0.024	0.002
closing → control	8.29	0.028	0.003
general → control	7.38	0.728	0.099
control → objection_handling	6.54	0.382	0.058
value_proposition → general	6.48	0.260	0.040

Analysis of conversation structure through graph-theoretic approaches reveals fundamental differences between successful and unsuccessful sales interactions. Table 2 presents the network metrics for both conversation types.

The table shows that successful calls contain substantially more conversational nodes (median 63 vs. 11, $p < 0.0001$) yet exhibit a lower network density (0.0224 vs. 0.0960, $p < 0.0001$), suggesting longer interactions with fewer returns to previously covered material. Successful calls also show lower centrality variance (0.0001 vs. 0.0034, $p < 0.0001$) and tenfold higher median coverage (39.2 vs. 3.9, $p < 0.0001$), indicating a more evenly distributed conversational structure and engagement across a wider range of sales phases. These structural differences suggest that successful calls follow more focused conversation pathways with systematic topic development, while unsuccessful calls exhibit higher connectivity among fewer nodes, indicating scattered topic exploration without sustained development of individual themes.

Beyond these structural metrics, examining the specific semantic transitions between the conversation phases reveals additional patterns that distinguish successful from unsuccessful interactions.

4.2.1. Semantic transition enrichment analysis

Semantic transition enrichment analysis reveals distinct conversation flow patterns between successful and unsuccessful calls. Table 3 presents the ten transitions with the highest enrichment factors, selected from the complete set of measured phase-to-phase transitions.

The transitions with the highest enrichment factors involve value proposition and confirmation phases. Value proposition to value proposition transitions show the highest enrichment (33.17×), occurring at frequency 0.016 in successful calls compared to 0.000 in unsuccessful calls. Confirmation to confirmation transitions show enrichment of 14.15×, with frequency 0.594 in successful calls versus 0.042 in unsuccessful calls. Multiple transitions involving movement toward value proposition phases show enrichment factors above 10×, including control to value proposition (16.59×), confirmation to value proposition (16.59×), and discovery to value proposition (11.06×).

These patterns indicate that successful calls involve more frequent sustained discussion within value proposition phases, more frequent sequential confirmation exchanges, and more frequent transitions from other phases toward value proposition discussion. Successful calls also show higher frequencies of transitions involving conversation control phases, including general to control (7.38×) and control to objection handling (6.54×).

Real-Time Prediction Performance

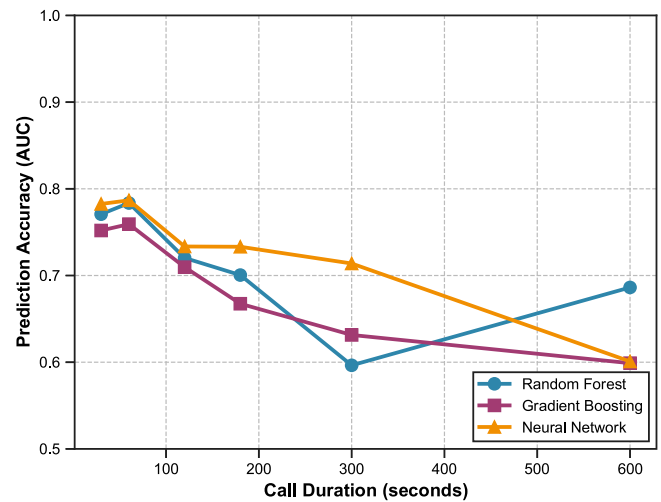


Fig. 4. Real-time prediction performance across conversation timeline. All three algorithms demonstrate peak performance at approximately 60 s, achieving 78.4% AUC. Performance declines after this optimal window, with MLP consistently outperforming Random Forest and Gradient Boosting across all time points.

4.2.2. Statistical validation results

All network metrics were analyzed using Mann–Whitney U tests due to non-normal distributions, particularly in unsuccessful calls with their sparse network structures. The analysis confirms highly significant differences ($p < 0.001$) with large to very large effect sizes: network density (Cohen’s $d = 1.78$), number of nodes ($d = 1.16$), and number of edges ($d = 1.16$). Bootstrap confidence intervals ($n = 1000$) demonstrate complete separation between groups, with non-overlapping 95% CIs for density [0.020, 0.025] vs. [0.095, 0.098] and nodes [78.6, 97.5] vs. [18.1, 19.6], strengthening confidence in these structural differences.

4.3. Temporal prediction performance results

Understanding when conversation outcomes become predictable enables identification of optimal coaching intervention timing. This section evaluates prediction accuracy at progressive time intervals to establish empirical windows where coaching systems achieve sufficient reliability for deployment.

4.3.1. Real-time outcome analysis

Real-time prediction of conversation outcomes (successful vs. unsuccessful) enables identification of optimal intervention windows for coaching systems. Progressive analysis across conversation timelines reveals distinct temporal patterns in outcome prediction accuracy, with performance peaking at 60 s and declining thereafter (see Fig. 4).

Temporal Performance Peak at 60 s: Conversation outcome prediction accuracy reaches the highest value at 60 s across all three algorithms (Random Forest, Gradient Boosting, and Multi-Layer Perceptron). Early prediction at 30 s demonstrates comparable performance (78.3% AUC), indicating that predictive features emerge within the first minute of conversation. These results indicate that intervention timing at approximately 60 s from conversation initiation achieves peak predictive performance in this dataset.

