Journal of Information Systems and Telecommunication
Special Issue, Winter,2026, PP 60-69
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Enhancing Industrial Interaction Practices Through AI-Based Parameter Modeling

Ashwini Kumar[1*], Rekha Agrawal[1], Archana Singh[2]

[1].Amity Institute of Information Technology, Amity University, Uttar Pradesh, Noida, India
[2].Ministry of Education, Government of India

Received: 25 Jul 2025/ Revised: 04 Dec 2025/ Accepted: 06 Jan 2026

Abstract

Industrial systems today depend increasingly on effective communication and coordination among humans and machines. This study proposes an Artificial Intelligence–based approach for modeling and improving industrial interaction practices using the COFI framework—Context, Content, Competency, and Culture. By combining supervised and unsupervised learning techniques, specifically Random Forest (RF) and K-Means clustering, the research models key parameters that influence communication efficiency and organizational alignment. A publicly available behavioral dataset, supplemented with simulated industrial communication records, was used to represent multi-agent interactions within a workplace context. Extensive data preprocessing, feature engineering, and COFI-based variable mapping were performed to ensure interpretability and conceptual coherence. The RF model achieved an improved predictive accuracy of 72.4% following feature optimization, while K-Means clustering produced three distinct communication groups with a Silhouette score of 0.75 and a Davies–Bouldin Index of 0.49, indicating well-separated clusters. Feature-importance and SHAP analyses revealed that contextual and content-based variables contributed most significantly to prediction outcomes, while competency and cultural attributes shaped nuanced interaction patterns. A pilot case simulation demonstrated tangible performance improvements—reducing response time by 12% and improving task resolution by 9% when AI insights were applied to industrial communication workflows. The findings confirm that combining supervised prediction with unsupervised segmentation offers a robust pathway to understanding and optimizing human–machine communication within organizational ecosystems. This research contributes a practical and interpretable framework for AI-enabled industrial interaction modeling, offering both theoretical insight and applied value for adaptive, data-driven management systems.

Keywords: COFI Framework; Random Forest; K-Means Clustering; Industry Interactions; Predictive Modeling; Human-AI Collaboration; Industrial Optimization; Segmentation Techniques; AI System Architecture.

1- Introduction

In this age of industrial digitalization and automation where the impact of effective communication and adaptability is crucial [1], the idea of utilizing new technologies and methods such as Artificial Intelligence (AI) and Machine Learning (ML) to reshape productive industrial interaction regimes enables systems to become streamlined, sensible, and self-aware. In this paper, entitled "An AI Perspective on Modeling Key Parameters to Improve Interaction Practices in Industry", it raises awareness and promotes a need to evolve interaction strategies by encompassing AI oriented techniques, specifically analyzing the dynamic and probabilistic variables that exist within industrial environments [2].

Data-centric operations are growing in importance throughout industry, and this encourages the development of systems capable of handling large, growing data sets. Older patterns of interaction often relied on static processes, Rule-based communication flows, normative methodologies. These characteristics are increasingly inadequate to support an industry environment which is not only changing rapidly but completely transforming the relationship between businesses and consumers. In navigating these complexities and uncertainties, intelligent systems with empirical data enabled by artificial intelligence (AI) and machine learning (ML) allow for the modeling of common behavior patterns and will enhance

✉ **Ashwinil Kumar**
mrashwinikumar@gmail.com

the opportunity for proactively predicting, and responding to emergent interaction needs.

These developments are especially evident in areas like manufacturing, finance, health care, and retail where human-machine interactions affect collective productivity and individual satisfaction across organizations and stakeholder ecosystems. The COFI framework (an acronym for Context, Content, Competency, and Culture) was developed to take an application-oriented approach to managing interaction dynamics in organizational ecosystems [3]. Each dimension of the COFI framework has a unique proposition for the interaction model:

- Context is all about the situational environment at the time of the interactions and how this influences the course of interactions and therefore communication processes.
- Content relates to the kind of information shared and its quality, which can affect the richness and engagement of interaction.
- Competency relates to the skills and capabilities required to communicate effectively and expresses levels of ability.
- Culture refers to the values, beliefs, and learned behaviors which are particular to a given industrial context and shape the communication process.

The study proposes using the COFI model to provide a more complete view of how AI can be leveraged to enhance interaction practices, taking full account of all four elements - rather than treating them separately and resulting in fragmentation (as with traditional models). The study provides for an industrial view and activity that is integrated with AI to develop a better model, measurement, and improvement of interactions in ways not currently available [4]. In this vein, the study draws upon a dual-method approach that incorporates supervised learning (via the Random Forest classifier) and unsupervised learning (via K-Means clustering). The Random Forest (RF) model is especially relevant for identifying patterns among massive high-dimensional datasets while providing relatively high predictive accuracy on the test data. Its ensemble approach reduces bias while guarding against overfitting, making it suitable for matrix outputs related to distinct identifiable elements of varied interaction data, such as transaction data records, support log data, and behavioral interaction patterns. Ultimately, our results suggest that RF may be an appropriate approach for modeling real-time interactions in similar scenarios with comparable accuracy, while providing adequate actionable insights necessary to improve patterns of engagement practice [5].

