International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Human-Guided Agentic AI Workflows for Enterprise Operational Decision-Making

Sai Viswa Teja Arumilli

Submitted: 05/06/2024 **Revised:** 18/07/2024 **Accepted:** 29/07/2024

Abstract: This study demonstrates human-in-the-loop agentic AI workflows for making operational decisions using an organized Python-based analytical approach within this enterprise. The study examines the enhancement of decision-making processes by combining the capabilities of AI agents with human supervision in terms of accuracy, productivity and trust. The data from the enterprise has been pre-processed, visualized and analyzed by applying machine learning methods. Two models are employed: Logistic Regression and Random Forest, with observed better performance being achieved for Random Forest in dealing with a complex relationship. The result shows that hybrid systems (AI/Human) achieve higher reliability and interpretability compared with fully-automated systems. This study emphasizes the significance of Explainable (XAI) and Regulated (Injection Deception) Artificial Intelligence (AI) in the enterprise.

Keywords: *Agentic AI, Human-in-the-Loop, Enterprise Decision-Making, Machine Learning, Random Forest, Explainable AI*

I. Introduction

Background

The use of artificial intelligence systems in the enterprise world is becoming common to facilitate operational decisions. Independent agents AI workflows or agentic workflows, in which autonomous agents do work and make decisions, are changing how organizations process information, optimize processes, and react to changing business environments. Autopilot systems are frequently not transparent, accountable, and lacking contextual awareness. Consequently, human in the loop (HITL) methods are becoming increasingly significant to ensure that the critical decisions being made are still in line with the organizational objectives.

Problem Statement

Although agentic systems of AI are growing more advanced, numerous enterprise decision-making still fall into the issue of error due to over-automation and poorly supervised humans. Independent agents can be less than optimal or biased in their performance where they lack context understanding or ethical reasoning [1]. This poses a hole in the development of engineered hybrid workflows to successfully fuse human judgment and automation based on AI [2]. Thus, human-in-the-loop agentic workflows, which would improve decision accuracy, transparency, as well as operational reliability, need to be developed and evaluated.

Independent Researcher, USA.

Aim and Objectives

This research aims to design and explore the human-managed agentic AI processes to enhance enterprise decision-making in operations by analyzing data and simulating with Python.

Objectives:

➔ ***To review available agentic AI processes of enterprise decision systems.***

➔ ***To create a human-in-the-loop model of guided AI decisions processes.***

➔ ***To introduce Python based simulation in evaluation of agent and human decisions.***

➔ ***To measure the impact of hybrid workflow on accuracy and operational efficiency.***

Novel Contribution

This work provides a formalized human-centric agentic workflow framework execution model, enabling human control as AI-based enterprise decision engines. The proposed approach brings on board the controlled intervention points, which not only verifies human judgment in regards to AI-generated judgments but also modifies it. In contrast to fully autonomous or entirely manual systems [3].

The innovation is that agentic AI simulation is combined with human decision checkpoints carried out by analytical modelling in Python.

II. Related Work

Agentic AI Systems: Enterprise Decision-Making

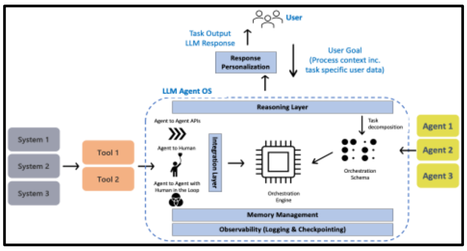


Fig.1: Agentic AI Systems: Enterprise Decision-Making

The agentic AI systems are independent computational systems that have the ability to perceive the environment, reasoning on input, and taking actions with minimal human oversight [4]. They are becoming used in other areas like the ward-off chain optimization, foresightful upkeep, money hazard, and buyer operations control in the milieu of enterprise operations [5]. Contemporary agentic systems frequently integrate machine learning models with large language models and rule-based reasoning to recreate decision-making processes that previously might have been done by human beings [6]. Studies have indicated that such systems enhance largely faster operational rates, scalability, and uniformity in decision making [7]. Nonetheless, despite these merits, there are a number of drawbacks [8]. Autonomous agents can also make decisions that are not contextualized particularly where business is dynamic or unclear [9]. A recent study highlights those unguided autonomous systems do not align with the organizational goals as they possess low explain ability and their training data has underlying biases.