Systematic Performance Degradation: Beyond the 60-second peak, prediction accuracy declines systematically across all algorithms. Performance drops to 73.2% AUC at 120 s, 71.5% AUC at 180 s, and 59.7–71.5% AUC at 300 s. This decay pattern indicates that

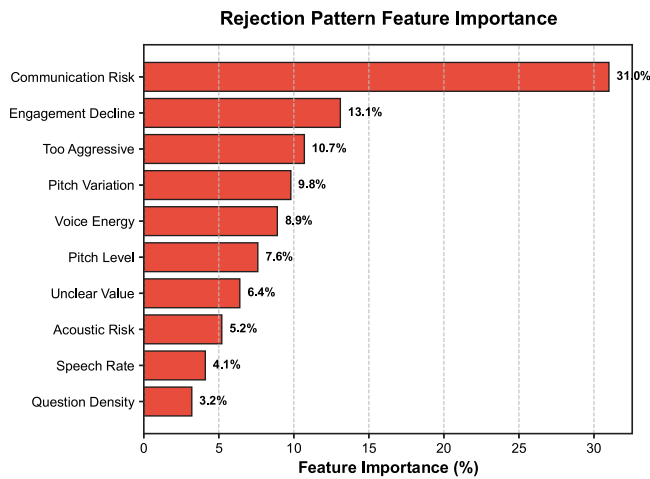


Fig. 5. Feature importance for rejection analysis results. Communication risk component dominates with 31.0% importance, followed by engagement decline rate (13.1%) and aggressive behavior patterns (10.7%).

early conversation phases contain the most predictive information for outcome determination.

Algorithm Comparative Performance: Multi-Layer Perceptron consistently outperforms Random Forest and Gradient Boosting across all temporal intervals, maintaining superior accuracy even during periods of general performance degradation. At the optimal 60-second window between the start of the conversation and the moment of maximum performance, MLP achieves 78.4% AUC compared to 78.0% for Random Forest and 76.0% for Gradient Boosting. This performance advantage persists throughout the temporal analysis, suggesting superior capability for temporal pattern recognition in conversation data.

Based on these empirical results, the temporal prediction confidence index parameters were established: $AUC_{base} = 0.784$ and $\alpha = 0.05$, representing the empirically derived decay rate observed in the temporal performance analysis.

4.4. Rejection analysis results

The unsuccessful call dataset (4929 unsuccessful calls) provides substantial empirical basis for rejection pattern analysis, enabling systematic understanding of failure mechanisms across conversation contexts and salesperson behaviors. This comprehensive multimodal analysis identified distinct subgroups within rejected calls based on salesperson behavioral patterns, achieving 95.7% AUC in distinguishing high-risk from low-risk rejection patterns. This enables systematic understanding of conversation breakdown mechanisms and behavioral failure modes that drive prospect disengagement. Feature importance analysis is presented in Fig. 5.

4.4.1. SHAP interpretability validation results

Given the high rejection analysis accuracy (95.7% AUC), SHAP (SHapley Additive exPlanations) [65] analysis was conducted to validate the interpretability of the model and confirm the reliability of feature importance. This analysis ensures that high performance reflects genuine classification patterns rather than overfitting artifacts (see Fig. 6).

Cross-validation between Random Forest and SHAP importance shows a strong correlation ($r = 0.94$), confirming consistent feature classification between methodologies (see Fig. 7).

The feature importance distribution reveals the multifaceted nature of rejection patterns and provides actionable insights for coaching system design. Communication Risk Component dominance (31.0%)

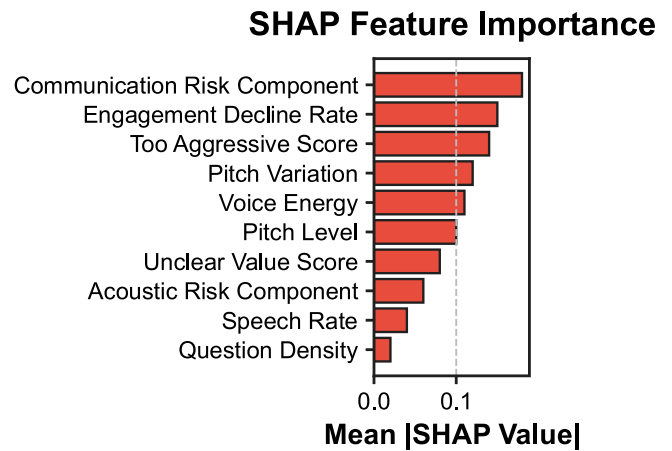


Fig. 6. SHAP summary plot — Feature impact on rejection classification (AUC = 0.956). SHAP values demonstrate that the communication risk component shows the strongest classification impact, with high values consistently indicating rejection characteristics.

indicates that linguistic and semantic issues – aggressive language, unclear value articulation, timing pressure – constitute the primary driver of conversation failure. The prominence of Engagement Decline Rate (13.1%) and Too Aggressive Score (10.7%) together suggests a common failure pattern: salespeople intensify pressure when sensing prospect resistance, accelerating rather than reversing engagement deterioration. Acoustic features collectively account for approximately 21% of predictive importance (pitch variation 7.6%, voice energy 7.1%, pitch level 6.8%), demonstrating that delivery quality significantly influences rejection beyond content alone. The Unclear Value Score (6.5%) highlights the criticality of articulating concrete benefits without hedging language. These findings establish that rejection stems from combined communication, acoustic, and behavioral deficits rather than single failure points, validating the multimodal framework design and informing the coaching system’s emphasis on integrated quality indicators spanning all three modalities.

5. Discussion

5.1. Interpretation of findings

The empirical results suggest that successful and unsuccessful B2B sales conversations differ systematically across multiple dimensions. Three key patterns emerge from the analysis: structural efficiency in conversation organization, early predictability of outcomes, and multimodal indicators of communication quality.

The network analysis findings reveal that sales communication effectiveness manifests through structural efficiency. Importantly, in sequential utterance graphs, network density is mathematically coupled with conversation length. Therefore, the interpretation of efficiency does not derive from density alone but from the combined pattern of lower density, substantially higher phase coverage, lower centrality variance, and enriched value- and confirmation-phase transitions observed in successful calls. Successful salespeople maintain focused conversation trajectories with systematic topic development within critical phases, while unsuccessful conversations exhibit scattered exploration without sustained theme development. Effective selling involves sustained discussion within value proposition and confirmation phases rather than rapid movement between topics, building agreement through extended sequences rather than isolated commitment requests. This pattern suggests that conversation efficiency operates through focused depth rather than comprehensive breadth.