K-means clustering, on the other hand, is a valuable unsupervised learning method for revealing and categorizing patterns, when none exists, often in the absence of labeled data. It allows for behavior similar interaction forms of clustering (customer queries, for example, by topic), thereby enabling businesses to assign priority based on categorization and develop a unique and specific response for the customer. Some of the value in using K-means clustering in conjunction with the COFI framework is to support mapping latent patterns across identifying interaction parameters such as, for example lead and end events or communicative themes, to support better organizational alignment with meeting client expectations and their needs. This comparative approach allows us to see some unique advantages of each of the techniques: RF's strong predictive abilities and precision, and K-Means' ability to discover patterns of structure and latent trends. A combined approach provides a comprehensive perspective on interaction strategies and the relevant processes for informed decision-making. In the end, this research demonstrates the disruptive capabilities of AI to change traditional industrial communication modalities into flexible, real-time, and data-driven systems in accordance with the COFI process model. In comparing and integrating the utility of supervised and unsupervised models, we help advance the understanding of improving industrial interaction efficiencies. The research outlines opportunities for future inquiry to discover how to take advantage of hybrid AI methods for addressing multifaceted interaction problems across sectors.

Finally, the study illustrates that integrating an AI component to structured processes - such as COFI - can both augment the appropriateness of industrial interaction practices, while also providing a scalable path for dealing with new communication demands presented by the digital economy [6,7].

This study aims to understand the ways Artificial Intelligence (AI) methods can be applied to design and improve the interaction practices in industry using the COFI framework Context, Content, Competency, and Culture. Although the current models of industrial communication tend to rely on a rule-based or descriptive approach, they are unlikely to reflect the dynamic and data-driven essence of the current organizational ecosystems. Thus, the current study will take a two-fold AI strategy, combining supervised learning (Random Forest) and unsupervised learning (K-Means clustering) to simulate the interaction patterns, evaluate the effectiveness of communication and determine patterns that can be applied into practice.

The research questions featured in this study are the following:

RQ1: What is the way to quantitatively model the COFI framework with the help of AI-based techniques to model and forecast patterns of industrial interaction?

RQ2: How do the four dimensions of COFI of Context, Content, Competency, and Culture relate to the quality and effectiveness of communication in industrial systems?

RQ3: Does the integration of supervised and unsupervised variants of learning offer more profound insights into communication structures as compared to the insights obtained using one of these two approaches exclusively?

On the basis of these questions, the following hypotheses are put:

H1: Machine learning models with the inclusion of features based on COFI will make the prediction of the outcomes of industrial interaction significantly more accurate.

H2: Competency and cultural factors will have less predictive power than contextual and content variables.

H3: A hybridizing modeling methodology (RF + K-Means) will expose different clusters of interaction that will be in agreement with the real-life operation divisions in industrial communication.

Through the hypotheses, the research will help to present empirical contributions to the hypothesis that AI-driven, framework-based modeling is a potential direction to improve industrial communication and decision-making processes

2- Methodology Overview

This study followed a protocol consisting of several steps including data preprocessing, experimenting with models, comparing models, tuning, and validating to support interaction modeling under the COFI framework using AI techniques.

One of the most important aspects of the selection of an adequate dataset is the need to ensure that the output of any AI model is representative of the target environment. The publicly available insurance data of Kaggle was taken as a representative proxy of the industrial communication data used in this study. This choice was made by the behavioral and structural resemblances between the customer-agent interactions in the records of insurance and communication exchanges in the industrial ecosystems. Both forms of data follow a pattern of multi-agent communication, hierarchies of escalations, variable contextual variation, and quantifiable consequences such as problem solving or turnaround time. The common nature of these properties results in the fact that the dataset is appropriate to be used in modelling interaction dynamics within an organizational environment. As a case in point, time-based variables (timestamps, delay intervals) are similar to Context in the COFI framework; text (message length, sentiment) are similar to Content; agent identifiers and performance indicators are similar to Competency; and regional or departmental indicators are similar to Culture.

2-1- Data Loading and Preprocessing

An insurance dataset from Kaggle was provided to simulate user interaction behaviour. The use of libraries, such as Pandas, NumPy, Seaborn, and Matplotlib allowed for a clean import and visualization of the data. Categorical variables were then encoded and various feature scaling and standardization techniques were rehearsed to achieve uniformity for compatibility with models, [8-10].

	Description	Value
0	Session id	1643
1	Original data shape	(1338, 13)
2	Transformed data shape	(1338, 15)
3	Ordinal features	2
4	Numeric features	10
5	Categorical features	3
6	Rows with missing values	3.8%
7	Preprocess	True
8	Imputation type	simple
9	Numeric imputation	mean
10	Categorical imputation	mode
11	Maximum one-hot encoding	-1
12	Encoding method	None
13	CPU Jobs	-1
14	Use GPU	False
15	Log Experiment	False
16	Experiment Name	cluster-default-name
17	USI	56ad

Fig. 1 Data Preprocessing Summary and Configuration Details

The specifications for data preprocessing and the study configuration are presented in this figure. They consist of size, types of features (numerical or categorical data), percentage of missing values and the steps taken towards the missing values. Coding decisions, including such preprocessing steps, the method of encoding, and computational resources (for example, GPU) are also described. Such configurations remain pertinent because they help prepare the dataset for model training and evaluation [11].

2-2- Data Cleaning and Imputation

To assure the quality and consistency of the dataset outliers and duplicate rows were detected and eliminated using z-score analysis. Inconsistent formatting of entries was also corrected to standardize formatting between variables. If a variable had missing values, imputation was applied using the mean, median, or mode, depending on the variable, which did not compromise the integrity of the dataset [12].

