International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1531~1542
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1531-1542

A risk-constrained SARSA–FIS hybrid decision architecture with adaptive exploration control

Joni Fat[1,2], Parwadi Moengin[1], Pudji Astuti[1], Sally Cahyati[3]
[1]Department of Industrial Engineering, Universitas Trisakti, Jakarta, Indonesia
[2]Electrical Engineering Program, Universitas Tarumanagara, Jakarta, Indonesia
[3]Department of Mechanical Engineering, Universitas Trisakti, Jakarta, Indonesia

Article Info

Article history:

Received Mar 4, 2026
Revised Apr 23, 2026
Accepted Apr 26, 2026

Keywords:

Algorithmic trading
Fuzzy inference system
Reinforcement learning
Risk-constrained learning
SARSA

ABSTRACT

Algorithmic trading systems operate in highly dynamic and uncertain environments where learning-based decision agents must balance adaptability with strict risk control. Reinforcement learning (RL) methods provide adaptive policy optimization but often suffer from unstable exploration and limited interpretability in financial markets. This study proposes a risk-constrained SARSA–FIS hybrid decision architecture with adaptive exploration control for algorithmic trading. The framework integrates a compact SARSA-based reinforcement learning environment with a Sugeno-type fuzzy inference system (FIS) that converts reinforcement signals into interpretable trading decisions. Exploration follows a decaying ε-greedy policy with a drawdown-triggered reset mechanism to maintain bounded risk exposure during learning. The system was implemented as a MetaTrader 5 Expert Advisor and evaluated on the GBPUSD currency pair using historical market data. Experimental results show that the hybrid framework improves trading performance compared with a rule-based baseline. During a six-month out-of-sample evaluation, the system achieved a net profit of 90 USD and a profit factor of 1.35, compared with 10 USD and 1.02 for the baseline. Extended one-year testing confirmed stable profitability and controlled drawdown behavior. The results demonstrate that integrating reinforcement learning, fuzzy decision mapping, and explicit risk constraints provides a practical approach for developing adaptive trading agents.

Corresponding Author:

Parwadi Moengin
Department of Industrial Engineering, Universitas Trisakti
Jakarta, Indonesia
Email: parwadi@trisakti.ac.id

1. INTRODUCTION

Algorithmic trading systems operate in environments that are uncertain, dynamic, and often non-stationary. In financial markets, price movements are affected by volatility clustering, sudden regime changes, and complex interactions between participants. An automated decision system must therefore be adaptive, but at the same time it must maintain strict risk control. Designing a learning architecture that balances adaptability, interpretability, and operational stability remains an important challenge in intelligent financial decision systems. Reinforcement learning (RL) has been widely used for sequential decision-making under uncertainty. Methods such as Q-learning and SARSA allow an agent to update its action-value function through direct interaction with the environment. These approaches provide online learning capability and adaptability to changing conditions. However, RL models often behave as black-box mechanisms and may lack interpretability. In addition, unconstrained exploration policies can generate unstable decisions, especially when applied in real-time financial environments where risk exposure must be controlled.

Fuzzy inference systems (FIS), particularly Sugeno-type models, offer an interpretable rule-based structure for reasoning under uncertainty. Through linguistic variables and structured rule bases, fuzzy systems can represent expert knowledge in an explicit manner. Nevertheless, classical fuzzy systems are usually static. Without an embedded learning mechanism, they cannot automatically adapt to evolving market states or changing volatility regimes. For these reasons, integrating RL with fuzzy inference becomes a promising direction. A hybrid structure can combine the adaptability of RL with the interpretability of fuzzy rules. However, many existing hybrids approaches mainly focus on prediction performance, while limited attention is given to risk-aware exploration control and bounded learning dynamics. In practical financial systems, decision policies must operate inside predefined risk limits in order to prevent destabilizing behavior during high-volatility periods.

This paper proposes a risk-constrained SARSA–FIS hybrid decision architecture with adaptive exploration control. The framework is built on a compact SARSA-based Markov decision process and a 27-rule Sugeno fuzzy inference layer that transforms reinforcement signals into structured trading decisions. The exploration mechanism follows a decaying ε-greedy policy with an additional drawdown-triggered reset, so that learning remains adaptive but does not exceed predefined risk boundaries. Furthermore, position-level profit and loss thresholds together with a maximum drawdown sentinel are incorporated to regularize the overall system behavior. The proposed architecture is implemented as a modular execution agent and evaluated using historical data from the GBPUSD currency pair. The experimental evaluation includes comparison with a rule-based baseline, ablation analysis, and multi-horizon back testing in order to assess profitability, risk-adjusted efficiency, and robustness across different market conditions.

Despite the growing adoption of RL in financial decision systems, several limitations remain. Existing RL approaches frequently rely on unconstrained exploration mechanisms that may generate unstable policies when deployed in volatile financial environments. Although FIS provide interpretability and structured reasoning, they are typically static and lack adaptive learning capability. Previous hybrid approaches combining RL and FIS primarily emphasize predictive performance or policy optimization, while limited attention has been given to regulating exploration behavior under explicit risk constraints. In practical trading systems, uncontrolled exploration can lead to excessive exposure and unstable decision dynamics. Consequently, there remains a need for a compact hybrid decision architecture that integrates adaptive RL, interpretable fuzzy reasoning, and explicit mechanisms for risk-constrained exploration.

This study proposes a risk-constrained SARSA–FIS hybrid decision architecture designed to address the limitations of conventional reinforcement learning systems in financial environments. The framework integrates a compact SARSA learning environment with a Sugeno-type FIS that maps reinforcement signals into interpretable trading decisions. In addition, an adaptive exploration control mechanism with a drawdown-triggered reset is introduced to regulate exploration dynamics under bounded risk conditions. The architecture is implemented as a modular execution agent in MetaTrader 5 and evaluated through multi-horizon back testing experiments on the GBPUSD currency pair. The significance of the proposed architecture lies in demonstrating that RL, interpretable fuzzy decision mapping, and explicit risk constraints can be integrated within a lightweight operational trading system. By combining adaptive learning with structured decision rules, the framework supports both policy adaptability and decision transparency. Such characteristics are particularly important for safety-critical artificial intelligence (AI) applications where learning algorithms must operate within predefined risk boundaries. The proposed SARSA–FIS architecture therefore contributes to the development of interpretable and risk-aware intelligent decision systems suitable for dynamic and uncertain environments.