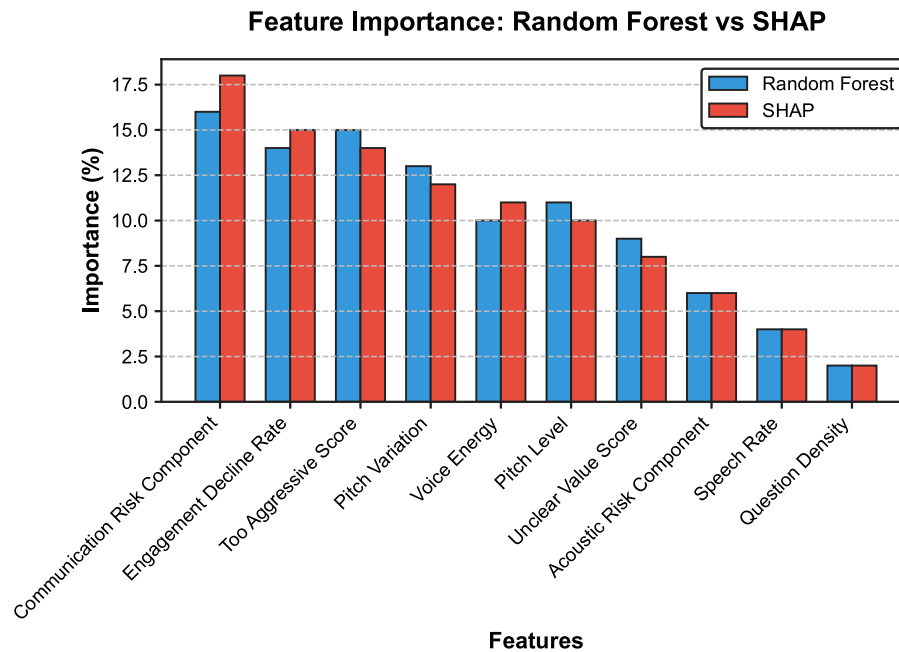


Fig. 7. Feature importance comparison results: Random Forest vs. SHAP. Strong correlation ($r = 0.94$) confirms consistent feature ranking across methodologies, validating the 95.7% AUC classification performance.

These structural findings are consistent with the stage progression analysis (Section 4.1), where successful calls demonstrated shorter opening phases and greater temporal allocation to discovery, proposal, and confirmation stages. The sustained value proposition and confirmation transitions observed in the network analysis provide a structural explanation for these stage allocation patterns, linking classical sales stage theory with graph-theoretic conversation structure.

The temporal prediction results (Section 4.3) demonstrate that conversation outcomes become predictable early in the interaction. Rather than implying causal psychological mechanisms, this finding indicates that the earliest phases of conversation contain the strongest discriminative signals for outcome prediction. The subsequent performance degradation at later time points suggests that early interaction phases contain the most informative behavioral patterns for classification. From a practical perspective, these findings establish an empirical intervention window where coaching systems can provide guidance based on reliable predictions while salespeople retain opportunity to adjust their approach. The temporal pattern indicates that coaching interventions achieve maximum effectiveness when deployed during early conversation phases when behavioral adjustment remains most feasible.

The rejection analysis (Section 4.4) reveals specific failure mechanisms within unsuccessful conversations. Communication risk dominates failure prediction, indicating that content quality and conversational approach constitute primary drivers of conversation breakdown—including aggressive language, unclear value articulation, and poor timing pressure. However, acoustic features collectively contribute substantial predictive importance, demonstrating that delivery quality significantly influences rejection beyond content alone. The prominence of engagement decline rate suggests a common failure pattern where salespeople intensify pressure when sensing resistance, accelerating rather than reversing disengagement. This multimodal failure signature indicates that conversation breakdown results from combined communication, acoustic, and behavioral deficits rather than single failure points, validating the integrated coaching framework design.

5.2. Comparison with prior work

This research extends prior work in multimodal conversation analysis, sales outcome prediction, and adaptive intervention timing by examining how multimodal communication patterns manifest within authentic B2B sales interactions.

Building on prior evidence demonstrating acoustic–semantic equivalence and the benefits of trimodal integration under duration-matched controls [17], the present study investigates how delivery characteristics, semantic content, and conversational structure appear within the structural and temporal organization of successful and unsuccessful sales calls. The observed patterns suggest that these modalities contribute complementary rather than redundant information: acoustic features reflect delivery characteristics associated with engagement and credibility; semantic features capture value-oriented content development; while linguistic features quantify conversation structure, providing the temporal framework within which acoustic and semantic signals achieve effective sequencing.

This multimodal performance aligns with established approaches in adjacent domains. Li et al. [67] demonstrated that combining acoustic and lexical features improves customer service sentiment analysis over single modalities. Prior work in persuasiveness prediction [68,69] established that verbal, acoustic, and visual features capture distinct aspects of communication effectiveness, with multimodal fusion consistently outperforming unimodal approaches. Zhang et al. [70] applied similar multimodal integration to deception detection, while Singla et al. [71] combined prosodic and lexical information for behavior coding in psychotherapy sessions. The current research extends this established multimodal paradigm to B2B sales contexts, demonstrating that communication effectiveness in high-stakes business interactions similarly manifests across multiple integrated channels rather than through any single dominant modality.

While traditional sales research established stage-based frameworks and identified key behaviors through observational methods [21–23], computational approaches to sales outcome prediction have remained limited. The current research operationalizes theoretical insights from classical sales literature through direct analysis of conversation audio and transcripts. The semantic network analysis reveals sustained value

proposition discussion and extended confirmation sequences in successful calls, providing quantitative support for theoretical arguments about thorough value development [25–27]. This translation of qualitative observations into measurable patterns enables automated coaching systems grounded in established sales theory.

