An AI-Agent-based, Neural Network system for automated polygraph scoring: Architecture, Design and Experimental Validation

Dana Rad [1,2,*], Nicolae Paraschiv [1], Csaba Kiss [3], Valentina Emilia Balas [4,5] and Gavril Rad [2,*]

[1] Doctoral School of Systems Engineering, Petroleum-Gas University of Ploiești, 100680 Ploiești, Romania, e-mail: dana@xhouse.ro, nparaschiv@upg-ploiesti.ro, balas@drbalas.ro

[2] Center of Research Development and Innovation in Psychology, Faculty of Educational Sciences Psychology and Social Work, Aurel Vlaicu University of Arad, 310032 Arad, Romania, e-mail: dana@xhouse.ro, radgavrilarad@gmail.com

[3] Faculty of Psychology and Educational Sciences, Hyperion University of Bucharest, 030615 Bucharest, Romania, e-mail: kiss.csaba@expertpoligraf.ro

[4] Faculty of Engineering, Aurel Vlaicu University of Arad, 310032 Arad, Romania, e-mail: balas@drbalas.ro

[5] The Academy of Romanian Scientists, Str. Ilfov 3, Bucharest, Romania, e-mail: balas@drbalas.ro

Abstract: *The automation of polygraph scoring represents a complex challenge at the intersection of psychology, physiological signal processing, expert-based reasoning and artificial intelligence. Traditional scoring approaches rely heavily on human expertise, introducing subjectivity and variability, while purely data-driven models often lack interpretability and domain alignment. This paper proposes an automated polygraph scoring system based on a hybrid architecture that integrates artificial neural networks with an AI agent responsible for orchestrating the processing pipeline. The system is designed using a systemic input–state–output framework, where multi-channel physiological signals (electrodermal activity, cardiovascular signals and respiration) are transformed into interpretable scores and decision outputs. The system was experimentally validated on a controlled dataset (N=20) using the Utah protocol. Performance evaluation included accuracy assessment, error analysis, and examination of misclassification cases, demonstrating the operational feasibility and consistency of the proposed approach.*

Keywords: AI Agent; Polygraph Scoring; Neural Networks; Physiological Signal Processing; Explainable AI; NeuroPolygraph

1 Introduction

Polygraph examination represents a complex interdisciplinary domain that integrates psychophysiology, signal processing, and decision systems. The fundamental premise of polygraph testing is that deceptive behavior induces measurable physiological responses, particularly within electrodermal, cardiovascular, and respiratory systems [1][2]. These responses are associated with autonomic nervous system activation and have been extensively studied in the context of deception detection and orienting response mechanisms [3][4].

Despite its widespread application, polygraph-based lie detection remains a debated topic, primarily due to variability in accuracy, dependence on protocol design, and challenges related to interpretation [5][6]. Meta-analytical evidence suggests that the validity of techniques such as the Comparison Question Test (CQT) depends significantly on methodological rigor and contextual factors [7][8]. Additionally, ethical and legal considerations continue to influence the acceptance and implementation of polygraph techniques across different domains [9][10].

From a technical perspective, polygraph systems rely on multi-channel acquisition of physiological signals, including electrodermal activity, heart rate, blood pressure, and respiratory patterns [2-4]. These signals are inherently noisy and affected by artifacts, inter-individual variability, and contextual influences, which complicates the extraction of reliable indicators of deception [6]. Consequently, the transformation of raw physiological data into decision-relevant information requires structured signal processing and robust analytical frameworks [11].

Recent advances in polygraph research have emphasized the need for systematic and reproducible scoring methodologies. Traditional approaches based on manual expert evaluation are increasingly complemented or replaced by computational models capable of processing complex physiological patterns [11]. In this context, the integration of machine learning techniques, particularly artificial neural networks, has shown promise in capturing nonlinear relationships between physiological features and deceptive behavior [12][13].

However, the application of data-driven models introduces new challenges related to interpretability, robustness, and system integration. Purely statistical or neural approaches may lack transparency and fail to incorporate domain-specific expertise, which remains critical in forensic applications [11]. This has led to the emergence of hybrid approaches that combine expert knowledge with computational intelligence.

From a systems engineering standpoint, polygraph scoring can be conceptualized as a structured pipeline in which physiological signals are progressively transformed into decision outputs. Similar to pipeline systems in engineering domains, such architectures require careful design to ensure stability, reliability, and efficient flow of information [14][15]. This perspective enables the modeling of polygraph scoring as a dynamic system with defined inputs, internal states, and outputs.