2. THEORETICAL FOUNDATION AND RELATED WORK

RL has been widely adopted for sequential decision-making problems under uncertainty. In financial trading research, early studies applied SARSA(λ), Q(λ), and TD(λ) agents for portfolio management and reported improvements compared to static rule-based benchmarks [1]. Further developments introduced directional change RL to improve robustness in high-frequency and volatile conditions [2]. More recent studies applied deep RL architectures, including convolutional and recurrent networks, for feature extraction and dynamic policy learning [3]. Hybrid ensemble approaches such as PPO, A2C, and SAC have also been proposed to increase adaptability across different market regimes [4]. Other works integrated recurrent and actor–critic structures to improve risk handling in non-stationary environment [5], [6]. Reviews of foreign exchange prediction models show that gated recurrent unit (GRU), long short-term memory (LSTM) and transformer-based models dominate recent research, reflecting the growing maturity of RL-based financial systems [7]. These developments demonstrate strong adaptability of RL, but they also introduce high computational complexity and limited interpretability.

In parallel, FIS have been applied in financial prediction due to their capability to manage uncertainty and encode expert reasoning. Neural–fuzzy models have shown competitive performance in stock and commodity forecasting tasks [8], [9]. Other approaches combined fuzzy logic with Kalman filtering and particle swarm optimization to improve robustness under noisy conditions [10]. RL and fuzzy hybrids have also been explored, especially in environments where interpretability is required together with adaptive learning [3]. In addition, stability analysis of fuzzy-based intelligent systems has been investigated, including asymptotic stability and periodic behavior under uncertainty and delay conditions [11]. These studies highlight that fuzzy systems provide transparency and structured reasoning, while RL provides adaptability.

Game-theoretic perspectives have also been applied to financial systems to model strategic interaction among agents. Mean-field games have been used to study optimal liquidation under crowding effects [12], while stochastic trading games have been proposed to represent heterogeneous beliefs in market dynamics [13]. Other works analyze financial stability and systemic resolution using game-theoretic frameworks [14], [15]. Game theory has also been used to explain liquidity formation in order-book markets and to model stock prices through Markov-based game structures [16], [17]. Although these approaches provide useful theoretical insights, most of them remain conceptual and are rarely embedded directly into lightweight operational trading architectures [18].

Overall, prior literature shows that RL provides adaptive policy optimization, fuzzy inference offers interpretability and structured reasoning, and game-theoretic analysis provides a perspective on bounded strategic behavior [19]. However, existing studies rarely combine these elements into a compact and risk-constrained decision architecture suitable for real-time execution. In particular, limited attention has been given to integrating adaptive exploration control with explicit drawdown constraints in hybrid reinforcement–fuzzy systems. Based on this theoretical background, the present study proposes a lightweight SARSA–FIS hybrid architecture with adaptive exploration control. The framework aims to combine online RL, interpretable fuzzy rule mapping, and bounded risk regulation within a unified decision structure [19]. The proposed method is positioned as a compact and computationally efficient alternative to deep RL approaches, while maintaining adaptability and operational stability under predefined risk limits.

3. METHOD

This section describes the experimental design and implementation of the risk-constrained SARSA–FIS hybrid decision architecture. The framework integrates a compact RL environment, a Sugeno-type FIS, adaptive exploration control, and multi-level risk constraints within a unified execution agent. The overall objective of the design is to create an interpretable and risk-aware learning framework capable of adapting to changing market conditions while maintaining controlled trading behavior.

3.1. Data and execution environment

The asset used in this study is the GBPUSD currency pair. The SARSA agent was trained on historical data from 30 June 2020 to 2 July 2024. Out-of-sample evaluation was conducted from 1 January to 30 June 2025. Additional extended testing on 2023 data was performed to assess robustness under different market regimes. The system was implemented as an expert advisor (EA) in MetaTrader 5 using the MQL5 programming language. All trading decisions operate on the M15 timeframe. Technical indicators used for feature extraction include $EMA(50)$, $EMA(10)$, $MACD(12,26,9)$, $ATR(14)$, and EMA-based standard deviation. For reproducibility, the Q-table, fuzzy rule matrix, and drawdown status are persisted to external text files after each learning update.

3.2. SARSA state–action environment

The RL component is formulated as a compact Markov decision process with three states and three actions. The state space S = {S1, S2, S3} is defined according to price distance from EMA(50) using a ±20 pip barrier. State S1 represents price below $EMA(50)$ minus 20 pips, S2 represents price within ±20 pips of the EMA, and S3 represents price above $EMA(50)$ plus 20 pips. The action space $A = \{Stay, Next, Jump\}$ is encoded as values 0 to 2 and selected through an ε-greedy policy. Transition dynamics are governed by a fixed 3×3×3 tensor that introduces controlled stochasticity in next-state sampling. The reward structure is defined as a shaped cube $R[s, a, s']$, balancing exploration penalties and trend-following incentives. Actions that maintain unfavorable states receive negative rewards, while transitions aligned with favorable price movements receive positive rewards. The SARSA update rule is applied at every market tick as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\,(r + \gamma\, Q(s',a') - Q(s,a)) \tag{1}$$

with discount factor γ=0.9.

The learning rate α is adaptive. It is scaled using $ATR(14)$ and EMA-based standard deviation to reflect volatility clustering and is bounded within the interval [0.01, 0.30]. In trending regimes, an additional bounded momentum adjustment derived from EMA deviation modifies α within ±0.1 to accelerate adaptation. Figure 1 illustrates the SARSA environment, showing the three states, available actions, and reward relationships that guide the learning process.