The network analysis approach extends applications of graph theory to conversational data. Chai et al. [72] demonstrated that graph-theoretic metrics such as density, centrality, and clustering can characterize discussion dynamics in educational contexts. The current research applies similar graph-based methods to dyadic sales conversations, revealing that successful calls exhibit substantially lower network density despite longer duration. This counterintuitive finding extends prior work on conversation structure to sales contexts, demonstrating that focused conversation trajectories rather than comprehensive topic coverage characterize effective communication. Where educational contexts may benefit from broad exploration across topics, sales effectiveness appears to operate through sustained development of specific value propositions.

The identification of an early optimal intervention window extends research on just-in-time adaptive interventions. Fuller et al. [73] established canonical frameworks for timing adaptive interventions in health contexts, emphasizing the balance between predictive confidence and actionable opportunity. The current findings provide empirical evidence of an early intervention window in sales conversations, contrasting with traditional post-call analysis approaches. The temporal performance patterns suggest that fundamental communication dynamics establish themselves rapidly in sales interactions, with implications for real-time coaching system design. Where health interventions may operate across extended timeframes, sales coaching systems require immediate responsiveness to capitalize on brief windows of behavioral plasticity.

5.3. Bias audit and fairness evaluation

To evaluate whether the prediction framework performs equitably across different practitioners, a comprehensive bias audit was conducted using salesperson identity as the grouping variable. The audit analyzed 5164 calls with complete trimodal data (acoustic, linguistic, and semantic features) using the same 5-fold stratified cross-validation methodology employed throughout this research.

5.3.1. Audit procedure

The bias audit followed established fairness evaluation practices [74]:

1. Trained Random Forest classifiers using class-balanced weights to account for outcome imbalance
2. Generated out-of-fold predictions for all calls via 5-fold cross-validation
3. Grouped predictions by salesperson identity
4. Calculated per-group metrics: accuracy, true positive rate (TPR), false positive rate (FPR), and AUC
5. Computed disparity metrics as max–min differences across groups with sufficient sample sizes ($n \geq 50$)
6. Evaluated temporal stability across the six-month data collection period

5.3.2. Audit results

Table 4 presents the bias audit results for the trimodal prediction framework.

The framework demonstrates consistent ranking ability across all salespeople, with per-salesperson AUC ranging from 0.829 to 0.993—comparable to the trimodal AUC of 0.830 reported in the foundational research [17]. False positive rates remain below 3% across all groups, indicating that the model rarely misclassifies unsuccessful calls as successful regardless of salesperson.

Table 4

Bias audit results: Trimodal framework performance.

Metric	Value	Assessment
Overall AUC	0.955 ± 0.007	Strong discrimination
Per-salesperson AUC	0.829–0.993	Consistent across groups
FPR disparity	0.027	Within threshold (<0.15)
TPR disparity	0.321	See interpretation
Temporal disparity	0.069	Stable over time

The observed TPR disparity (0.321) reflects underlying data characteristics rather than algorithmic discrimination. Win rates vary substantially across salespeople (0.3% to 33.5%), creating fundamentally different prediction challenges for each group. Salespeople with very low win rates have fewer positive examples from which the model can learn group-specific patterns, resulting in lower recall for those groups. Importantly, the model maintains strong ranking ability (AUC > 0.82) and low false positive rates across all groups.

Temporal analysis confirmed stable performance across the six-month data collection period, with monthly AUC ranging from 0.919 to 0.988 (disparity = 0.069). This stability indicates that the identified patterns are robust to temporal variation in sales conditions.

5.3.3. Compliance-by-design provisions

The framework incorporates several design choices that support responsible deployment:

- **Objective features only:** Analysis relies exclusively on acoustic, semantic, and linguistic features—no demographic, biometric, or affective inference
- **Class-balanced training:** Training procedures account for outcome imbalance through balanced class weights
- **Advisory architecture:** The system provides decision support while preserving human autonomy over final coaching decisions
- **Transparent predictions:** All model outputs are interpretable through SHAP analysis, enabling practitioners to understand the basis for recommendations

These provisions align with EU AI Act requirements for high-risk AI systems used in employment-related contexts, ensuring that the framework can be deployed with appropriate oversight and accountability.

6. AI coaching system design

6.1. Evidence-based alert system

This system specification operationalizes the multimodal equivalence and weighting principles established in [17] using the structural and temporal findings of the present analysis. The alert system translates empirical findings into actionable coaching guidance through 15 calibrated quality indicators organized across four categories: acoustic quality (5 indicators: pitch stability, volume consistency, speaking pace, pause quality, voice energy), semantic quality (4 indicators: value persistence, value steering, value dominance, value confirmation), linguistic quality (3 indicators: network density, transition quality, temporal confidence), and integrated multimodal metrics (3 indicators: co-dominance score, multimodal quality index, early warning composite score). Each indicator provides color-coded performance zones (Green/Yellow/Red) with specific coaching priorities derived from empirical thresholds. Complete threshold specifications are provided in Supplementary Materials Section 2, Table S1.

Decision-Support Mechanism: The alert system translates quantitative indicators into decision support through a three-stage process: (1) *System Output*—color-coded alerts (Green/Yellow/Red) signal performance levels across acoustic, semantic, and linguistic dimensions; (2) *Salesperson Decision*—coaching priorities guide immediate corrective

actions (e.g., “Stabilize voice tone” when $PSI < 0.6$, “Focus on value” when $VPI < 10$); (3) *Business Benefit*—corrections address empirically validated success discriminators (network density $\rho = 0.0224$ vs. 0.096 , $p < 0.001$; rejection risk $AUC = 95.7\%$), addressing patterns associated with conversation failure. For example, when EWCS enters Red Zone (>0.7), the system signals high rejection risk, prompting the salesperson to employ value-steering and confirmation-building strategies validated in successful calls, aligning behavior with patterns observed in successful calls. Post-call analytics enable systematic performance tracking, identifying persistent weaknesses for targeted training interventions.

6.2. Validation framework

This research employs the comprehensive validation methodology established for this data set in the foundational research [17], including stratified 5-fold cross-validation, duration-matched controls, and cross-algorithm consistency testing. Network analysis used Mann–Whitney U tests with Bonferroni correction for multiple comparisons, achieving highly significant differences ($p < 0.001$) with large effect sizes (Cohen’s $d = 1.16$ – 1.78). SHAP analysis validated the 95.7% AUC rejection prediction performance, confirming genuine predictive patterns rather than overfitting artifacts.