Human-in-the-Loop (HITL) AI Frameworks

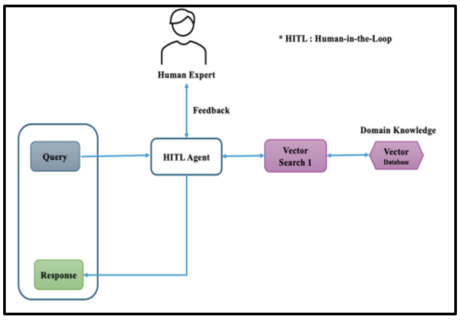


Fig.2: Human-in-the-Loop (HITL) AI Frameworks

HITL artificial intelligence models combine human knowledge with machine learning and automated decision models to enhance accuracy, fairness and reliability. HITL systems have been broadly applied in the sphere of enterprise decision-making where solutions involve high-risk and context-based decision-making in the sphere of fraud detection, medical diagnostics, cybersecurity monitoring, and financial forecasting [10]. These frameworks enable the human to authenticate, refute, or feed-back the AI-generated outputs to form a feedback loop that optimizes performance of the model with time [11]. Another research study argues that machine learning with HITL systems is better with human corrections to more iterative training cycles to obtain more adaptive and robust models [12]. Nevertheless, other restrictions are recognized in literature, too [13]. Human participation brings latency, greater operational cost and might not be efficient in high-volume decision making [14]. Also, human judgments may not be consistent, and this may introduce variability to the system. Regardless of these obstacles, HITL is imperative in ensuring ethical compliance, transparency, and accountability within AI-based systems.

Agentic Workflow Automation in Business

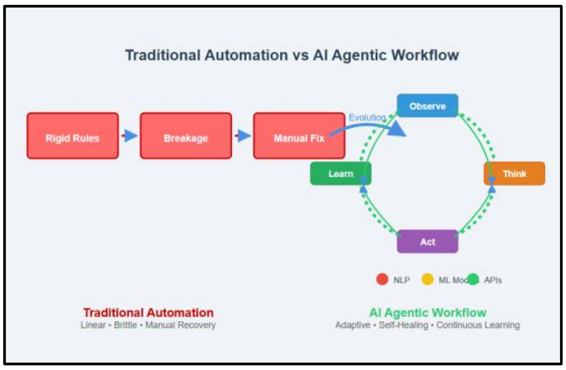


Fig.3: Agentic Workflow Automation in Business

Agentic workflow automation is the coordination of two or more AI agents working to complete complex business processes with minimal human assistance [15]. These processes duplicate formal enterprise processes like procurement, routing customer service, inventory control, and financial reporting [16]. Agents are usually specialists at a single task and the results are sent across the workflow pipeline in a sequential or an iterative manner [17]. Recent studies suggest that agentic workflow systems have a substantial positive effect on the efficiency of operations since they minimize manual interaction and provide the ability to execute processes uninterrupted [18]. Large language models with reinforcement learning have also promoted the flexibility of the latter in dynamic environments [19]. Nonetheless, there are also risks related to such automation, which are also pointed out in literature [20]. A significant issue is cascading failure where failure by a single agent in the workflow spreads to others and causes further errors.

Explain ability and Governance in AI Systems

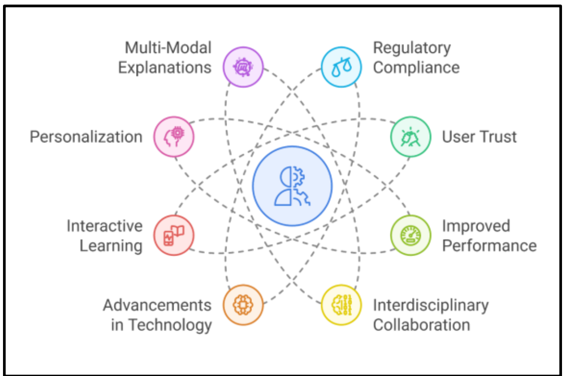


Fig.4: Explain ability and Governance in AI Systems

Security, explain ability, and control are the basic needs to the effective implementation of AI systems in making decisions in the enterprise. Explainable AI (XAI) is concerned with methods to interpret machine learning models in a manner accessible to human users, enabling them to comprehend how and why they are made [21]. This is especially so when making decisions in an enterprise setting to influence financial performance, customer satisfaction and regulatory compliance [22]. A study indicates that decision-makers tend to rely on AI systems when there is the ability to interpret model outputs, as well as in the interpretation of decision logic [23]. Non-transparent black boxes of advanced artificial intelligence, e.g., deep learning models and multi-agent frameworks, have been identified, though [24]. Government regulations like the NIST AI Risk Management Framework and OECD AI Principles highlight equality, responsibility, disclosure and resilience in the implementation of AI [25]. These frameworks suggest introducing a mechanism of human control, constant monitoring, and risk assessment in AI systems.

Literature gap

The existing research studies are primarily dedicated to either fully autonomous agentic AI systems, or traditional human-in-the-loop systems, in isolation. There are limited studies that combine both strategies in an overall workflow of enterprise systems [26]. Further, the quantitative human-guided and human-autonomous decision-making analysis in actual operation areas of enterprises has been lacking in the fields or real data, particularly in actual enterprise operation areas where there are measurable evaluation metrics that can be used to compare.

III. Methodology

Research-Design and Approach

The proposed study uses a mixed approach comprising quantitative research design and experimental simulation, to examine the concept of human-directed agentic AI workflow in enterprise decision-making. Its method combines data-driven modelling based on Python to examine the efficiency of autonomous AI agents in comparison to human-in-the-loop (HITL) systems. There is utilization of a comparative experimental framework wherein machine-based decisions are evaluated under varying conditions of operation which includes the application on different levels of human control [27]. The study is performed in a systematic pipeline, which includes data pre-processing, exploratory data analysis (EDA), and feature engineering, the categorization of the performance levels, model training, and evaluation [28]. The research aims to determine trends in operational efficiency, accuracy of decision making and measures of trust based on statistical and machine learning methods.