In parallel, the development of intelligent systems has introduced agent-based architectures capable of managing complex workflows and integrating multiple analytical components. Multi-agent systems and agentic AI frameworks provide mechanisms for distributed decision-making, process orchestration, and adaptive system behavior [16][17]. These architectures are particularly relevant in contexts where heterogeneous data sources and processing modules must be coordinated within a unified system.

A critical dimension of modern AI systems is the incorporation of human-in-the-loop paradigms, which ensure that human expertise remains integrated into automated decision processes. Human-in-the-loop learning enhances system safety, interpretability, and ethical compliance, particularly in high-stakes applications such as forensic evaluation [18][19]. This approach allows for a balance between automation and expert oversight, reducing the risks associated with fully autonomous decision systems.

Furthermore, recent developments in explainable artificial intelligence (XAI) highlight the importance of transparency and interpretability in AI-assisted decision-making. In domains involving sensitive judgments, such as deception detection, the ability to explain system outputs is essential for trust, validation, and practical adoption [20].

Building on these theoretical and technological advancements, there is a growing need for integrated architectures that combine physiological signal processing, expert knowledge, artificial neural networks, and intelligent orchestration mechanisms. Such systems must not only achieve high predictive performance but also ensure reproducibility, traceability, and interpretability.

In this context, the present study proposes an AI-agent-based neural system for automated polygraph scoring, designed as a modular and system-oriented architecture. By integrating multi-channel physiological data processing, neural modeling, and agent-based orchestration within a human-in-the-loop framework, the proposed approach aims to advance the reliability, transparency and operational feasibility of polygraph scoring systems.

2 Methodology

2.1 Objectives

This research aims to design and validate an AI-agent-based neural system for automated polygraph scoring, conceptualized as a structured and modular information processing architecture. The study seeks to transform multi-channel physiological signals into interpretable and reproducible decision outputs by integrating signal processing, expert knowledge formalization, and artificial intelligence techniques.

The first objective is to develop a systemic architecture based on an input–processing–output framework, in which physiological signals (electrodermal activity, cardiovascular responses, and respiration) are acquired, preprocessed, and transformed into structured representations suitable for automated scoring. This includes the implementation of standardized preprocessing procedures such as filtering, baseline correction, normalization, and artifact detection, ensuring signal stability and comparability.

The second objective is to formalize expert polygraph knowledge into a structured scoring subsystem (SKILL), capable of generating intermediate, explainable scores based on predefined rules. This component aims to preserve domain-specific expertise while enabling transparency and traceability in the decision-making process.

The third objective is to integrate an artificial neural network model within the scoring pipeline, in order to capture nonlinear relationships between physiological features and deception-related responses. The model is designed to operate under controlled conditions, with constraints aimed at preventing overfitting and ensuring generalization.

The fourth objective is to implement an AI agent responsible for orchestrating the processing workflow, managing data validity conditions, coordinating subsystem interactions, and ensuring full traceability of the scoring process. The agent-based approach aims to enhance system robustness, adaptability, and operational control.

The fifth objective is to develop a decision fusion mechanism that combines expert-based scores and neural model outputs into a unified decision, accompanied by a confidence measure and the possibility of defining uncertainty regions (gray zones).

Finally, the research aims to experimentally validate the proposed system on a controlled dataset using a standardized polygraph protocol. The evaluation focuses on performance metrics, error analysis, and system stability, with the goal of assessing the feasibility, reliability, and practical applicability of the proposed architecture.

These objectives contribute to advancing automated polygraph scoring by shifting the focus from isolated algorithmic solutions to integrated system design, emphasizing reproducibility, interpretability, and robustness in real-world conditions.

2.2 Data collection and participants

The data used in this study were collected within a controlled experimental framework designed to ensure standardization, reproducibility, and consistency of physiological recordings during polygraph examinations. The dataset consists of a sample of N=20 participants, selected to allow the preliminary validation of the proposed automated scoring system under controlled conditions.

All participants underwent polygraph testing using a standardized protocol, specifically the Utah protocol, which is widely recognized for its structured sequence of relevant, comparison, and neutral questions. This protocol ensures temporal alignment between stimulus presentation and physiological responses, allowing for consistent segmentation and analysis of post-item reactions.

Physiological data were acquired using a multi-channel polygraph system capable of recording the primary indicators of autonomic activity. The recorded signals included electrodermal activity (EDA), cardiovascular signals, and respiratory activity, all synchronized with event markers corresponding to the timing of each question. Data acquisition was performed using a Lafayette LX6 polygraph system, ensuring high-resolution multi-channel recording and standardized signal capture.