- **Validated Performance Metrics:** 83.0% AUC and 74.4% F1 score under duration-matched validation
- **Equal Weighting:** Constrained optimization reveals that approximately equal weighting (1/3 each) across acoustic, semantic, and linguistic modalities maximizes ensemble effectiveness
- **Multi-Metric Robustness:** Acoustic-semantic equivalence confirmed across AUC (0.5% difference) and F1 (0.2% difference)
- **Cross-Algorithm Consistency:** Validated across Random Forest, XGBoost [75], and Logistic Regression
- **Theoretical Architecture:** Design specifications for potential real-time applications pending implementation validation

Implementation recommendations for phased organizational roll-out, including pilot testing protocols, training integration strategies, and continuous optimization procedures, are provided in Supplementary Materials Section 4.

7. Conclusion

This research developed and empirically validated an AI-assisted coaching framework for B2B sales conversations through multimodal analysis integrating acoustic delivery, semantic content, and linguistic structure. Network analysis established that successful conversations achieve efficiency through focused progression rather than comprehensive topic coverage, revealing systematic structural differences between successful and unsuccessful interactions. Temporal prediction modeling identified an early optimal intervention window that enables coaching systems to provide reliable guidance while salespeople retain opportunity for behavioral adjustment. Rejection pattern analysis with interpretability validation confirmed that conversation failure stems from combined communication, acoustic, and behavioral factors rather than single failure points. These empirical findings were operationalized into evidence-based quality indicators with calibrated performance thresholds, human-centered advisory architecture, and regulatory compliance provisions aligned with the EU AI Act.

Future research should pursue four key directions to extend and validate this framework. Prospective field studies deploying the system in controlled organizational settings with randomized assignment would measure actual win-rate improvement, behavior change trajectories, and user adoption patterns through longitudinal tracking. Cross-language validation studies should test whether core structural findings – network efficiency patterns, optimal intervention timing, and acoustic-semantic equivalence – generalize across linguistic and

cultural contexts or require context-specific recalibration. Industry-specific threshold refinement analyzing communication patterns across distinct verticals (healthcare, technology, manufacturing) would determine whether sector-specific calibrations improve prediction accuracy beyond the general framework. Finally, integration research developing standardized APIs connecting the framework to existing CRM platforms, call recording systems, and learning management tools would enable seamless deployment and reduce implementation friction for organizations adopting evidence-based coaching systems.

7.1. Limitations and future research opportunities

While this research establishes comprehensive evidence-based frameworks for AI-assisted sales coaching, several characteristics of the current study present opportunities for future investigation:

Language and Cultural Generalization: The present analysis focuses exclusively on Finnish B2B sales conversations. While the mathematical framework provides language-agnostic architectural specifications, empirical parameters (e.g., optimal speaking rate of 123 WPM, semantic transition patterns) require calibration for different linguistic and cultural contexts. Future research extending this framework to additional languages would strengthen cross-cultural validity while testing the generalizability of core structural findings such as network efficiency advantages.

Industry Context Expansion: Current findings derive from B2B sales interactions within a specific organizational context. Extending validation to diverse industry verticals (healthcare, technology, manufacturing) would test whether identified patterns represent universal sales communication principles or context-specific phenomena. Such expansion would enable development of industry-specific threshold calibrations while identifying domain-invariant coaching principles.

Deployment Validation: The framework provides theoretical specifications and empirical foundations but requires controlled implementation studies to validate operational performance. Real-world deployment testing would assess system response times, alert threshold calibration under production conditions, and actual coaching effectiveness through randomized controlled trials. Such validation studies represent essential next steps for translating analytical findings into measurable business outcomes.

Longitudinal Performance Tracking: Current cross-sectional analysis identifies patterns distinguishing successful from unsuccessful conversations but does not track individual salesperson improvement trajectories. Longitudinal studies monitoring coaching intervention effects over time would quantify learning curves, measure sustained behavior change, and identify optimal training protocols for different skill development stages.

These limitations simultaneously define boundaries of current findings and establish clear directions for continued investigation. The validated analytical framework, comprehensive mathematical specifications, and robust empirical evidence provide infrastructure upon which future research can systematically address these opportunities while expanding scientific understanding of AI-assisted sales coaching systems.

7.2. Design science contribution and impact

This research demonstrates how multimodal communication analytics can be systematically integrated into deployable coaching frameworks through validated design specifications. By establishing empirically grounded thresholds, optimal intervention timing, and computationally efficient mathematical formulations, the framework bridges the gap between communication science and operational decision support systems. The design prioritizes interpretability, regulatory compliance, and human autonomy—translating abstract performance concepts into concrete, measurable indicators that support professional development while preserving individual decision-making agency.

The transition from subjective evaluation to objective optimization offers potential advancement in sales performance management through coaching systems that genuinely support salesperson development and empowerment. The framework's advisory rather than prescriptive architecture ensures that coaching recommendations enhance human capability rather than replacing human judgment. The evidence-based thresholds and multimodal indicators enable self-directed improvement and skill development through objective feedback on communication effectiveness.

Implementation studies building on these validated design specifications can translate theoretical frameworks into operational systems that respect human autonomy, promote continuous learning, and support systematic skill development. The comprehensive architectural specifications – including performance thresholds, intervention timing, mathematical formulations, and regulatory compliance provisions – provide a foundation for organizations to develop empowering, development-focused coaching systems that enhance salesperson capability and professional growth.

CRediT authorship contribution statement

Joona Mäntyvaara: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Paavo Nevalainen:** Writing – review & editing, Validation. **Kirill Glavatskiy:** Writing – review & editing, Validation. **Jukka Heikkonen:** Supervision, Project administration, Funding acquisition.