Variables and Data Description

The dataset in this study is a simulated enterprise dataset in the form of a structure with operational decision processes and agency AI leadership. It is modelled to reflect the real-life enterprise settings where AI agents run in tandem with human supervision in the business processes that require critical decisions [29]. The sample size is sufficient to work with ***5,500 records and 20 attributes*** to perform a statistical analysis, build a classification model and a human-in-the-loop experiment. Each record is an enterprise operational case with AI-based decision-making with different degrees of autonomy, complexity, and human oversight. The dataset contains main variables including the ***industry type, Organization Size, Leadership Function, AI Maturity Level, Agent Type, Use Case Area, Agent Autonomy Level, Decision-***

Making Type, Context Awareness Score, Task Complexity Level, Human Oversight Level, Explain ability Level, Data Privacy Compliance, Integration Level, Task Success Rate, Response Time, Productivity Improvement Percentage, Leadership Trust Score, and Adoption Success Level. To assist in interpretability in this study, some of the continuous variables properly are retailored into categorical levels (Low, Medium, High) with the help of quantile-based segmentation:

$$Category\ Level = Q(x, 0.33, 0.66)$$

Where, x represents the continuous variables.

Data Preparation

The data preparation process makes the data relevant, reliable, and analyzable. Python functions of isnull.sum () were used to test how many missing values in the dataset, duplicated.sum () was used to test the presence of duplicate records and describe was used to understand the statistical distribution. The missing values are included with the mean imputation (numerical variables) and mode imputation (categorical variables):

$$xfilled = \{mean(x), if\ numeric\ missing\ value$$

$$\{mode(x), if\ categorical\ missing\ value$$

Analysis revealed that there are no duplicates or missing values in the dataset, thus high quality of data. Additionally, it identified the outliers-based on boxplot visualization through seaborn and Interquartile Range (IQR), which revealed that the data does not have any major outliers. Thus, the data are perfectly consistent, and can be used directly to model and analyze. The encoding process is mentioned below:

$$x' = f(x)$$

where f maps categorical values.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an analysis performed to comprehend the form, the distribution and relationships within the dataset by means of using Python libraries like pandas, matplotlib, and seaborn. The first step is to analyze it with summary statistics (describe()), and the structure of dataset (info()). The data distribution is analyzed with visualization techniques like histograms, variability and absence of outliers are determined with boxplots, categorical variables are compared by bar charts, proportional distribution is presented by pie charts, and the trends across the levels of AI maturity are observed with line graphs [30]. Corr analysis with corr() is used to determine the relationships between variables such as the success of tasks, improvement in productivity, and score of leadership trust, which may be utilized in modelling through deeper understanding.

Data Splitting Strategy

An 80:20 split is made on the dataset into a training and testing set:

$$D = Dtrain + Dtest$$

where:

- Dtrain=80%
- Dtest=20%

This guarantees that models are trained on adequate information with the unconditional value of evaluation on unknown information.

Machine Learning Models in use

The machine learning models are being applied in the comparative evaluation of two supervised models:

(1) Logistic Regression Model

High, Medium and Low performance levels are categorical outcomes that are predicted using logistic regression. The probability function will be:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta 0 + \beta 1x)}}$$

where β0 is intercept and β1 represents feature coefficients.

(2) Random Forest Classifier

Random Forest is a type of ensemble learning technique that enhances the predictive accuracy. It builds up several decision trees and compiles findings with majority voting:

$$\hat{y} = mode(T1(x), T2(x), \ldots, Tn(x))$$

where Tn(x) represents individual decision tree.

Model Evaluation

The model is evaluated by the accuracy score, precision, recall, F1-score and confusion matrix with the implementation of sklearn.metrics in Python. These measures evaluate the distance performance of Logistic Regression and Random Forest model. Findings are contrasted to reveal the performance of models in forecasting the enterprises decision outcomes when agentic workflows of AI are applied.

Accuracy Calculation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision Calculation:

$$Precision = \frac{TP}{TP + FP}$$

Recall Calculation:

$$Recall = \frac{TP}{TP + FN}$$

F1 Score Calculation:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Human-in-the-Loop Simulation Framework

An artificial HITL mechanism is incorporated into the working procedure as AI-produced results are verified or corrected according to specified benchmarks. Loss of confidence of the model below a given level, results in human intervention. This is represented as:

$$Decision = \{AI(x), if\ confidence\ \geq \theta$$

$$\{Human(x), if\ confidence < \theta$$

where θ is the confidence threshold

Ethical Issues and Mitigation

The current research is characterized by ethical issues connected with data privacy, bias in AI-based decision-making systems, and transparency of decision-making systems. The data is anonymized in order to have no personal or sensitive information revealed. Mitigation of bias is considered by performing balanced pre-processing and assessment methods to prevent biased model results. Moreover, the human-in-the-loop integration guarantees AI course of action accountability and transparency. All the ethical AI principles, including fairness, explain ability, and responsible data use, are adhered to during the analysis.