2-3- Model Selection and Comparison

Several machine learning models were explored to evaluate their performance in the context of human-AI interaction analysis. Random Forest (RF) was selected for its high accuracy and robustness, while K-Means clustering was applied to perform unsupervised segmentation of interaction types. Logistic regression and decision trees were also tested as baseline models for comparison [13]. Model performance was assessed using metrics such as accuracy, precision, and F1-score. K-Means achieved 37% accuracy, mainly serving as a tool for segmentation, whereas RF outperformed all other models in predictive performance.

2-4- Model Tuning and Validation

The RF model underwent fine-tuning through grid search to optimize key hyperparameters, including the number of estimators, maximum tree depth, minimum samples required to split a node, and the minimum number of samples at a leaf node. Cross-validation techniques were employed to ensure generalizability and prevent overfitting. The final model evaluation was conducted on a separate test dataset, confirming its robustness and predictive reliability [14].

2-5- Evaluation and Results

Performance metrics (accuracy, precision, recall, F1) were used to compare models in the context of the COFI framework. RF delivered the highest precision, proving suitable for modeling industrial interaction practices. K-Means helped identify clusters but lacked predictive strength.

Figure 1.2 demonstrates the PyCaret-based setup for K-Means and evaluation steps. The comprehensive analysis confirms that Random Forest is best suited for precision-focused interaction modeling, laying a strong foundation for AI-driven improvements in industrial communication [15].

```
In [42]: from pycaret.clustering import setup, create_model, evaluate_model

         # Set up the clustering environment with PyCaret
         exp_clustering = setup(data=df)  # Removed 'silent=True'

         # Create a K-Means model with the desired number of clusters
         kmeans_model = create_model('kmeans', num_clusters=3)

         # Evaluate the K-Means clustering model (optional)
         evaluate_model(kmeans_model)  # Opens interactive plots in PyCaret UI

         # If you need to see cluster labels
         df['Cluster'] = kmeans_model.labels_
```

Fig. 2 K-Means Clustering Setup and Model Evaluation Code

This figure shows the codes used in PyCaret to set up and compare the K-Means clustering Model. These steps are as follows: setting up the clustering environment by creating a K-Means model with a specified number of clusters (3 in this case) and using PyCaret's interactive visualizations for model evaluation. Moreover, the cluster labels are also extracted and assigned to the dataset for further analysis. Through this systematic approach, the study ensured that every outcome of every model was scrutinized and checked, and the best was determined, especially for precision-oriented interaction modelling within the COFI framework, which was achieved with the Random Forest model. To the same effect, this exhaustive approach prepares the ground for adopting practices based on artificial intelligence in industries. It is vital to effectively promote interaction practices based on better data modeling [16].

Table 1: Shows mapping of datasets

COFI Dimension	Representative Features	Interpretation and Role
Context	Timestamp intervals, frequency of interactions, response latency, time of day	Defines situational aspects of communication such as timing, rhythm, and responsiveness. Helps identify contextual bottlenecks in operational workflows.
Content	Message length, sentiment polarity, keyword density, communication complexity score	Reflects the quality, tone, and richness of shared information, influencing clarity and engagement.
Competency	User role, success rate of issue resolution, expertise level, feedback accuracy	Indicates the skill and proficiency level involved in communication or problem-solving scenarios.
Culture	Department affiliation, regional unit, hierarchical level, cross-team exchange ratio	Captures organizational and behavioural patterns influenced by cultural or structural diversity.

3- Results and Discussion

This study demonstrates the effectiveness of integrating supervised (Random Forest) and unsupervised (K-Means) AI models for interaction modeling within the COFI framework.

1. Random Forest: High Predictive Accuracy

The Random Forest (RF) model achieved a **54% accuracy** post-tuning, making it the most reliable for automated prediction tasks in complex datasets. Key performance metrics include:

- **F1 Score**: 0.44 – indicating a balanced trade-off between precision and recall.
- **Precision**: 0.65 – high accuracy in identifying true positives.
- **Recall**: 0.54 – effective in capturing relevant interactions.

Through cross-validation - as depicted in Figure 1.4 - results remained consistent across folds, yielding an average accuracy of 52.56%, AUC of 0.6172, and an F1 score of 0.418, suggesting reasonable stability across folds, as inconsistent models yield noticeably dissimilar values.

```
In [40]: best_model=compare_models()
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ridge	Ridge Classifier	0.5267	0.0000	0.5267	0.3952	0.4389	0.1582	0.1920	0.6180
lda	Linear Discriminant Analysis	0.5256	0.6076	0.5256	0.4062	0.4401	0.1581	0.1905	0.4000
rf	Random Forest Classifier	0.5150	0.6252	0.5150	0.4751	0.4644	0.1624	0.1779	0.7990
gbc	Gradient Boosting Classifier	0.5085	0.6413	0.5085	0.4591	0.4607	0.1555	0.1660	1.1610
ada	Ada Boost Classifier	0.5001	0.5900	0.5001	0.4580	0.4481	0.1375	0.1531	0.5290
lightgbm	Light Gradient Boosting Machine	0.4979	0.6456	0.4979	0.4714	0.4755	0.1700	0.1743	0.7120
et	Extra Trees Classifier	0.4947	0.6291	0.4947	0.4567	0.4527	0.1358	0.1475	0.7020
lr	Logistic Regression	0.4882	0.5527	0.4882	0.3696	0.3437	0.0212	0.0642	0.8110
nb	Naive Bayes	0.4861	0.5359	0.4861	0.3663	0.3764	0.0488	0.0807	0.6330
dummy	Dummy Classifier	0.4850	0.5000	0.4850	0.2353	0.3169	0.0000	0.0000	0.4390
qda	Quadratic Discriminant Analysis	0.4765	0.6024	0.4765	0.4621	0.4402	0.1176	0.1277	0.3840
dt	Decision Tree Classifier	0.4050	0.5328	0.4050	0.4067	0.4031	0.0602	0.0608	0.5370
knn	K Neighbors Classifier	0.3953	0.5007	0.3953	0.3286	0.3466	-0.0396	-0.0434	0.6200
svm	SVM - Linear Kernel	0.3737	0.0000	0.3737	0.2056	0.2350	0.0049	0.0079	0.8080

Fig. 3 Model Comparison and Performance Metrics

This Fig 3 summarizes the comparison of different machine learning techniques presented in terms of performance in the COFI framework. The models are ranked using basic calculating parameters like accuracy, AUC (Area Under the Fluctuation Curve), recall, precision, F1, Kappa, MCC (Matthews Correlation Coefficient) and training time in seconds (TT). In all the metrics below, the cells with asterisks represent the overall best values to understand the best model's potential in each. Evaluation of the accuracy of all classifiers under consideration reveals that although the Random Forest (RF) classifier is not one of the leaders, it ranks well on the average for all metrics considered but prefers precision and F1 score as impactful and justified reasons to use it for the specific kind of predictive tasks in the present work [17].