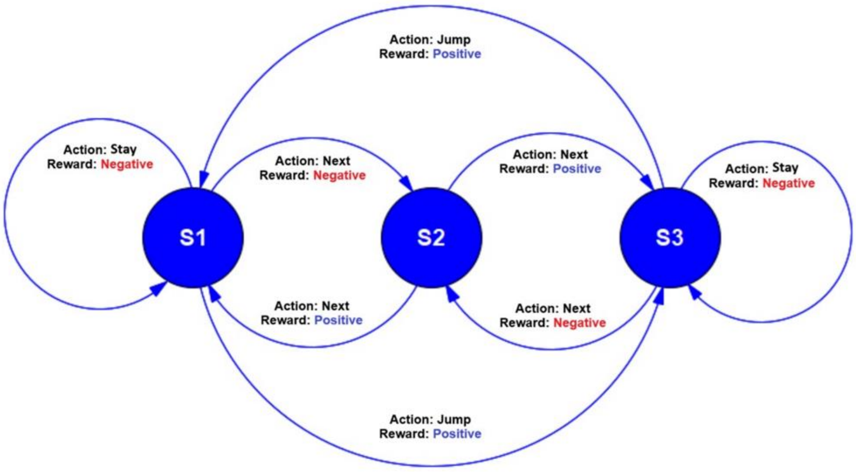

Figure 1. SARSA environment with three states (S1–S3), available actions, and corresponding reward signals

3.3. Adaptive exploration control

Exploration follows a decaying ε-greedy policy defined as

$$\varepsilon_t = max(\varepsilon_{\min}, \varepsilon_0 \delta^t) \quad (2)$$

where $\epsilon_0 = 0.30$, $\epsilon_{\min} = 0.05$, and $\delta = 0.999$. To enforce risk-constrained learning, a drawdown-triggered reset mechanism is introduced. If floating equity drawdown exceeds 20% of the running peak, ε is reset to ϵ_0. After reset, exponential decay resumes. This mechanism allows renewed exploration during regime shifts while maintaining bounded risk exposure.

3.4. Fuzzy inference integration

The reinforcement signals generated by SARSA are mapped into structured trading decisions using a Sugeno-type FIS consisting of 27 rules arranged in a 3×3×3 structure. The inputs of the FIS are the normalized current-state Q-value, the normalized predicted next-state Q-value, and the immediate reward. Each input variable is partitioned into three linguistic levels: Low, Moderate, and High, defined using trapezoidal and triangular membership functions.

The rule base maps the triplet (Current State, Predicted State, Reward) to Sugeno consequents in the set {0.5, 1.5, 2.5}, corresponding to Sell, Hold, and Buy signals. Defuzzification is performed using the weighted average method, producing a continuous control signal that is subsequently discretized. Output values below 1.0 indicate Sell, values around 1.5 indicate Hold, and values above 2.0 indicate Buy. The fuzzy rule matrix can be persisted to external storage for incremental tuning and reproducibility. Representative rule mappings are summarized in Table 1, illustrating the relationship between Current State (CS), Predicted State (PS), Reward (RW), and the corresponding Sugeno output.

Table 1. Representative Sugeno rules mapping CS, PS, and RW to trading signals

Current State	Predicted State	Reward	Output	Signal
Low	Low	Low	0.5	Sell
Moderate	Moderate	Moderate	1.5	Hold
High	High	High	2.5	Buy
Low	High	Moderate	1.5	Hold
High	Low	High	0.5	Sell

3.5. Execution logic and risk constraints

Trading is executed exclusively on GBPUSD. The Sugeno output serves as the primary signal gate and must be confirmed by $EMA(10)$ crossover and MACD histogram momentum before trade entry. Risk management is implemented at multiple levels. Each position includes a take-profit threshold of 10 USD, with an auxiliary closure at 12 USD to secure extended gains. A cut-loss threshold of -8 USD is applied when the fuzzy system indicates reversal conditions. At the account level, a maximum drawdown kill-switch is implemented via an external file. If activated, all positions are closed and trading is halted. To mitigate trade clustering, a mandatory interval of four hours is imposed between consecutive trades. These constraints ensure that learning and execution operate within predefined risk boundaries.

3.6. Baseline and comparative configurations

Three configurations are evaluated under identical market data, transaction costs, and risk constraints. The baseline configuration applies rule-based EMA and MACD filters with identical profit and loss parameters but without learning or fuzzy inference. The RL-only configuration applies SARSA directly using argmax $Q(s, a)$ for action selection. The hybrid configuration integrates SARSA, Sugeno FIS, and adaptive exploration control. This design isolates the contribution of hybrid learning and risk-constrained exploration.

3.7. System architecture and reproducibility

The architecture is implemented as a modular EA consisting of a SARSA learning module, fuzzy inference module, risk manager, and trade executor. At each market tick, technical indicators are computed and passed to the SARSA module. Updated Q-values and rewards are processed by the fuzzy system to generate a trading score. The signal is validated by confirmation filters and processed by the risk manager before execution. Executed trades are fed back into the learning module, enabling continuous adaptation.

Figure 2 presents the system architecture and signal flow from market data acquisition to execution and feedback. For reproducibility, the Q-table and fuzzy weight matrix are written to disk after each learning step. Upon reinitialization, the most recent state is restored, enabling continuation of training across sessions.