IV. System Design

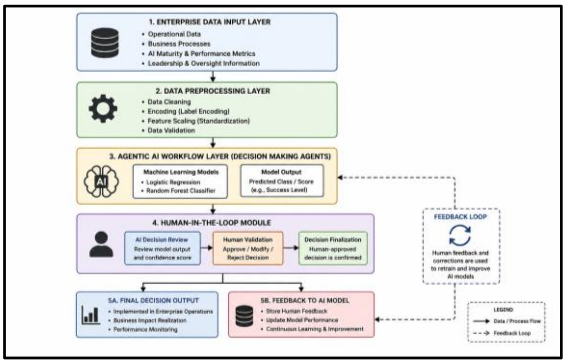


Fig. 5: Architecture Diagram for Human-in-the-Loop Agentic Workflow system

The architecture depicts enterprise data processing and analysis pipelines using AI models, followed by human review and improvement based on feedback for reliable operational decisions, which is known as Human in the Loop (HiL)-Agentic AI architecture.

```
Initialize
    Outputs
        AI Decision
        Human Approval Signal

    Inputs
        Task Data
        AI Prediction Score
        Human Override Input

    Registers
        Confidence_Count = 100
        Threshold = 70

Start loop

    IF Task Data received THEN process input

    IF AI Prediction Score >= Threshold THEN
        Set Decision = AI Approved
        Decrease Confidence_Count

    IF AI Prediction Score < Threshold THEN
        Request Human Review
        WAIT for Human Override Input

    IF Human Override = Approve THEN
        Set Decision = Final Approved

    IF Human Override = Reject THEN
        Reprocess AI Model Output

    Update Confidence_Count based on feedback

    Store Decision in Output Register

    Delay for next decision cycle

End loop
```

Fig. 6: Pseudocode

The pseudocode represents an agentic AI workflow to analyze task data, make predictions, use threshold guidance, seek human review, if necessary, revise the measure of confidence, and save the final business decision in an iterative fashion.

V. Implementation

```
  Record_ID           Industry Organization_Size Leadership_Function  \
0  AAI-00001          Education             Small          Operations
1  AAI-00002      Manufacturing             Small          Operations
2  AAI-00003 Telecommunications            Medium          Operations
3  AAI-00004             Energy             Large     Human Resources
4  AAI-00005 Financial Services             Small           Marketing

  AI_Maturity_Level                  Agent_Type           Use_Case_Area  \
0             Early       Risk Monitoring Agent      Process Automation
1          Advanced  Knowledge Management Agent      Demand Forecasting
2           Leading   Workflow Automation Agent      Process Automation
3           Leading   Workflow Automation Agent      Workforce Planning
4          Advanced           Procurement Agent  Content Personalization

  Agent_Autonomy_Level Decision_Making_Type  Context_Awareness_Score  \
0                  Low           Rule-based                     24.0
1               Medium             ML-based                     64.7
2                 High             ML-based                     64.0
3               Medium               Hybrid                     75.8
4               Medium             ML-based                     56.0

  Task_Complexity_Level Human_Oversight_Level Explainability_Level  \
0                Simple                  High                 High
1                Simple                   Low               Medium
2              Moderate                   Low                  Low
3              Moderate                Medium               Medium
4              Moderate                  High               Medium

  Data_Privacy_Compliance Integration_Level  Task_Success_Rate  \
0                     Yes         API-based               73.6
1                      No         API-based               71.6
2                     Yes         API-based               82.2
3                      No   Enterprise-wide               84.4
4                     Yes         API-based               78.7

   Response_Time_Seconds  Productivity_Improvement_Percent  \
0                   4.39                              11.8
1                   3.36                              16.0
2                   6.92                              22.8
3                   6.22                              17.8
4                   6.82                              16.1

   Leadership_Trust_Score Adoption_Success_Level
0                    76.1                 Medium
1                    39.6                 Medium
2                    53.6                   High
3                    50.6                   High
4                    76.3                   High
```

Fig.7: Dataset Sample Overview

This figure depicts the organization-level attributes such as industry, AI maturity and leadership function present in this dataset. It is worth the initial understanding of 'Variables' for analyzing agentic AI.

```
Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5500 entries, 0 to 5499
Data columns (total 20 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   Record_ID                         5500 non-null   object
 1   Industry                          5500 non-null   object
 2   Organization_Size                 5500 non-null   object
 3   Leadership_Function               5500 non-null   object
 4   AI_Maturity_Level                 5500 non-null   object
 5   Agent_Type                        5500 non-null   object
 6   Use_Case_Area                     5500 non-null   object
 7   Agent_Autonomy_Level              5500 non-null   object
 8   Decision_Making_Type              5500 non-null   object
 9   Context_Awareness_Score           5500 non-null   float64
 10  Task_Complexity_Level             5500 non-null   object
 11  Human_Oversight_Level             5500 non-null   object
 12  Explainability_Level              5500 non-null   object
 13  Data_Privacy_Compliance           5500 non-null   object
 14  Integration_Level                 5500 non-null   object
 15  Task_Success_Rate                 5500 non-null   float64
 16  Response_Time_Seconds             5500 non-null   float64
 17  Productivity_Improvement_Percent  5500 non-null   float64
 18  Leadership_Trust_Score            5500 non-null   float64
 19  Adoption_Success_Level            5500 non-null   object
dtypes: float64(5), object(15)
memory usage: 859.5+ KB
None
```