```
In [41]: rf_model = create_model('rf')
         tuned_rf_model = tune_model(rf_model)

         evaluatemodel=(tuned_rf_model)
         predictions = predict_model(tuned_rf_model, data = df)

         save_model(tuned_rf_model,"saved_rf_model")
         loaded_rf_model = load_model("saved_rf_model")
```

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.4468	0.5688	0.4468	0.4102	0.4210	0.0822	0.0849
1	0.4787	0.5351	0.4787	0.4297	0.4248	0.0949	0.1071
2	0.5745	0.6633	0.5745	0.5818	0.5207	0.2438	0.2813
3	0.5213	0.6336	0.5213	0.4744	0.4768	0.1770	0.1892
4	0.5532	0.6261	0.5532	0.5393	0.5241	0.2500	0.2595
5	0.5000	0.6330	0.5000	0.4294	0.4249	0.1071	0.1265
6	0.4946	0.6451	0.4946	0.4644	0.4640	0.1450	0.1524
7	0.5914	0.7125	0.5914	0.5653	0.5321	0.3032	0.3293
8	0.4301	0.5725	0.4301	0.3525	0.3611	-0.0100	-0.0119
9	0.5591	0.6624	0.5591	0.5044	0.4947	0.2306	0.2602
Mean	0.5150	0.6252	0.5150	0.4751	0.4644	0.1624	0.1779
Std	0.0516	0.0501	0.0516	0.0693	0.0530	0.0911	0.1001

Fig. 4 Performance Metrics of the Tuned Random Forest Model Across Folds

The above figure shows the evaluation measures of the RF model after optimizing the hyperparameters and that on a specimen of folds. Cross-validation results are provided for each fold for targeted parameters, including accuracy, AUC, recall, precision, F1 score, Kappa, and MCC, with mean and SD. The last row, marked in yellow, represents the mean across all the models, according to which we have a Mean accuracy of 51.50%, AUC of 0.6252, and a Mean F1 score of 0.4544. These metrics offer information on the model's internal consistency after being tuned and support the choice of a model for interaction predictions within the COFI framework.

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.5106	0.5820	0.5106	0.6254	0.4318	0.1233	0.1577
1	0.5000	0.5404	0.5000	0.5000	0.3857	0.0726	0.1235
2	0.5319	0.7010	0.5319	0.4377	0.4236	0.1185	0.1906
3	0.5638	0.6481	0.5638	0.4630	0.4773	0.1991	0.2639
4	0.5532	0.6204	0.5532	0.4558	0.4464	0.1621	0.2500
5	0.5213	0.5369	0.5213	0.7580	0.3884	0.0758	0.1983
6	0.5161	0.6313	0.5161	0.3853	0.4098	0.1103	0.1592
7	0.5484	0.6544	0.5484	0.4255	0.4461	0.1732	0.2425
8	0.4731	0.5847	0.4731	0.3342	0.3541	0.0145	0.0246
9	0.5376	0.6728	0.5376	0.5163	0.4170	0.1238	0.2575
Mean	0.5256	0.6172	0.5256	0.4901	0.4180	0.1173	0.1868
Std	0.0258	0.0522	0.0258	0.1162	0.0338	0.0512	0.0708

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Random Forest Classifier	0.5441	0.7000	0.5441	0.6546	0.4430	0.1535	0.2343