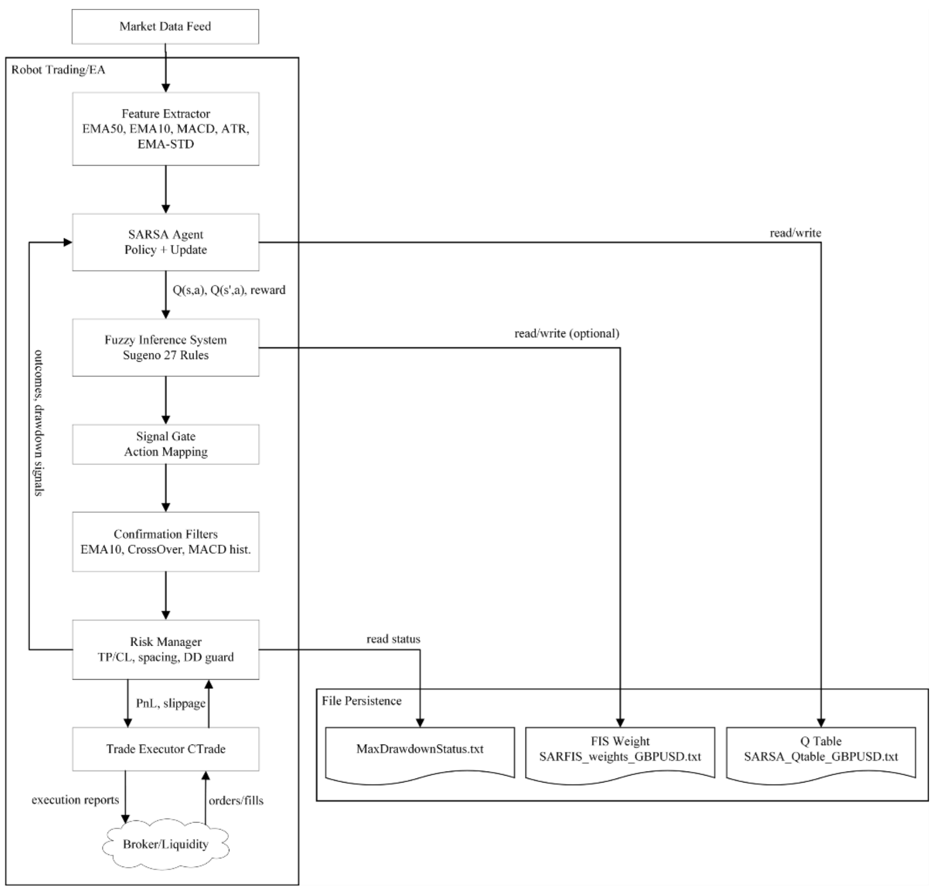

Figure 2. Component-level architecture of the SARSA–FIS hybrid system

The dynamic operational sequence of the framework is illustrated in Figure 3, showing the pipeline from market data acquisition to signal generation, risk validation, execution, and feedback-driven learning. The experimental setup of the trading system, including data flow, learning modules, and execution environment, is illustrated in Figures 2 and 3 to ensure reproducibility.

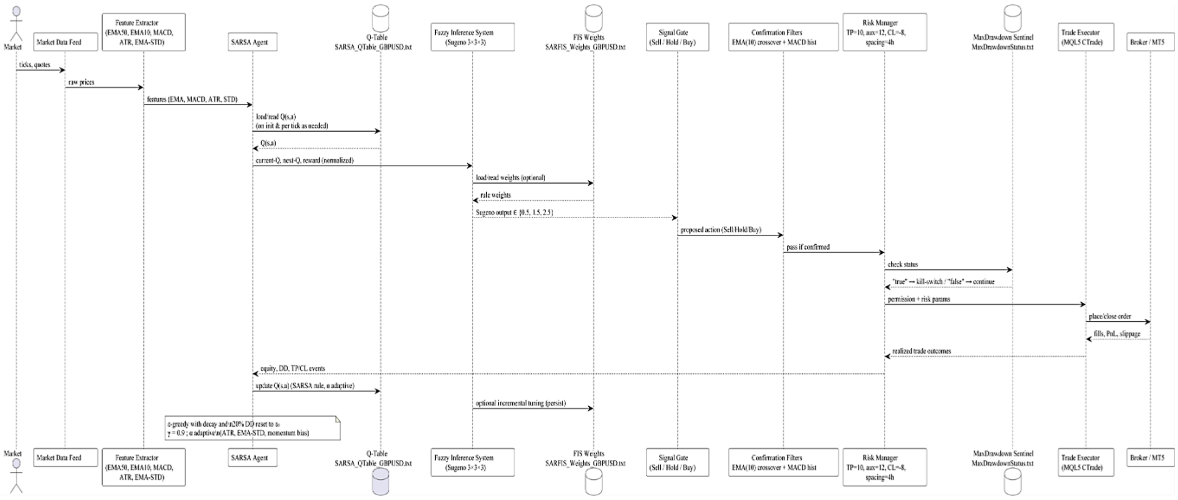

Figure 3. System flow diagram of the SARSA–FIS hybrid trading agent

4. RESULTS AND DISCUSSION

Back testing experiments were conducted on the GBPUSD currency pair to evaluate the performance of the risk-constrained SARSA–FIS hybrid framework. The six-month out-of-sample evaluation covers the period from 1 January to 30 June 2025. An extended one-year evaluation on 2023 data was performed to assess robustness under different volatility regimes. All configurations were tested under identical market data, transaction costs, and execution conditions.

4.1. Six-month out-of-sample evaluation (2025)

Figure 4 presents the comparative equity curves of the hybrid SARSA–FIS framework and the baseline rule-based configuration during the 2025 evaluation period. The hybrid framework exhibits a stable upward balance trajectory with limited equity retracements. The system achieved a net profit of approximately 90 USD, whereas the baseline configuration produced approximately 10 USD under the same conditions. The equity growth of the hybrid model shows smoother progression and more consistent recovery following moderate drawdowns.

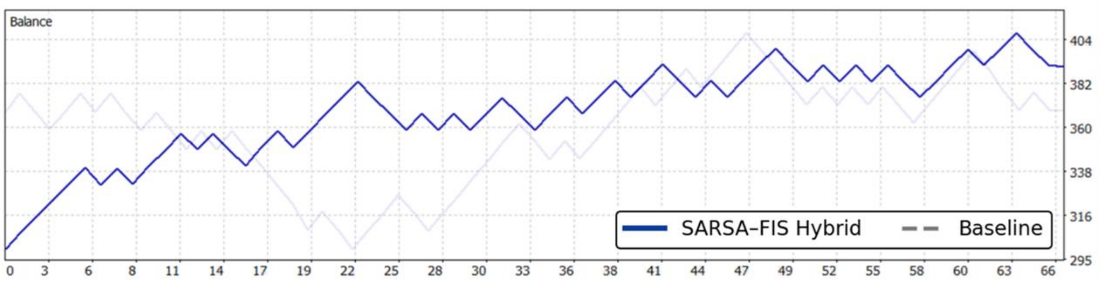

Figure 4. Comparative balance curves of the SARSA–FIS hybrid framework and the baseline system during the six-month 2025 evaluation period

Trade-level behavior is illustrated in Figure 5, which shows the distribution of individual trade outcomes over time. The SARSA–FIS framework demonstrates clear clustering of trade results around the predefined take-profit (+10 USD) and cut-loss (−8 USD) thresholds. This pattern indicates systematic enforcement of the intended risk–reward structure. In contrast, the baseline configuration displays wider dispersion, with several trades extending into deeper adverse movements before closure. The tighter clustering observed in the hybrid system reflects disciplined exit control and bounded downside exposure.