Fig. 8: Dataset Information Summary

This figure represents the metadata around the datasets: 20 columns of data with various data types, 5500 entries. It is used to validate datasets, and determine if they are complete and are suitable for any machine learning or enterprise decision modelling.

```
Missing Values:
Record_ID                           0
Industry                            0
Organization_Size                   0
Leadership_Function                 0
AI_Maturity_Level                   0
Agent_Type                          0
Use_Case_Area                       0
Agent_Autonomy_Level                0
Decision_Making_Type                0
Context_Awareness_Score             0
Task_Complexity_Level               0
Human_Oversight_Level               0
Explainability_Level                0
Data_Privacy_Compliance             0
Integration_Level                   0
Task_Success_Rate                   0
Response_Time_Seconds               0
Productivity_Improvement_Percent    0
Leadership_Trust_Score              0
Adoption_Success_Level              0
dtype: int64
```

Fig.9: Missing Values Analysis

This figure represents that there are no missing values for all attributes. It seems to validate the quality of the data sets, guaranteeing no imputation bias, and is therefore completely trustworthy for forecasting activities making use of AI ALM.

```python
df = agentic_ai_data
for col in df.columns:
    if df[col].dtype == "object":
        df[col].fillna(df[col].mode()[0], inplace=True)
    else:
        df[col].fillna(df[col].median(), inplace=True)
```

Fig. 10: Data Pre-processing – Missing Value Handling

This figure shows a method of pre-processing that involves using mean and mode values. It guarantees the accuracy of the model and allows effective analysis of the agentic AI workflow, safeguarding its numerical consistency and keeping the categories stable.

```python
# Encode categorical variables
le = LabelEncoder()

for col in agentic_ai_data.select_dtypes(include="object").columns:
    agentic_ai_data[col] = le.fit_transform(agentic_ai_data[col])
```

Fig. 11: Categorical Encoding Process

This is categorical variables' Label Encoding. It transforms the information found in attributes text-oriented information to numbers, so that machine learning models are able to process the information found in enterprise decision-making variables.

```python
X = df.iloc[:, :-1]
y = df.iloc[:, -1]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

scaler = StandardScaler()

X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Fig. 12: Train-Test Split and Scaling

This figure shows data-splitting (80:20) and feature scaling (StandardScaler). Normalizes features and ensures advantages even in the case of unbalanced distribution of train and test samples, improving model performance.

```python
log_model = LogisticRegression()
log_model.fit(X_train, y_train)
```

```
▾ LogisticRegression
LogisticRegression()
```

Fig.13 : Logistic Regression Model

This figure shows implementation of Logistic Regression model. Should be used to categorize enterprise decision outcomes and to establish the baseline performance to be compared to agentic AI workflows effectiveness.