Ethics statement

This research was conducted as part of the PATA (AI-Enabled Customer Experiences) project, a Business Finland co-funded research initiative with Haaga-Helia University of Applied Sciences and University of Turku as joint data controllers. The study analyzed sales call recordings collected during normal business operations by participating organizations.

Legal basis and consent: Personal data processing was conducted under GDPR Article 6(1)(a) based on explicit consent. Each participating salesperson provided written informed consent specifically authorizing the processing, analysis, and research use of their call recordings for AI model development and sales technology research. Consent forms documented that participation was voluntary, that consent could be withdrawn at any time without providing reasons, and that withdrawal would result in deletion of unprocessed recordings. Participants were provided with a full privacy notice describing processing activities, data controller contacts, retention periods, and their rights under GDPR Articles 13–14.

Customer privacy protection: To protect customer privacy under GDPR, the participating organization removed the customer audio channel (R-channel) from stereo recordings prior to research team access. Only the salesperson audio channel (L-channel) was included in the research dataset. This design ensures no customer personal data was processed or analyzed in this research.

De-identification: Before transfer to the research team, salesperson names were removed from recording filenames and replaced with numerical identifiers. All analysis was conducted through automated machine-based processing; researchers did not listen to recordings manually. Results are reported at aggregate level only, and individual participants cannot be identified in published outputs.

Data protection and retention: Data were stored on University of Turku's secure Seafile cloud storage system within Finland. No data transfers occurred outside the EU/EEA. Recordings will be deleted upon completion of analysis or by project end (December 2026), whichever occurs first; only aggregated statistical results will be retained.

Funding

Business Finland supported this co-research project. Funding number: 6684/31/2023. The funding institution had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank participating sales organizations for providing authentic B2B conversation data and the TurkuNLP team for FinBERT development. This research was part of the PATA co-research project (AI-enabled customer experiences), in cooperation with the University of Turku, Finland, the Haaga-Helia University of Applied Sciences, Finland, and the University of Newcastle, Australia.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.array.2026.100755>.

Data and documentation availability

Restricted materials: Raw audio recordings, verbatim transcripts, and feature embeddings cannot be shared due to non-disclosure agreements with participating sales organizations and GDPR requirements protecting personally identifiable information. Feature vectors may contain patterns enabling salesperson re-identification and therefore remain restricted under the same confidentiality provisions as source data.

Available materials: The following resources are available to researchers: (1) Python analysis scripts for network analysis, temporal prediction, and rejection modeling; (2) model architecture specifications and hyperparameter configurations; (3) complete mathematical formulations in Supplementary Materials; (4) methodology documentation describing data processing procedures.

Reproducibility specifications: Key software versions and configurations: Python 3.11, scikit-learn 1.6.1 (Random Forest: `n_estimators = 100`, `class_weight='balanced'`, `random_state=42`; 5-fold stratified cross-validation), transformers 4.55.2 with TurkuNLP/bert-base-finnish-cased-v1 for semantic embeddings (768-dimensional), librosa 0.11.0 with PYIN pitch tracking (`fmin=50Hz`, `fmax=400Hz`, `frame_length=2048`, `hop_length=512`). Complete hyperparameter tables and pseudocode for data preparation pipelines are provided in Supplementary Materials.

Access procedure: Analysis code and specifications are available from the corresponding author upon reasonable request for non-commercial research purposes. Requests should include institutional affiliation and intended research use. Expected response time is 10 business days. The methodology documentation is publicly available: Acoustic–Semantic Equivalence Methodology and Replication Guide.pdf

<https://seafile.utu.fi/f/1786a1fc71014bf8a25f/>.