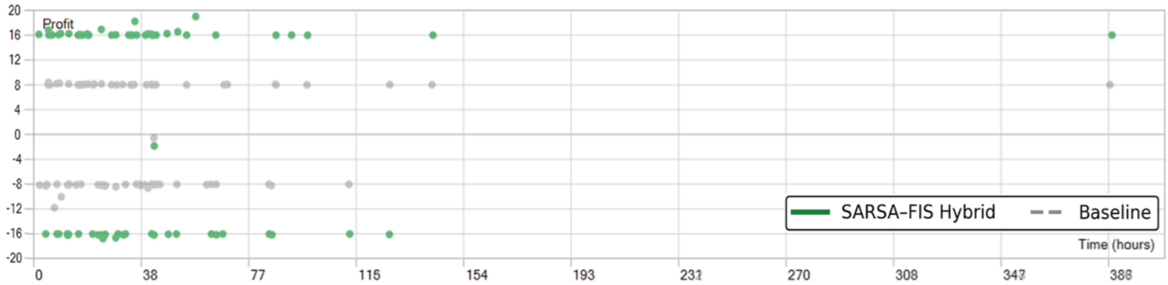

Figure 5. Profit–time distribution of the SARSA–FIS hybrid framework, shown together with the baseline system on a common reference axis to highlight differences in trade dispersion and risk containment

Comparative performance metrics for the six-month period are summarized in Table 2. The hybrid framework achieved a profit factor of 1.35 and a higher daily Sharpe ratio relative to the baseline, while maintaining comparable maximum drawdown. These results indicate improved risk-adjusted performance without increasing drawdown severity.

Table 2. Comparative performance metrics of the baseline (no AI) and SARSA–FIS hybrid framework on GBPUSD

Metric	NO AI	SARSA–FIS Hybrid
Total net profit (USD)	10	90
Profit factor	1.02	1.35
Maximum drawdown (%)	5.3	5.5
Sharpe ratio (daily)	0.02	0.18
Average trade return (%)	0.06	0.52
Win rate (%)	52.1	58.7
Exposure time (%)	42.5	47.2
Trade count	170	176

4.2. Extended one-year evaluation (2023)

To assess robustness beyond the six-month window, the SARSA–FIS framework was back tested over the full year of 2023. This period includes varying volatility regimes and macroeconomic conditions. Figure 6 presents the equity curve for the 2023 evaluation. The system achieved a net profit of 137.06 USD with a profit factor of 1.27 and a Sharpe ratio of 1.25. Maximum drawdown during this period was 25.39%. The annual equity trajectory shows gradual upward progression with structured recovery phases following drawdowns, indicating stable behavior over a longer horizon.

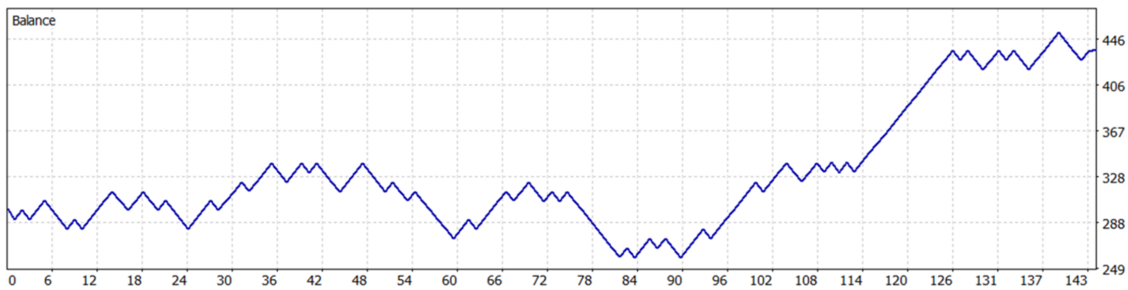

Figure 6. One-year balance curve of SARSA–FIS hybrid framework (2023)

Aggregate performance metrics for 2023 are summarized in Table 3. The system executed 144 trades with a win rate of 56.25%. Average profit and loss per trade were 7.97 USD and −8.08 USD, respectively, indicating stable trade expectancy under varying market conditions. The balanced gross profit and loss values further confirm controlled trade management and consistent application of the predefined risk thresholds.

Table 3. Performance metrics of SARSA–FIS hybrid framework during one-year evaluation (2023)

Metric	Value
Total net profit (USD)	137.06
Gross profit (USD)	645.90
Gross loss (USD)	–508.84
Profit factor	1.27
Sharpe ratio	1.25
Maximum drawdown (%)	25.39
Total trades	144
Win rate (%)	56.25
Average profit trade (USD)	7.97
Average loss trade (USD)	–8.08

4.3. Sensitivity analysis of exploration and risk parameters

To further evaluate the stability of the proposed architecture, a sensitivity analysis was conducted on the main learning and risk-control parameters. The analysis focuses on the exploration rate and drawdown reset threshold that regulate adaptive exploration during learning. The initial exploration rate parameter varied within a range of moderate values while maintaining the same decay schedule, and the drawdown reset threshold was adjusted within a bounded interval around the baseline configuration. The experimental results indicate that the SARSA–FIS framework maintains stable performance across these parameter variations. Profit factor values remain consistently above the baseline configuration, while maximum drawdown levels remain within controlled limits. These observations suggest that hybrid learning architecture is not highly sensitive to moderate parameter adjustments. The adaptive exploration mechanism effectively regulates policy updates while maintaining bounded risk exposure. Overall, the sensitivity analysis confirms that the proposed architecture exhibits stable learning behavior and robust performance across different parameter settings. This stability is particularly important for practical deployment, where parameter tuning may vary depending on market conditions and system configuration.