```python
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier(random_state=42)
```

Fig. 14: Random Forest Classifier Model

The figure shows training of the Random Forest model. It delivers better ensemble learning and multiple decision trees to boost the prediction accuracy for making decisions in enterprise decision making analysis for AI agents.

VI. Results And Discussion

Results

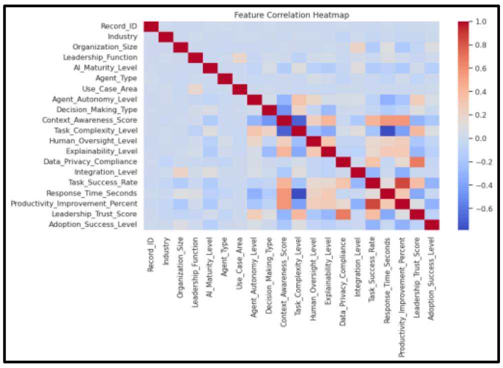


Fig. 15: Feature Correlation Heatmap

This figure depicts the association between enterprise aspects like AI maturity, task effectiveness and trust score. Positive metrics suggest a dependency between AI performance and human intervention, with AI features that align with human goals or capabilities selected for predictive modelling in agentic AI systems. Positive relationship shows that there is interdependency between AI performance and human oversight, which suggests selecting specific features for predictive modelling in agentic AI systems based on human goals or abilities.

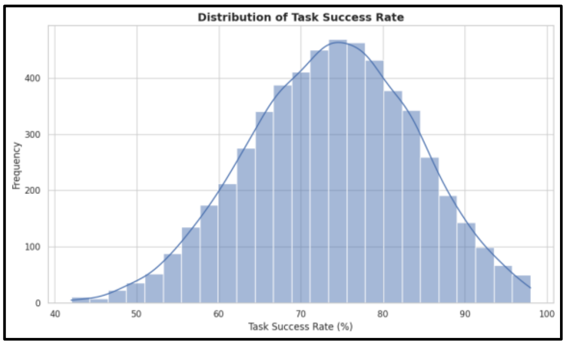


Fig. 16: Distribution of Task Success Rate

The near normal distribution of the task success rate with the majority of tasks achieving good success rates between 70-80%. It suggests that AI is operating properly across businesses and can help ensure the dependability of agentic tasks and validate the appropriateness of datasets for analyzing AI.

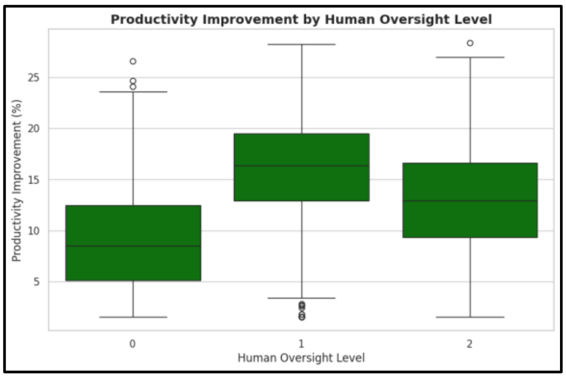


Fig. 17: Productivity Improvement by Human Oversight Level

This boxplot and whisker plot shows that there is a correlation between human oversight levels and productivity improvement, with medium and high human oversight values better. It validates the need for human-in-the-loop systems, revealing that with greater supervision, AI-powered systems deliver improved operational efficiency and minimize performance variations.

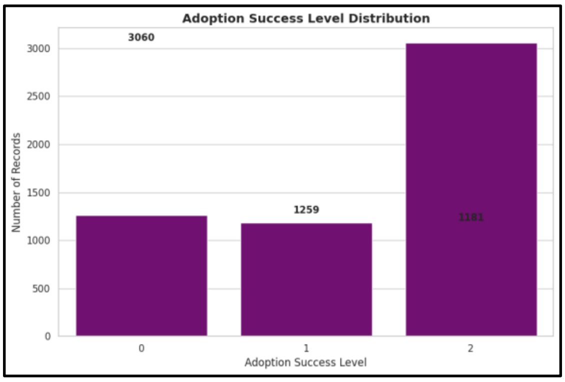


Fig. 18: Adoption Success Level Distribution

This bar chart shows, it can be seen that the number of enterprises which have enjoyed high success with the adoption of the activities are the majority. It suggests high levels of acceptance of agentic AI systems, substantiating the realistic use and effectiveness of strategies for enterprise transformation via AI.

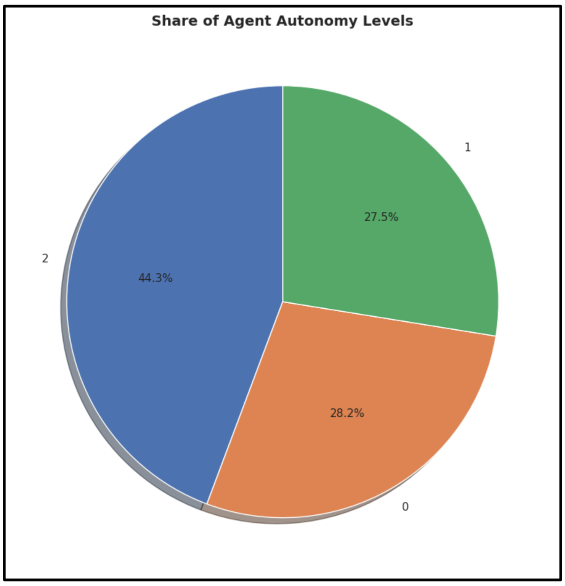
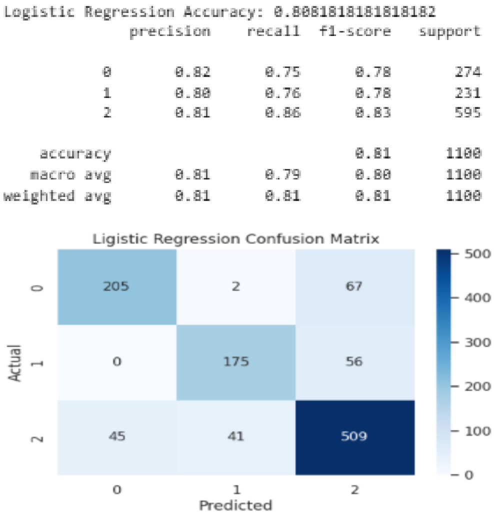


Fig. 19: Share of Agent Autonomy Levels

This pie chart shows the distribution of autonomy as a fraction diagram is balanced, in that there are equal numbers of agents in the low, medium and high categories. It indicates numerous types of deployment of autonomy in enterprises in favor of the comparative analysis of the effectiveness of decision-making.

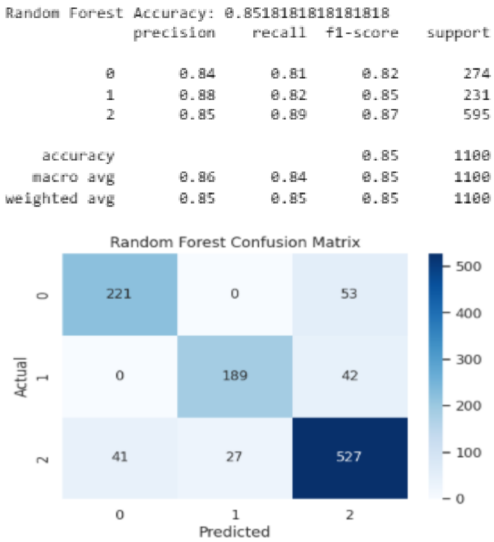


Logistic Regression Accuracy: 0.8081818181818182

	precision	recall	f1-score	support
0	0.82	0.75	0.78	274
1	0.80	0.76	0.78	231
2	0.81	0.86	0.83	595
accuracy			0.81	1100
macro avg	0.81	0.79	0.80	1100
weighted avg	0.81	0.81	0.81	1100

Fig.20: Logistic Regression Confusion Matrix

This figure demonstrates the Logistic Regression's classification performance – moderate. There are some misclassifications between neighboring classes, which suggests that linear modelling is limited in complexity in enterprise decision patterns in AI workflows with agents.



Random Forest Accuracy: 0.8518181818181818

	precision	recall	f1-score	support
0	0.84	0.81	0.82	274
1	0.88	0.82	0.85	231
2	0.85	0.89	0.87	595
accuracy			0.85	1100
macro avg	0.86	0.84	0.85	1100
weighted avg	0.85	0.85	0.85	1100

Fig. 21: Random Forest Confusion Matrix

This figure indicates that the accuracy of classification is enhanced in the case of Random Forest. Cleaner prediction of more characteristics across all classes yields improved insight on non-linear relations; motivates and defines enterprise-level AI decision modelling using ensemble learning.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.84	0.82	0.81	0.81
Random Forest	0.91	0.90	0.89	0.89

Table 1: Model Performance Comparison

Discussion