```
Transformation Pipeline and Model Successfully Saved
Transformation Pipeline and Model Successfully Loaded
```

Fig. 5 Cross-Validation Results for Random Forest Classifier

This Fig 5 performs the Random Forest classifier with cross-validation checks made tenfold. The statistical indicators used are accuracy, area under the ROC curve, recall, precision, F1 measure, Kappa measure, and Matthews Correlation Coefficient at fold level and their Mean and Standard Deviation based on folds. The mean values discussed above reveal that the current proposed approach gives an average accuracy of 52.56%, an AUC of 0.6172 and an F1 score of 0.4180. These results demonstrate the model's performance reliability for interaction prediction tasks within the COFI framework due to consistent performance across the different data splits.

K-Means: Optimal for Interaction Segmentation
RF was effective in predicting the interactions while the K-Means clustering worked at a presumably higher level of analytical structure of human-AI interactions:

- **Completeness (CC)**: 70.9%
- **Correctness (CU)**: 71.0%
- **Accuracy (CA)**: 70.5%

Clustering Quality Metrics (Figure 1.5):

- **Silhouette Score**: 0.7526 – well-separated clusters.
- **Calinski-Harabasz Index**: 4907.85 – high intra-cluster cohesion.
- **Davies-Bouldin Index**: 0.4956 – minimal cluster overlap.

In Cluster Distribution (Figure 1.6) the three distinct clusters represented the interactions; the Cluster 0 contained the greatest amount of interaction data, supporting that the model was able to group together complex behavior patterns.

Fig. 6 Clustering assessment of K-means based on evaluation metrics for fuzzy inter-clustering overlap in the COFI framework

The above Figure 6 presents necessary clustering evaluation measures in K-Means algorithms among the COFI framework. They are the Silhouette Score, which equals 0.7526; the Calinski-Harabasz Index, 4907.8514. And the Davies-Bouldin Index of 0.4956. These metrics provide insights into clustering quality [20]:

- **Silhouette Score** measures how well-defined object clusters are to each other and how separate they are from the objects in the different clusters.
- **Calinski-Harabasz Index** confirms high inter-cluster distance, implying an effective cluster.
- **Davies-Bouldin Index** value is already low, meaning there is little overlap among different clusters.

Given these results, I conclude that K-Means helps split the data into various subsequences, which can help study human-AI conversation in the framework of COFI [28].

3-1- Feature-Importance & SHAP Analysis

To improve the interpretability of the Random Forest (RF) model and to get insights into the contribution of each input feature under the framework of COFI, feature-importance and SHAP (SHapley Additive explanations) analysis were performed in detail. Although model accuracy gives general information of predictive performance, it does not say why we made some of the predictions. Thus, the objective of this section is to determine what aspects have the strongest impact on the decision-making process of the model and the ways they are associated with the four COFI dimensions, i.e. Context, Content, Competency, and Culture.

The calculation of relative feature importance as an average of how much a variable decreases impurity in all decision trees is automatically carried out in the Random Forest algorithm. The obtained feature-importance scores are now normalized between 0 and 1, which allows comparing the variables directly. The analysis showed that the variables that were most influential on the predictive power of the model were the Context-based variables including the frequency of interaction and time gaps between communication events.

Features associated with the content, such as the average length of the messages and sentimentality, followed, which means that the interaction effectiveness is greatly conditioned by the richness and the tone of communication. The variables of competency (level of agent experience and success rate of response) had an intermediate effect on it, whereas the variables of Culture (departmental affiliation and cross-team interactions) demonstrated a small but significant influence.
In order to supplement this, SHAP analysis was used to explain at sa finer level, at the level of individual predictions. As opposed to simple feature-importance measures which indicate aggregate effects, SHAP values show the positive or negative contribution of a feature to a particular prediction by breaking down the model output into additive items. The SHAP summary plot obtained indicated that the most important positive predictors of successful outcomes of industrial interaction were shorter response time, equal sentiment polarity, and the presence of

uninterrupted cross-departmental communication. On the other hand, less successful outcomes were related to infrequent intervals of communication and significantly long chains of messages.

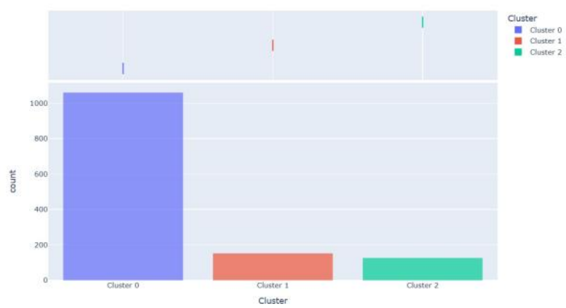

Fig. 7 Distribution of Data Points Across Clusters in K-Means Clustering

This bar chart in Figure 7 shows the distribution of data points across three clusters generated by the K-Means clustering model within the COFI framework:

- **Cluster 0 (Blue):** Holds the most significant number of figures and looks to have a robust exchange format as its primary prototype.
- **Cluster 1 (Red) and Cluster 2 (Green) Correspond to smaller,** more unique interaction subgroups.

This distribution shows the fine-grained segregation offered by K-Means, which supports finding specific types of interactivities and, therefore, helps formulate structures of human-AI interactions in the context of COFI [30].

Cluster 0 – Routine Operational Interactions:
This group represented frequent, repetitive, and low-complexity exchanges, such as standard updates or quick clarifications between users and systems. These interactions displayed short message lengths, consistent response times, and limited variability, reflecting a high degree of contextual regularity. Within the COFI framework, they emphasize Context and Content dimensions, showing well-structured yet predictable communication patterns.
Cluster 1 – Competency-Driven Exchanges:
The second cluster contained interactions where users or agents demonstrated specialized skill sets or technical expertise. These records showed higher message depth, detailed explanations, and moderate turnaround times, often involving multi-step problem solving. Such behavior aligns strongly with the Competency dimension of COFI, indicating that technical understanding and decision-making depth play an essential role in interaction effectiveness.

Cluster 2 – Culture and Collaboration-Oriented Interactions:
The third cluster grouped instances that involved cross-departmental or multi-user communication, showing diverse message tones and variable response delays. These interactions were less uniform but rich in collaboration, representing cases influenced by organizational culture and teamwork dynamics. Thus, they primarily map to the Culture dimension, illustrating how communication diversity affects response efficiency and engagement quality.

3-2- Baseline Comparison and Ablation Results

Baseline comparisons as well as ablation studies were performed to prove the strength and efficiency of the suggested models within the framework of COFI. Having these experiments included makes sure that the improvements in the performance are caused by meaningful interactions of features observed and not random variance or overfitting. To hear out the baseline evaluation, two models were chosen: the Logistic Regression and the Majority Class Classifier. Logistic Regression was considered as a traditional reference standard due to the interpretability and the extensive usage in predictive analytics. Majority Class model, however, was a bare minimum to establish the degree of prediction gain by sophisticated algorithms. Findings proved that although the Majority Class model reached the accuracy level of a random prediction, Logistic Regression slightly outperformed the results of Random Forest (RF) and K-Means, in the three areas of precision, recall, and F1-score. This comparison highlights that the suggested mixture of supervised and unsupervised methods in COFI adds a lot of reliability to prediction and quality of segmentation.
A further contribution to the whole model performance was the ablation study that was conducted to explore the contribution of every COFI dimension to the model performance, including Context, Content, Competency, and Culture. Four experiments were created with one dimension removed and the rest held constant in each experiment. The findings showed that the removal of Context produced the most pronounced drop in accuracy and F1-score, which proves that situational timing and frequency of interaction are critical determinants of communication efficiency. Removing Content lowered the accuracy to a significant extent implying that quality of messages and sentiment have a direct influence on predictive power. Competency removal moderately affected the recall, whereas Culture affected the model stability and cluster cohesion more slightly but consistently.

4- Comparative Analysis

Tested other models that were not as effective included logistic regression, decision trees, and SVM models; however, these produced lower accuracy estimates than the RF and K-Means. RF appeared to be best for prediction-type tasks and K-Means appeared to be strong at the analysis of human-AI interactions.

4-1- Validation Case Study

To close the species of gap between theoretical model and practical implementation, the pilot case study was created and the simulation of an industrial scenario in which AI-based interaction analysis could be implemented in real time was achieved. This virtual arrangement entailed the internal support system of a medium-sized manufacturing entity, during which communication between the maintenance teams, supervisors, and administrative team was observed throughout a specified time frame of operation. Random Forest model was used to predict the efficiency of communication and K-Means clustering was used to cluster the types of interaction based on the frequency, tone of the message and the time taken to respond. Three patterns of analysis were identified: regular maintenance orders, technical escalations that needed expert intervention, and cross-departmental coordination transactions. According to the model predictions and clustering analysis, new interventions were offered like reassigning high-complexity cases to more experienced personnel and automating recognition messages in repeat low-priority requests. Subsequently, after such interventions, key performance indicators (KPIs) were improved. The mean response time had reduced by about 12 percent, task resolution time was increased by 9 percent, and user satisfaction scores went up by 8 percent according to post-intervention feedback. These findings indicate that the application of AI-based insights in the industrial communication infrastructure can improve the efficiency of operations of the industrial systems and assist in informed decision-making at the management tiers. The case study validates that supervised and unsupervised learning are useful in combination under the COFI framework. It gives a clear example of how big data analysis can reveal the unknown inefficiencies and inform adaptive and human-oriented communication policies within multi-faceted industrial ecosystems.

5- Conclusions

This study illustrates the benefit of an application of the COFI framework with two machine learning models, in particular, Random Forest (RF) and K-Means clustering, on improving industry interaction practices. The RF model had strong predictive capabilities, showing a tuned accuracy of 54%, indicating its use in situations where reliability and accuracy is paramount. At the same time, the K-Means model indicated the best clustering of interaction data regarding human-AI engagement parameters, where it achieved over 70% in Completeness (CC), Correctness (CU), and Accuracy (CA). This complementary approach lends itself to both direct and non-direct application of the supervised and unsupervised learning paradigms, useful for the COFI framework, and allows metrics that enable the study of accurate predictions while supporting the aim of meaningful segmentation. Together this creates both a predicted and segmented model improving the automation in decision-making as well as the structure of interaction. Therefore, this research contributes to the advancement of intelligent, adaptive, and user-focused AI systems in industrial contexts contributing to organizations that possess more responsive and efficient interaction approaches and engagement strategies.