4.4. Robustness and consistency analysis

Robustness is evaluated by comparing results across non-overlapping evaluation horizons. In both the six-month and one-year tests, the SARSA–FIS hybrid framework demonstrates persistent profitability, bounded drawdown behavior, and improved Sharpe ratios relative to the baseline configuration. The equity curves in Figures 4 and 6 show sustained upward trajectories rather than isolated performance spikes or short-lived gains. The clustering of trade outcomes observed in Figure 5 confirms consistent enforcement of take-profit and cut-loss constraints, indicating that adaptive learning does not lead to destabilizing exposure. The persistence of positive profit factors across different time horizons suggests that performance improvements are not limited to a specific market phase. Overall, the results demonstrate that the risk-constrained SARSA–FIS hybrid architecture improves risk–reward efficiency while maintaining disciplined exposure control. The integration of adaptive RL, structured fuzzy decision mapping, and drawdown-regulated exploration contribute to stable and reproducible performance under varying market conditions. Beyond these empirical observations, the stability characteristics observed in the experimental results also provide insight into the sustainability implications of the proposed decision architecture.

4.5. Sustainability implications

The empirical results presented in the previous subsections provide indicative evidence that the proposed SARSA–FIS hybrid framework contributes to sustainable trading behavior. In the context of algorithmic trading, sustainability refers to the ability of a decision system to maintain consistent profitability while preserving disciplined risk control and stable operational behavior across different market conditions. First, the profitability outcomes observed across both evaluation horizons indicate persistent positive performance. During the six-month 2025 evaluation, the SARSA–FIS hybrid framework achieved a net profit of approximately 90 USD with a profit factor of 1.35. In the extended one-year 2023 evaluation, the system produced a net profit of 137 USD with a profit factor of 1.27. Although the absolute profit values remain moderate, the consistency of profitability across non-overlapping periods suggests that the performance is not limited to a single market phase. Similar observations have been reported in RL studies showing that adaptive trading agents can improve financial decision-making performance relative to conservative rule-based benchmarks [1], [4].

Second, the risk–reward characteristics of the framework further support the sustainability interpretation. The system maintained bounded maximum drawdowns of approximately 5.5% during the 2025 evaluation and about 25.4% during the longer 2023 evaluation. At the same time, the hybrid architecture achieved improved risk-adjusted performance, reflected by Sharpe ratios of 0.18 and 1.25 for the respective periods. Trade-level behavior indicates that favorable price movements are captured while adverse movements remain bounded by predefined risk thresholds. Such balanced behavior is consistent with previous studies on fuzzy–RL hybrid systems that emphasize robustness and stability under uncertain environments [3], [20]. Third, the operational characteristics of the trading agent also indicate stable system behavior. Most trades follow a short intraday holding pattern, suggesting controlled exposure and disciplined turnover [21], [22]. This operational structure reduces prolonged market exposure and allows the system to adapt more effectively to changing volatility regimes. From a strategic perspective, such disciplined behavior can be interpreted within the context of repeated strategic interactions in financial markets, where agents must maintain stable strategies under varying competitive conditions [12], [13]. Finally, the integration of explicit risk constraints strengthens the sustainability perspective of the proposed architecture. Mechanisms such as capped drawdowns, predefined profit and loss thresholds, and confirmation filters function as practical regulators that prevent destabilizing trading behavior [23]. These constraints can be interpreted as payoff-regularizing mechanisms that guide the learning agent toward stable decision policies. Similar ideas have been discussed in financial stability research, where bounded decision rules help maintain systemic robustness under uncertainty [14], [15].

From a responsible AI perspective, the proposed framework also supports transparency and controlled deployment [24]. The use of a FIS enhances interpretability of the decision process, while the RL component enables adaptive responses to changing market dynamics. These design principles align with recent discussions on responsible and sustainable AI deployment in financial systems [25]–[28]. Overall, the combination of consistent profitability, bounded risk exposure, and stable operational behavior suggests that the SARSA–FIS hybrid framework provides a sustainable approach to algorithmic trading under varying market conditions.

4.6. Limitations and deployment considerations

Although the experimental evaluation demonstrates stable and profitable behavior, several limitations should be acknowledged. First, experimental validation is conducted using historical back testing rather than real-time trading environments. Live market deployment may introduce additional operational factors such as execution latency, slippage, and liquidity constraints that cannot be fully captured through historical simulations. These factors may influence the realized performance of the trading agent in practical settings. Second, the proposed architecture relies on a compact state representation consisting of three states derived from price deviations relative to the EMA indicator. While this simplified structure improves computational efficiency and interpretability, it may limit scalability when applied to more complex decision environments involving high-dimensional feature spaces or multi-asset portfolios. Extending the framework to larger state spaces or multiple assets may require more advanced learning mechanisms or hierarchical architecture. Third, the fuzzy inference layer employs a predefined rule structure designed to maintain interpretability and decision transparency. However, manual rule design may become increasingly complex as the dimensionality of the input space grows. Automated rule adaptation or rule-learning mechanisms may therefore be required when applying the architecture to more complex domains. Despite these limitations, the proposed architecture provides a practical foundation for developing interpretable reinforcement learning systems with explicit risk constraints. Future research will investigate extensions to multi-asset trading environments, integration with deep reinforcement learning models, and automated fuzzy rule optimization to improve scalability and deployment capability in real-world financial systems.

5. CONCLUSION

This study proposed and experimentally evaluated a risk-constrained SARSA–FIS hybrid decision architecture for algorithmic trading. The framework integrates reinforcement learning with a Sugeno-type fuzzy inference system and explicit risk-control mechanisms within a unified execution agent. Architecture combines the adaptability of SARSA learning with the interpretability of fuzzy decision rules, while adaptive exploration control and drawdown constraints regulate system stability during learning. Back testing experiments on the GBPUSD currency pair indicate that the proposed framework achieves improved performance compared with a rule-based baseline configuration across both short-term and extended evaluation horizons. During the six-month out-of-sample evaluation in 2025, the hybrid system generated a net profit of approximately 90 USD with a profit factor of 1.35. In the one-year evaluation on 2023 data, the system achieved a net profit of 137 USD with a profit factor of 1.27 and a Sharpe ratio of 1.25. The results also show disciplined drawdown behavior and consistent enforcement of predefined risk thresholds. Trade-level outcomes demonstrate that the hybrid framework maintains bounded loss exposure while capturing favorable price movements more effectively than the baseline configuration.