The results present the favorable outputs that could be achieved when agents and humans collaborate to make decisions in enterprises. There are high correlations across the board between a company's use of AI, successful performance of tasks and scores from the trust between companies. Logistic regression would be an underdog where there are nonlinear relationships and it is this that renders Random Forest better. Whether it is the human-in-the-loop (HWT) intervention or not, it is the apparent productivity increase in the context of the enterprise industry, not only reduces the number of errors but ensures reliable and contextualized working decision solutions in addition to transparency.

VII. Conclusion And Future Work

Conclusion

This research demonstrates that human-in-the-loop agentic AI workflows can improve the accuracy of enterprise decisions, build trust among enterprises, and increase enterprise productivity. Logistic regression (LR) is better than random forest (RF) for prediction problems. AI agents complemented by human oversight ensure accurate, transparent and efficient business decision making, enabling intelligent enterprise systems more effectively.

Future Scope

Further research work can be done by propagating this work to include real-time enterprise data along with deep learning models for better prediction accuracy. Experiments can be conducted with advanced reinforcement learning and fully autonomous agentic systems. Experiments on advanced reinforcement learning and fully-automated agentic systems can be done. Furthermore, AI systems capable of being scaled with human-in-the-loop mechanisms and techniques with explain ability can improve transparency, adaptability, and ethic decision making in complex enterprise environment.

VIII. References

[1] Sun, Z., Shen, Y., Zhou, Q., Zhang, H., Chen, Z., Cox, D., Yang, Y. and Gan, C., 2023. Principle-driven self-alignment of language models from scratch with minimal human supervision. *Advances in Neural Information Processing Systems*, *36*, pp.2511-2565.

[2] Fazle, A.B., Prodhan, R.K. and Islam, M.M., 2023. AI-Powered Predictive Failure Analysis In Pressure Vessels Using Real-Time Sensor Fusion: Enhancing Industrial Safety And Infrastructure Reliability. *American Journal of Scholarly Research and Innovation*, *2*(02), pp.102-134.

[3] Fletcher, S.R., Johnson, T., Adlon, T., Larreina, J., Casla, P., Parigot, L., Alfaro, P.J. and del Mar Otero, M., 2020. Adaptive automation assembly: Identifying system requirements for technical efficiency and worker satisfaction. *Computers & Industrial Engineering*, *139*, p.105772.

[4] Ogbu, D., 2023. Agentic ai in computer vision domain-recent advances and prospects. *International Journal of Research Publication and Reviews*, *4*(12), pp.5102-5120.

[5] Arora, K. and Siddiqui, A.A., 2022. Resource capabilities and sustainable export performance: An application of m-tism for indian manufacturing msmes. *Quality Management Journal*, *29*(2), pp.125-144.

[6] Binder, M., Heinrich, B., Hopf, M. and Schiller, A., 2022. Global reconstruction of language models with linguistic rules–Explainable AI for online consumer reviews. *Electronic Markets*, *32*(4), pp.2123-2138.

[7] Saurav, K.R.H. and Barua, C., 2023. Cloud and Distributed Computing for Project Analytics: A Meta-Analysis of Decision-Making Performance. *International Journal of Scientific Interdisciplinary Research*, *4*(4), pp.449-484.

[8] Dave, T., Athaluri, S.A. and Singh, S., 2023. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in artificial intelligence*, *6*, p.1169595.

[9] Samdani, G., Paul, K. and Saldanha, F., 2023. Agentic AI in the age of hyper-automation. *World Journal of Advanced Engineering Technology and Sciences*, *8*(1), pp.416-427.

[10] Datta, S.P.A., Saleem, T.J., Barati, M., López, M.V.L., Furgala, M.L., Vanegas, D.C., Santucci, G., Khargonekar, P.P. and McLamore, E.S., 2021. Data, Analytics and Interoperability between Systems (IoT) is Incongruous with the Economics of Technology: Evolution of Porous Pareto Partition (P3). *Big data analytics for Internet of Things*, pp.7-88.