References

- [1] Rich GA. The constructs of sales coaching: Supervisory feedback, role modeling, and trust. *J Pers Sell Sales Manag* 1998;18(1):53–63.
- [2] Onyemah V. The effects of coaching on salespeople's attitudes and behaviors: A contingency approach. *Eur J Mark* 2009;43(7/8):938–60.
- [3] Ingram TN, LaForge RW. *Sales management: Analysis and decision making*. Routledge; 2014.
- [4] Kopp S, Krämer NC, Gratch J. The social companion: Modeling interpersonal communication in human-AI interaction. *Comput Linguist* 2018;44(1):25–58.
- [5] Cron WL, Marshall GW, Singh J, Spiro RL, Sujana H. Salesperson selection, training, and development: Trends and implications. *J Pers Sell Sales Manag* 2005;25(2):123–36.
- [6] Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J. Challenging games help students learn: An empirical study on engagement, flow, and learning. *Comput Hum Behav* 2016;54:170–9.
- [7] Seaborn K, Fels DI. Gamification in theory and action: A survey. *Int J Hum-Comput Stud* 2015;74:14–31.
- [8] Dixon M, Adamson B. *The challenger sale: Taking control of the customer conversation*. Portfolio/Penguin; 2011.
- [9] Gartner, Inc. *Market guide for conversation intelligence applications in sales*. 2023, <https://www.gartner.com/>.
- [10] Peterson RA, Cannito MP, Brown SP. An exploratory investigation of voice characteristics and selling effectiveness. *J Pers Sell Sales Manag* 1995;15(1):1–15. <http://dx.doi.org/10.1080/08853134.1995.10754008>.
- [11] Jokiniemi S, Halinen A, Pullins EB, Hosoi K. Rapport building in B2B sales interactions: The process and explananda. *J Pers Sell Sales Manag* 2023. <http://dx.doi.org/10.1080/08853134.2023.2216935>.
- [12] Levinson SC. Turn-taking in human communication—Origins and implications for language processing. *Trends Cogn Sci* 2016;20(1):6–14. <http://dx.doi.org/10.1016/j.tics.2015.10.010>.
- [13] Nahum-Shani I, Smith SN, Spring B, Collins LM. Just-in-time adaptive interventions (JITAs) in mobile health: Key components and future directions. *Ann Behav Med* 2022;56(6):549–62. <http://dx.doi.org/10.1093/abm/kaac004>.
- [14] Koedinger KR, Kim J, Jia JZ, McLaughlin EA, Bier NL. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In: *Proceedings of the 2nd ACM conference on learning at scale*. 2013, p. 111–20. <http://dx.doi.org/10.1145/2460296.2460355>.
- [15] European Commission. *Regulation (EU) 2024/1689 of the European parliament and of the council laying down harmonised rules on artificial intelligence (AI act)*. 2024, <https://eur-lex.europa.eu/>.
- [16] Amershi S, Weld DS, Vorvoreanu M, Fournier A, Nushi B, Collisson P, Suh J, Iqbal ST, Bennett PN, Inkpen K, Teevan J. Guidelines for human-AI interaction. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, p. 1–13. <http://dx.doi.org/10.1145/3290605.3300233>.
- [17] Authors. *Acoustic-Semantic equivalence: A dual-channel framework for B2B sales communication effectiveness and training allocation*. *J Bus Res* 2025. submitted for publication.
- [18] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. *MIS Q* 2004;28(1):75–105. <http://dx.doi.org/10.2307/25148625>.
- [19] Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. *J Manage Inf Syst* 2007;24(3):45–77. <http://dx.doi.org/10.2753/MIS0742-1222240302>.
- [20] Gregor S, Hevner AR. Positioning and presenting design science research for maximum impact. *MIS Q* 2013;37(2):337–55. <http://dx.doi.org/10.25300/MISQ/2013/37.2.01>.
- [21] Dodge HR. *An integrative model of the personal selling process*. *South Bus Rev* 1989;14(2):1–8.
- [22] Moncrief WC, Marshall GW. The evolution of the seven steps of selling. *Ind Mark Manag* 2005;34(1):13–22. <http://dx.doi.org/10.1016/j.indmarman.2004.06.001>, URL <https://www.sciencedirect.com/science/article/pii/S0019850104000525>.
- [23] Rackham N. *SPIN selling*. McGraw-Hill Education; 1988.
- [24] Steward MD, Narus JA, Roehm ML, Ritz W. From transactions to journeys and beyond: The evolution of B2B buying process modeling. *Ind Mark Manag* 2019;83:288–300. <http://dx.doi.org/10.1016/j.indmarman.2019.05.002>.
- [25] Terho H, Haas A, Eggert A, Ulaga W. It's almost like taking the sales out of selling—Towards a conceptualization of value-based selling in business markets. *Ind Mark Manag* 2012;41(1):174–85. <http://dx.doi.org/10.1016/j.indmarman.2011.11.011>.
- [26] Terho H, Eggert A, Haas A, Ulaga W. How salespeople make value-based selling happen in business markets. *J Pers Sell Sales Manag* 2015;35(1):42–60. <http://dx.doi.org/10.1080/08853134.2014.943416>.
- [27] Terho H, Haas A, Eggert A, Ulaga W. Linking value-based selling to sales performance: A multilevel investigation. *J Acad Mark Sci* 2017;45(2):133–50. <http://dx.doi.org/10.1007/s11747-015-0431-4>.
- [28] Cialdini RB. *Influence: The psychology of persuasion*. New York: HarperCollins; 1984.
- [29] Pöyry EI, Parvinen P, McFarland RG. Generating leads with sequential persuasion: Should sales influence tactics be consistent or complementary? *J Pers Sell Sales Manag* 2017;37(2):89–99. <http://dx.doi.org/10.1080/08853134.2016.1275973>.
- [30] Niemi J, Hirvonen L. Money talks: Customer-initiated price negotiation in business-to-business sales interaction. *Discourse Commun* 2019;13(1):95–118. <http://dx.doi.org/10.1177/1750481318801629>.
- [31] De Haan E, Nilsson VO. *Psychology of executive coaching: Theory and application*. Routledge; 2022.
- [32] Noe RA, Clarke ADM, Klein HJ. Learning in the twenty-first-century workplace: Perspectives on the future of human resource development. *Annu Rev Organ Psychol Organ Behav* 2017;4:245–75. <http://dx.doi.org/10.1146/annurev-orgpsych-032516-113108>.
- [33] Intille SS. A new research challenge: Persuasive technology to motivate healthy aging. *IEEE Trans Inf Technol Biomed* 2004;8(3):235–7. <http://dx.doi.org/10.1109/TITB.2004.834384>.
- [34] Riedl MO, Harrison B. Using AI to influence human decision-making in real time. *AI Mag* 2020;41(2):27–39. <http://dx.doi.org/10.1609/aimag.v41i2.5263>.
- [35] Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;119(2):254–84. <http://dx.doi.org/10.1037/0033-2909.119.2.254>.
- [36] London M. *The power of feedback: Giving, seeking, and using feedback for performance improvement*. Routledge; 2015.
- [37] Wasserman S, Faust K. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press; 1994, <http://dx.doi.org/10.1017/CBO9780511815478>.
- [38] Scott J. *Social network analysis*. 3rd ed.. London: SAGE Publications; 2011.
- [39] Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *Science* 2009;323(5916):892–5. <http://dx.doi.org/10.1126/science.1165821>.
- [40] Stolcke A, Ries K, Coccaro N, Shriberg E, Bates R, Jurafsky D, Taylor P, Martin R, Van Ess-Dykema C, Meteer M. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Comput Linguist* 2000;26(3):339–73. <http://dx.doi.org/10.1162/089120100561737>.
- [41] Young S, Gašić M, Thomson B, Williams JD. POMDP-based statistical spoken dialog systems: A review. *Proc IEEE* 2013;101(5):1160–79. <http://dx.doi.org/10.1109/JPROC.2012.2225812>.
- [42] Chen H, Liu X, Yin D, Tang J. A survey on dialogue systems: Recent advances and new frontiers. *Acem Sigkdd Explor Newsl* 2017;19(2):25–35.
- [43] Gao J, Galley M, Li L. Neural approaches to conversational AI. In: *The 41st international ACM SIGIR conference on research & development in information retrieval*. 2018, p. 1371–4.
- [44] Raheja V, Tetreault J. Dialogue act classification with context-aware self-attention. 2019, arXiv preprint [arXiv:1904.02594](https://arxiv.org/abs/1904.02594).
- [45] Murray G, Renals S, Carletta J. Extractive summarization of meeting recordings. In: *Proceedings of eurospeech*. Lisbon, Portugal; 2005, p. 593–6.
- [46] Gama J, Žliobaite I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. *ACM Comput Surv* 2014;46(4):1–37.
- [47] Bifet A, Holmes G, Kirkby R, Pfahringer B. MOA: Massive online analysis for data stream mining. *J Mach Learn Res* 2010;11:1601–4.
- [48] Hulten G, Spencer L, Domingos P. Mining time-changing data streams. In: *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*. 2001, p. 97–106.
- [49] Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?”: Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016, p. 1135–44.
- [50] Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: *Proceedings of the 34th international conference on machine learning*. ICML, *Proceedings of machine learning research*, vol. 70, PMLR; 2017, p. 1321–30.
- [51] Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform* 2016;3(2):119–31.
- [52] Radford A, Kim JW, Xu T, Brockman G, McLevey C, Sutskever I. Robust speech recognition via large-scale weak supervision. 2023, arXiv preprint [arXiv:2212.04356](https://arxiv.org/abs/2212.04356).
- [53] Stolcke A, Ries K, Coccaro N, Shriberg E, Bates R, Jurafsky D, Taylor P, Martin R, Ess-Dykema M, Meteer C. Dialogue act modeling for automatic tagging and recognition of conversational speech. In: *Proceedings of the 1st meeting of the North American chapter of the association for computational linguistics*. 2000, p. 1–8. <http://dx.doi.org/10.3115/1075218.1075274>.
- [54] Jurafsky D, Martin JH. *Speech and language processing*. 3rd ed.. Prentice Hall; 2023, URL <https://web.stanford.edu/~jurafsky/slp3/>. Draft version available online.
- [55] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947;18(1):50–60. <http://dx.doi.org/10.1214/aoms/1177730491>.
- [56] Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed.. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- [57] Virtanen A, Kanerva J, Ilo R, Luoma J, Luotolahti J, Salakoski T, Ginter F, Pyysalo S. Multilingual is not enough: BERT for finnish. 2019, arXiv preprint [arXiv:1912.07076](https://arxiv.org/abs/1912.07076). Preprint.