The results further indicate that adaptive learning, interpretable decision rules, and explicit risk constraints can be combined within a lightweight trading architecture without introducing unstable behavior. The SARSA component enables continuous policy adaptation, while the fuzzy inference layer provides structured decision mapping that improves transparency and operational stability. Together with drawdown-regulated exploration control, these mechanisms support consistent performance across different market conditions. Overall, the findings demonstrate that a compact SARSA–FIS hybrid architecture can provide a practical approach for developing adaptive and risk-aware trading agents. The proposed framework contributes to the design of intelligent decision systems that balance learning capability, interpretability, and disciplined risk management. Future work will extend the architecture to multi-asset environments, investigate alternative reinforcement learning variants, and further explore adaptive mechanisms for improving robustness in highly volatile financial markets.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Doctoral Program in Industrial Engineering, Universitas Trisakti, Jakarta, Indonesia, where this research was conducted. The author also acknowledges the Electrical Engineering Program, Universitas Tarumanagara, Jakarta, Indonesia, for academic and institutional support.

FUNDING INFORMATION

This research was supported by the Institute for Research and Community Service (LPPM), Universitas Tarumanagara, Jakarta, Indonesia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Joni Fat	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓
Parwadi Moengin	✓	✓			✓					✓		✓		
Pudji Astuti	✓	✓		✓						✓		✓		
Sally Cahyati				✓		✓				✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

INFORMED CONSENT

Not applicable. This study does not involve human participants or personal data.

ETHICAL APPROVAL

Not applicable. This study does not involve human participants, animals, or personal identifiable data.

DATA AVAILABILITY

The data and implementation resources supporting the findings of this study are available in the public repository: *https://github.com/jonifat/SARSA-FIS*. The repository includes the core algorithm implementation, fuzzy rule configuration, experimental workflow, and reproducibility documentation described in this study. Derived data supporting the results, including configuration files and algorithmic workflow descriptions, are also provided in the Appendix of this article. Additional experimental details or execution configurations may be obtained from the corresponding author upon reasonable request.

REFERENCES

[1] P. C. Pendharkar and P. Cusatis, "Trading financial indices with reinforcement learning agents," *Expert Systems with Applications*, vol. 103, pp. 1–13, 2018, doi: 10.1016/j.eswa.2018.02.032.

[2] M. E. Aloud and N. Alkhamees, "Intelligent algorithmic trading strategy using reinforcement learning and directional change," *IEEE Access*, vol. 9, pp. 114659–114671, 2021, doi: 10.1109/ACCESS.2021.3105259.

[3] Di. Fengqian and L. Chao, "An adaptive financial trading system using deep reinforcement learning with candlestick decomposing features," *IEEE Access*, vol. 8, pp. 63666–63678, 2020, doi: 10.1109/ACCESS.2020.2982662.

[4] F. Li, Z. Wang, and P. Zhou, "Ensemble investment strategies based on reinforcement learning," *Scientific Programming*, vol. 2022, 2022, doi: 10.1155/2022/7648810.

[5] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading," *Expert Systems with Applications*, vol. 173, Jul. 2021, doi: 10.1016/j.eswa.2021.114632.

[6] L. Tabaro *et al.*, "Algorithmic trading using double deep q-networks and sentiment analysis," *Information (Switzerland)*, vol. 15, no. 8, 2024, doi: 10.3390/info15080473.

[7] M. S. Islam, E. Hossain, A. Rahman, M. S. Hossain, and K. Andersson, "A review on recent advancements in FOREX currency prediction," *Algorithms*, vol. 13, no. 8, 2020, doi: 10.3390/A13080186.

[8] R. Bisoi and P. K. Dash, "A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter," *Applied Soft Computing Journal*, vol. 19, pp. 41–56, 2014, doi: 10.1016/j.asoc.2014.01.039.

[9] D. Kuruppuarachchi, H. Lin, and I. M. Premachandra, "Testing commodity futures market efficiency under time-varying risk premiums and heteroscedastic prices," *Economic Modelling*, vol. 77, pp. 92–112, 2019, doi: 10.1016/j.econmod.2017.12.005.

[10] C. Li and C. H. Tu, "Complex neural fuzzy system and its application on multi-class prediction — A novel approach using complex fuzzy sets, IIM and multi-swarm learning," *Applied Soft Computing Journal*, vol. 84, 2019, doi: 10.1016/j.asoc.2019.105735.

[11] S. Rao and T. Zhang, "Global asymptotic stability and asymptotically periodic oscillation in fractional-order fuzzy Cohen-Grossberg neural networks with delays," *Engineering Letters*, vol. 32, no. 1, pp. 12–20, 2024.

[12] P. Cardaliaguet and C. A. Lehalle, "Mean field game of controls and an application to trade crowding," *Mathematics and Financial Economics*, vol. 12, no. 3, pp. 335–363, 2018, doi: 10.1007/s11579-017-0206-z.

[13] P. Casgrain and S. Jaimungal, "Mean-field games with differing beliefs for algorithmic trading," *Mathematical Finance*, vol. 30, no. 3, pp. 995–1034, 2020, doi: 10.1111/mafi.12237.

[14] G. Mitrache, "A game theory approach to optimizing the banking and financial resolution framework," *Review of Financial Studies*, vol. 3, no. 4, pp. 66–83, 2018.

[15] K. Janda and P. Marek, "Financial sustainability-game theory analysis of options approach for a Czech Bank," *EconStor Preprints, ZBW - Leibniz Information Centre for Economics*, no. 249677, 2022.

[16] J. C. Freitas, A. A. Pinto, and Ó. Felgueiras, "Game theory for predicting stocks' closing prices," *Mathematics*, vol. 12, no. 17, 2024, doi: 10.3390/math12172676.

[17] D. Evangelista, Y. Saporito, and Y. Thamsten, "Price formation in financial markets: a game-theoretic perspective," *arXiv preprint*, 2022.