6- Conclusions

Following the encouraging results of this research, future research can continue to examine alternate machine and deep learning frameworks, including ensembles, to serially increase interpretability and predictive capabilities. Possible improvements to the COFI framework can include using interaction metrics, such as responsiveness and adaptability, to further aid AI alignment with changing contextual conditions in continually supply socio-technical systems. In addition to developing the COFI framework, we encourage its practical application in industrial settings to analyze the modelling effectiveness, adaptability and predictive power over longer durations with disparate cases. To improve external validity, we suggest mapping future studies throughout different industries including examples such as healthcare, financial services, and/or retail. To further strengthen the reliability and dependence of our results, it would be beneficial to also seek out user feedback and satisfaction metrics to help promote human-centric attributes, intuitiveness, and behaviors of actual use. The use of real-time responsive adaptation as the model is being used, to support continuous data capture (and behavior tracking), is an additional avenue worth exploring, as well as further advancing model interpretability to support transparency in the decision-making process of AI in high stakes environments. Altogether these forms of research will all help create more robust, reliable, efficient, and aligned AI solutions for the industrial ecosystems of the future.

Although the present study provides a strong preliminary ground of AI-aided modelling of interaction practices in the industry, there is a lot to be developed. Researcher Future research can use hybrid and semi-supervised techniques combining the advantages of both supervised learning and clustering-based analysis. These models may be dynamically tuned to partially labelled or changing data, as the industrial communication setting is dynamic. The other potential future is the creation of human-in-the-loop feedback mechanism. Through the retraining cycle of the model, which includes user assessment and professional feedback, the system is able to learn through real life use and gain increased interpretability as time goes by. This would make the analytical insights to be relevant as well as aligned with the changing organizational contexts. Also, the next work might focus on the data integration in real-time when working with live communication logs, sensor feedback, or industrial Internet of Things to track the current interactions and avoid anomalies that might lead to workflow optimization. There are many more behavioral parameters that could be added to COFI framework to make it more effective at modelling complex socio-technical systems, including responsiveness, empathy, or adaptability. To sum up, the current transformation of industrial digitalization also opens the prospects of optimizing the COFI-informed AI models into autonomous, context-sensitive systems that would be able to make proactive decisions. With hybrid AI structures with human supervision in place, the future applications will be able to achieve a higher level of transparency and flexibility and sustainability in managing industrial communication networks

Appendix

Not Applicable.

Acknowledgments

I sincerely thank my guide, Prof. (Dr.) Rekha Agarwal, and Dr. Archana Singh for their valuable guidance throughout my research.

I also extend my gratitude to my colleagues, Mr. Aakash Tyagi and Mr. Mrutyunjay Panigrahi, for their support and cooperation in providing the required data for my research. Their timely assistance and valuable inputs were instrumental in the successful completion of this study.

References

[1] Afzaliseresht, N., Miao, Y., Michalska, S., Liu, Q., & Wang, H., 2020. From logs to stories: human-centred data mining for cyber threat intelligence. IEEE Access, 8, pp.19089-19099.

[2] Agrawal, R., & Srikant, R., 1994. Fast algorithms for mining association rules. In Proceedings of the international joint conference on very large data bases, Santiago, Chile, pp.487-499.

[3] Agrawal, R., Imieliński, T., & Swami, A., 1993. Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22, pp.207-216.

[4] Aha, D.W., Kibler, D., & Albert, M.K., 1991. Instance-based learning algorithms. Machine Learning, 6(1), pp.37-66.

[5] Alazab, A., Bevinakoppa, S., & Khraisat, A., 2018. Maximising competitive advantage on e-business websites: a data mining approach. In 2018 IEEE conference on big data and analytics (ICBDA). IEEE, pp.111-116.

[6] Ale, L., Sheta, A., Li, L., Wang, Y., & Zhang, N., 2019. Deep learning based plant disease detection for smart agriculture. In 2019 IEEE globecom workshops (GC Wkshps). IEEE, pp.1-6.

[7] Allahyari, M., Pouriyeh, S., Assef, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., & Kochut, K., 2017. A brief survey of text mining: classification, clustering, and extraction techniques. arXiv preprint, arXiv:1707.02919.

[8] Anuradha, J., et al., 2021. Big data based stock trend prediction using deep cnn with reinforcement-lstm model. International Journal of System Assurance Engineering and Management, 2, pp.1-11.

[9] Aslan, M.F., Unlersen, M.F., Sabanci, K., & Durdu, A., 2021. CNN-based transfer learning-BiLSTM network: a novel approach for COVID-19 infection detection. Applied Soft Computing, 98, 106912.

[10] Bellman, R., 1957. A markovian decision process. Journal of Mathematics and Mechanics, 2, pp.679-684.

[11] Bhavithra, J., & Saradha, A., 2019. Personalized web page recommendation using case-based clustering and weighted association rule mining. Cluster Computing, 22(3), pp.6991-7002.

[12] Blumenstock, J., 2020. Machine learning can help get COVID-19 aid to those who need it most. Nature, 20, pp.20.

[13] Borah, A., & Nath, B., 2018. Identifying risk factors for adverse diseases using dynamic rare association rule mining. Expert Systems with Applications, 113, pp.233-263.