- [58] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAAACL-HLT 2019. 2019, p. 4171–86.
- [59] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32. <http://dx.doi.org/10.1023/A:1010933404324>.
- [60] Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001;29(5):1189–232. <http://dx.doi.org/10.1214/aos/1013203451>.
- [61] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed.. Springer series in statistics, Springer; 2009, <http://dx.doi.org/10.1007/978-0-387-84858-7>.
- [62] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016, URL <http://www.deeplearningbook.org>.
- [63] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986;323(6088):533–6. <http://dx.doi.org/10.1038/323533a0>.
- [64] Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27(8):861–74. <http://dx.doi.org/10.1016/j.patrec.2005.10.010>.
- [65] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;30.
- [66] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence. *IJCAI*, Vol. 2, 1995, p. 1137–43.
- [67] Li R, Wu Z, Stolcke A, Metze F. Acoustic and lexical sentiment analysis for customer service calls. In: ICASSP 2019 - IEEE international conference on acoustics, speech and signal processing. IEEE; 2019, p. 7305–9. <http://dx.doi.org/10.1109/ICASSP.2019.8683503>.
- [68] Park S, Shim HS, Chatterjee M, Sagae K, Morency L-P. Multimodal analysis and prediction of persuasiveness in online social multimedia. *ACM Trans Interact Intell Syst (TiiS)* 2016;6(3):1–25. <http://dx.doi.org/10.1145/2982142>.
- [69] Nojavanasghari B, Baltrušaitis T, Hughes C, Morency L-P. Deep multimodal fusion for persuasiveness prediction. In: Proceedings of the 18th ACM international conference on multimodal interaction. *ICMI*, ACM; 2016, p. 284–8. <http://dx.doi.org/10.1145/2993148.2993198>.
- [70] Zhang Z, Levitan R, Hirschberg J. Multimodal deception detection using automatically extracted acoustic, visual, and lexical features. In: Proceedings of interspeech 2020. ISCA; 2020, p. 435–9. <http://dx.doi.org/10.21437/Interspeech.2020-2257>.
- [71] Singla K, Chen Z, Peri R, Gibson J, Narayanan S. Using prosodic and lexical information for learning utterance-level behaviors in psychotherapy conversations. *IEEE J Sel Top Signal Process* 2019;13(2):449–63. <http://dx.doi.org/10.1109/JSTSP.2019.2896466>.
- [72] Chai L, Li M, Wang X. Applying graph theory to examine the dynamics of student discussions. *CBE Life Sci Educ* 2019;18(3):ar38. <http://dx.doi.org/10.1187/cbe.18-10-0195>.
- [73] Fuller D, Spring B, Pellegrini C. Just-in-time adaptive interventions (JITAs) in mobile health: Key components and design principles. *Ann Behav Med* 2018;52(6):446–62. <http://dx.doi.org/10.1093/abm/kax033>.
- [74] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv* 2021;54(6):1–35. <http://dx.doi.org/10.1145/3457607>.
- [75] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '16, New York, NY, USA: Association for Computing Machinery; 2016, p. 785–94. <http://dx.doi.org/10.1145/2939672.2939785>.

Joona Mäntyvaara is a Doctoral Researcher at the Department of Computing, University of Turku, and a Sales Lecturer and Sales AI Researcher at Haaga-Helia University of Applied Sciences, Finland. His research investigates how artificial intelligence advances business-to-business (B2B) sales operations by enhancing communication effectiveness, decision-making, and performance management through human-centered and evidence-based approaches. With over two decades of experience in B2B sales and marketing across technology and professional services sectors, he manages the Business Finland-funded PATA project, developing AI-driven frameworks for sales operations. His work emphasizes the responsible integration of AI that augments – rather than replaces – human expertise in professional selling.