[18] M. Ullah, G. R. Lakhan, A. Channa, and S. Gul, "Game theory and stock investment," *Multicultural Education*, vol. 7, no. 6, pp. 40–54, 2021, doi: 10.5281/zenodo.4900016.

[19] J. Fat, P. Moengin, P. Astuti, and S. Cahyati, "Sustainability in forex trading: a review in search of the SARSA-FIS hybrid method as a novelty," in *Buletin Ekonomi Moneter dan Perbankan/Monetary and banking economics bulletin*, 2025, vol. 28, pp. 71–93, doi: 10.59091/2460-9196.2369.

[20] S. Mahdavi, S. Rahnamayan, and K. Deb, "Opposition based learning: A literature review," *Swarm and Evolutionary Computation*, vol. 39, pp. 1–23, 2018, doi: 10.1016/j.swevo.2017.09.010.

[21] P. Grądzki and P. Wójcik, "Is attention all you need for intraday Forex trading?," *Expert Systems*, vol. 41, no. 2, 2024, doi: 10.1111/exsy.13317.

[22] Y. Qiu, Y. Qiu, Y. Yuan, Z. Chen, and R. Lee, "QF-TraderNet: Intraday trading via deep reinforcement with quantum price levels based profit-and-loss control," *Frontiers in Artificial Intelligence*, vol. 4, 2021, doi: 10.3389/frai.2021.749878.

[23] D. Yang, Y. Yang, J. Luo, Z. Wang, and H. Sha, "Research on the impact of algorithmic trading on market volatility," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-15020-w.

[24] D. Peters, K. Vold, D. Robinson, and R. A. Calvo, "Responsible AI—Two frameworks for ethical design practice," *IEEE Transactions on Technology and Society*, vol. 1, no. 1, pp. 34–47, 2020, doi: 10.1109/TTS.2020.2974991.

[25] S. Küfeoğlu, "SDG-8: Decent work and economic growth," in *Sustainable Development Goals*, 2022, pp. 331–348.

[26] R. Verdecchia, J. Sallou, and L. Cruz, "A systematic review of Green AI," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 4, 2023, doi: 10.1002/widm.1507.

[27] A. Dixit and R. Chaudhary, "Sustainability: An essential concept for current and upcoming technological practices," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 804, no. 1, doi: 10.1088/1757-899X/804/1/012023.

[28] A. K. M. A. Ullah, S. Sultana, F. Faisal, M. M. I. Rahi, M. A. Alam, and M. G. R. Alam, "A brief review of responsible AI and socially responsible investment in financial and stock trading," *TechRxiv*, 2021, doi: 10.36227/techrxiv.15093762.v1.

BIOGRAPHIES OF AUTHORS

Joni Fat received the S.T. (*Sarjana Teknik*) degree in electrical engineering from Universitas Tarumanagara, Jakarta, Indonesia, in 2002, the M.E. (*Magister Ekonomi*) degree from Universitas Trisakti in 2009, and the M.T. (*Magister Teknik Elektro*) degree from Universitas Trisakti in 2019. He is currently a lecturer and the Secretary of the Electrical Engineering Study Program at Universitas Tarumanagara. He is also a doctoral candidate in the Industrial Engineering Program at Universitas Trisakti. He has received several research grants from Universitas Tarumanagara and national funding agencies, including the Directorate General of Higher Education (DIKTI). His research interests include artificial intelligence, algorithmic trading systems, responsible AI, sustainability in financial systems, game theory, Internet of Things (IoT), microcontroller and microprocessor systems, and antenna design. In this study, he contributed to conceptualization, system development, software implementation, experimental evaluation, and manuscript preparation. He can be contacted at email: jonif@ft.untar.ac.id.

Parwadi Moengin received the Drs. degree in mathematics from Universitas Negeri Semarang, Indonesia, in 1991, the M.Si. degree in applied mathematics (specialized in mathematical modeling) from Institut Teknologi Bandung in 1993, and the Ph.D. degree in operations research from Universiti Putra Malaysia in 2003. He is currently a professor in operations research and the Head of the Doctoral Program in industrial engineering at Universitas Trisakti, Jakarta, Indonesia. His research focuses on mathematical modeling, operations research, optimization, and sustainable system design. He has supervised numerous graduate and doctoral students and has contributed extensively to research in systems modeling and industrial optimization. In this study, he contributed to the conceptualization of the research framework, methodological design, and critical review of the manuscript. He can be contacted at email: parwadi@trisakti.ac.id.

Pudji Astuti received the Dra. degree in mathematics from Universitas Gadjah Mada, Yogyakarta, Indonesia, the M.T. degree in industrial engineering and management from Universitas Indonesia, and the Ph.D. degree in agro-industrial technology from Institut Pertanian Bogor. She is currently a senior lecturer in the Department of Industrial Engineering at Universitas Trisakti. Her research interests include artificial intelligence applications, algorithmic trading systems, responsible AI, sustainability in financial systems, and computational decision-making. She has participated in multiple research projects funded by national and institutional grants and has contributed to interdisciplinary research involving artificial intelligence and sustainable system design. In this study, she contributed to conceptual development, methodological supervision, validation of the research framework, and manuscript preparation. She can be contacted at email: pudji@trisakti.ac.id.

Sally Cahyati received the S.T. degree in mechanical engineering from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, the M.T. degree in mechanical engineering from Universitas Indonesia, and the doctoral degree in agricultural engineering from Institut Pertanian Bogor (IPB). She is currently a senior lecturer in the Department of Mechanical Engineering at Universitas Trisakti. She also serves as the Head of the CADCAM Studio and the Head of the Waste to Value Trisakti Research Center. She has received numerous research grants, including Hibah PTM BIMA Kemensaintek DIKTI, Hibah Matching Fund DIKTI, Hibah BRIN, and Hibah Dana Padanan programs. Her research interests include artificial intelligence applications, sustainability engineering, smart manufacturing systems, digital-based design and manufacturing, and AI–IoT integration. In this study, she contributed to system validation, implementation evaluation, and manuscript review. She can be contacted at email: sally@trisakti.ac.id or sallycahyati@trisakti.ac.id.