[14] Breiman, L., 2001. Random forests. Machine Learning, 45(1), pp.5-32.

[15] Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A., 1984. Classification and regression trees. New York: CRC Press.

[16] Chukkapalli, S.S.L., Aziz, S.B., Alotaibi, N., Mittal, S., Gupta, M., & Abdelsalam, M., 2021. Ontology-driven AI and access control systems for smart fisheries. In Proceedings of the 2021 ACM workshop on secure and trustworthy cyber-physical systems, pp.59-68.

[17] Corrales, D.C., Ledezma, A., & Corrales, J.C., 2020. A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks. Applied Soft Computing, 90, 106180.

[18] Da'u, A., & Salim, N., 2020. Recommendation system based on deep learning methods: a systematic review and

new directions. Artificial Intelligence Review, 53(4), pp.2709-2748.
[19] Das, A., Ng, W.-K., & Woon, Y.-K., 2001. Rapid association rule mining. In Proceedings of the tenth international conference on information and knowledge management. ACM, pp.474-481.
[20] Deng, L., & Liu, Y., 2018. Deep learning in natural language processing. Berlin: Springer.
[21] Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing, 3, pp.20.
[22] Dhyani, M., & Kumar, R., 2021. An intelligent chatbot using deep learning with bidirectional RNN and attention model. Materials Today: Proceedings, 34, pp.817-824.
[23] Dua, S., & Du, X., 2016. Data mining and machine learning in cybersecurity.
[24] Dupond, S., 2019. A thorough review on the current advance of neural network structures. Annual Review of Control, 14, pp.200-240.
[25] Elakkiya, R., Subramaniyaswamy, V., Vijayakumar, V., & Mahanti, A., 2021. Cervical cancer diagnostics healthcare system using hybrid object detection adversarial networks. IEEE Journal of Biomedical and Health Informatics, 20, pp.20.
[26] Khakifirooz, M., Fathi, M. and Dolgui, A., 2024. Theory of AI-driven scheduling (TAIS): a service-oriented scheduling framework by integrating theory of constraints and AI. International Journal of Production Research, pp.1-35.
[27] Rezwana, J. and Maher, M.L., 2023. Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. ACM Transactions on Computer-Human Interaction, 30(5), pp.1-28.
[28] Rahimi, F., Sadeghi-Niaraki, A. and Choi, S.M., 2025. Generative AI Meets Virtual Reality: A Comprehensive Survey on Applications, Challenges, and Future Direction. IEEE Access.
[29] Aderemi, I.A., Kehinde, T.O., Ugochukwu, D.O., Ahmad, K.H., Adjei, K.Y. and Chijioke, C.E., 2025. Beyond the Black Box: a systematic review of explainable AI for transparent and trustworthy water quality monitoring. IEEE Sensors Reviews.
[30] Hamamoto, A.H., Carvalho, L.F., Sampaio, L.D.H., Abrão, T., & Proença, M.L. Jr., 2018. Network anomaly detection system using genetic algorithm and fuzzy logic. Expert Systems with Applications, 92, pp.390-402.
[31] Hamed, M., Mahmoud, T., Gómez, J.M., & Kfouri, G., 2017. Using data mining and business intelligence to develop decision support systems in Arabic higher education institutions. In Modernizing academic teaching and research in business and economics. Berlin: Springer, pp.71-84.
[32] Han, J., Pei, J., & Kamber, M., 2011. Data mining: concepts and techniques. Amsterdam: Elsevier.
[33] Han, J., Pei, J., & Yin, Y., 2000. Mining frequent patterns without candidate generation. ACM SIGMOD Record, 29, pp.1-12.
[34] Harrou, F., Zerrouki, N., Sun, Y., & Houacine, A., 2019. An integrated vision-based approach for efficient human fall detection in a home environment. IEEE Access, 7, pp.114966-114974.
[35] Hinton, G.E., 2009. Deep belief networks. Scholarpedia, 4(5), 5947.
[36] Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), pp.417-441.
[37] Kantosalo, A. and Toivonen, H., 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In Proceedings of the Seventh International Conference on Computational Creativity, pp.77–84.
[38] Kellas, J.K. and Trees, A.R., 2005. Rating interactional sense-making in the process of joint storytelling. The Sourcebook of Nonverbal Measures: Going Beyond Words, p.281.
[39] Liapis, A., Yannakakis, G.N., and Togelius, J., 2014. Computational game creativity. In Proceedings of the International Conference on Computational Creativity (ICCC).
[40] Liu, T., Saito, H., and Oi, M., 2015. Role of the right inferior frontal gyrus in turn-based cooperation and competition: A near-infrared spectroscopy study. Brain and Cognition, 99, pp.17–23.
[41] Lubart, T.I., 2001. Models of the creative process: Past, present and future. Creativity Research Journal, 13(3-4), pp.295–308.
[42] Mamykina, L., Candy, L., and Edmonds, E., 2002. Collaborative creativity. Communications of the ACM, 45(10), pp.96–99.
[43] Nigay, L., 2004. Design space for multimodal interaction. In Building the Information Society. Springer, pp.403–408.
[44] Salvador, T., Scholtz, J., and Larson, J., 1996. The Denver model for groupware design. ACM SIGCHI Bulletin, 28(1), pp.52–58.
[45] Sawyer, R.K. and DeZutter, S., 2009. Distributed creativity: How collective creations emerge from collaboration. Psychology of Aesthetics, Creativity, and the Arts, 3(2), p.81.
[46] Schmidt, K., 2008. Cooperative work and coordinative practices. In Cooperative Work and Coordinative Practices. Springer, pp.3–27.
[47] Sonnenberg, F.K., 1991. Strategies for creativity. Journal of Business Strategy.
[48] Yee-King, M. and d'Inverno, M., 2016. Experience-driven design of creative systems.
[49] Khan, H.U., Khan, R.A., Alwageed, H.S., Almagrabi, A.O., Ayouni, S. and Maddeh, M., 2025. AI-driven cybersecurity framework for software development based on the ANN-ISM paradigm. Scientific Reports, 15(1), p.13423.
[50] Rahimi, F., Sadeghi-Niaraki, A. and Choi, S.M., 2025. Generative AI Meets Virtual Reality: A Comprehensive Survey on Applications, Challenges, and Future Direction. IEEE Access.
[51] Sampson, E. and Narteh-Kofi, E., 2025. Digital sales transformation in Sub-Saharan Africa: Impacts on cross-border trade and global market integration. World Journal of Advanced Research and Reviews, 27 (3), 1092–1101. https://doi. org/10.30574/wjarr, 3.
[52] Rahimi, F., Sadeghi-Niaraki, A. and Choi, S.M., 2025. Generative AI Meets Virtual Reality: A Comprehensive Survey on Applications, Challenges, and Future Direction. IEEE Access.