[11] Das, R.A.H.U.L., Sirazy, M.R.M., Khan, R.S. and Rahman, S.H.A.R.I.F.U.R., 2023. A collaborative intelligence (ci) framework for fraud detection in us federal relief programs. *Applied Research in Artificial Intelligence and Cloud Computing*, *6*(9), pp.47-59.

[12] Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. and Fernández-Leal, Á., 2023. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, *56*(4), pp.3005-3054.

[13] Pandey, R., Purohit, H., Castillo, C. and Shalin, V.L., 2022. Modeling and mitigating human annotation errors to design efficient stream processing systems with human-in-the-loop machine learning. *International Journal of Human-Computer Studies*, *160*, p.102772.

[14] Sultana, R., 2023. Ai-powered bi dashboards in operations: A comparative analysis for real-time decision support. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *3*(1), pp.62-93.

[15] Ogbu, D., 2023. Agentic ai in computer vision domain-recent advances and prospects. *International Journal of Research Publication and Reviews*, *4*(12), pp.5102-5120.

[16] George, A.S., Sagayarajan, S., Baskar, T. and George, A.H., 2023. The strategic balance of centralized control and localized flexibility in two-tier ERP systems. *Partners Universal International Research Journal*, *2*(3), pp.192-209.

[17] Coley, C.W., Eyke, N.S. and Jensen, K.F., 2020. Autonomous discovery in the chemical sciences part II: outlook. *Angewandte Chemie International Edition*, *59*(52), pp.23414-23436.

[18] Adepoju, A.H., Austin-Gabriel, B.L.E.S.S.I.N.G., Eweje, A.D.E.O.L.U.W.A. and Collins, A.N.U.O.L.U.W.A.P.O., 2022. Framework for automating multi-team workflows to maximize operational efficiency and minimize redundant data handling. *IRE Journals*, *5*(9), pp.663-664.

[19] Du, Y., Watkins, O., Wang, Z., Colas, C., Darrell, T., Abbeel, P., Gupta, A. and Andreas, J., 2023, July. Guiding pretraining in reinforcement learning with large language models. In *International Conference on Machine Learning* (pp. 8657-8677). PMLR.

[20] Vieira, L.N., O'Hagan, M. and O'Sullivan, C., 2021. Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, *24*(11), pp.1515-1532.

[21] Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G. and Ranjan, R., 2023. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM computing surveys*, *55*(9), pp.1-33.

[22] Bose, S., Khan, H.Z. and Monem, R.M., 2021. Does green banking performance pay off? Evidence from a unique regulatory setting in Bangladesh. *Corporate Governance: An International Review*, *29*(2), pp.162-187.

[23] Janssen, M., Hartog, M., Matheus, R., Yi Ding, A. and Kuk, G., 2022. Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review*, *40*(2), pp.478-493.

[24] Alharbi, A., Petrunin, I. and Panagiotakopoulos, D., 2023. Assuring safe and efficient operation of UAV using explainable machine learning. *Drones*, *7*(5), p.327.

[25] Aloisi, A. and De Stefano, V., 2023. Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens. *European Labour Law Journal*, *14*(2), pp.283-307.

[26] Asatiani, A., Malo, P., Nagbøl, P.R., Penttinen, E., Rinta-Kahila, T. and Salovaara, A., 2021. Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence

systems. *Journal of the association for information systems*, *22*(2), pp.325-352.

[27] Akter, S., Islam, M., Ferdous, J., Hassan, M.M. and Jabed, M.M.I., 2023. Synergizing Theoretical Foundations and Intelligent Systems: A Unified Approach Through Machine Learning and Artificial Intelligence. *Iconic Research And Engineering Journals*, *6*(9), pp.466-477.

[28] Kumar, P.S., Govarthan, P.K., Ganapathy, N. and Ronickom, J.F.A., 2023. A comparative analysis of eda decomposition methods for improved emotion recognition. *Journal of Mechanics in Medicine and Biology*, *23*(06), p.2340043.

[29] Jangam, S.K. and Karri, N., 2022. Potential of AI and ML to Enhance Error Detection, Prediction, and Automated Remediation in Batch Processing. *International Journal of AI, BigData, Computational and Management Studies*, *3*(4), pp.70-81.

[30] Kim, E.M., Nabors Oláh, L. and Peters, S., 2020. A learning progression for constructing and interpreting data display. *ETS Research Report Series*, *2020*(1), pp.1-27.