For the experimental design, the 20 participants were divided into five groups of four individuals, each group being exposed to a controlled scenario involving a simulated theft task. In each group, two banknotes of 50 RON were introduced, and the ground truth condition was predefined such that either one participant stole both banknotes or two participants each stole one banknote, while the remaining participants were instructed to respond truthfully. This design ensured a clear and controlled ground truth for deception versus non-deception conditions.

Data collection was conducted under controlled environmental conditions to minimize external sources of variability, such as temperature fluctuations, movement artifacts, or sensor instability. Participants were instructed to maintain a stable posture and to avoid voluntary movements or behaviors that could interfere with signal quality.

Prior to analysis, the recorded signals underwent a validation process to ensure data integrity. Segments affected by significant artifacts—such as abrupt movements, signal loss, or irregular physiological patterns—were identified and either excluded or flagged for reduced reliability within the scoring process.

The relatively small sample size reflects the exploratory and validation-oriented nature of the study, focusing on the feasibility and functional performance of the proposed system rather than on large-scale generalization. However, the controlled design and the explicit definition of ground truth provide a reliable foundation for evaluating the system's behavior and decision accuracy.

2.3 AI-Agent-orchestrated processing pipeline

The proposed system is implemented as an AI-agent-orchestrated processing pipeline, designed to transform raw multi-channel physiological signals into structured scoring outputs and final decision variables. The architecture follows a modular and hierarchical design, integrating signal processing, expert knowledge formalization, and neural inference within a unified computational framework.

From a systems engineering perspective, the entire scoring process is conceptualized as a sequential transformation in which physiological signals are progressively structured, reduced, and mapped into a decision space. This transformation can be formally expressed as:

$$\mathbf{y = D \circ M \circ E \circ P(u(t), w(t))}$$

where $u(t)$ represents the multi-channel physiological signals acquired during the polygraph examination, $w(t)$ denotes perturbations such as noise and artifacts, and y corresponds to the final decision output. The operators P, E, M, and D define the successive processing stages of the system.

The first stage, corresponding to the preprocessing operator P, ensures the stabilization and conditioning of raw signals. Channel-specific filtering is applied to electrodermal, cardiovascular, and respiratory signals, followed by baseline correction and detrending procedures. In addition, intra-subject normalization is implemented to reduce inter-individual variability, while artifact detection mechanisms identify and flag segments affected by movement, signal loss, or irregular physiological patterns. The output of this stage is a conditioned signal $x(t)$, aligned with event markers corresponding to each stimulus.

The second stage, defined by the feature extraction operator EEE, transforms the conditioned signal into a finite-dimensional representation:

$$z=[z1,z2,...,zn]$$

This transformation involves the computation of features that characterize the physiological response, including amplitude, latency, duration, area under the curve, and variability indices. Additionally, inter-channel synchronization patterns are considered to capture the coordinated dynamics of autonomic responses. The role of this stage is to reduce the dimensionality of the signal while preserving the information relevant for classification and scoring.

The third stage, represented by the modeling operator MMM, integrates both expert-based and data-driven components. A rule-based subsystem (SKILL) formalizes polygraph expertise into explicit scoring rules, generating interpretable intermediate outputs. In parallel, a neural network model is employed to capture nonlinear relationships between the extracted features and deception-related responses. The neural inference is defined as:

$$s=f_\theta(z)$$

where θ denotes the learned parameters of the model. This dual structure allows the system to combine interpretability with predictive capability, ensuring both transparency and performance.

The final stage, corresponding to the decision operator *D*, converts the inferred score into a decision output. This process includes threshold-based classification, probabilistic interpretation, and the definition of an uncertainty region ("gray zone") for borderline cases. The output can be expressed as:

$$y=(class,p,u_c)$$

where *class* represents the decision label, p is the associated probability, and u_c denotes the uncertainty measure.

A central component of the architecture is the AI agent, which orchestrates the entire processing pipeline. The agent is responsible for managing the execution flow, validating input data, monitoring intermediate outputs, and ensuring the traceability of all processing steps. It coordinates the interaction between the preprocessing, feature extraction, modeling, and decision modules, enabling consistent system behavior under varying conditions.

From an implementation perspective, the system was developed as an integrated application in which the AI agent operates as a supervisory layer over the computational pipeline. The interface and orchestration logic were constructed using Gemini Pro in conjunction with the Antigravity framework [21], enabling structured interaction between components, dynamic control of processing stages, and real-time monitoring of system states. This implementation ensures modularity, reproducibility, and extensibility of the system, allowing future integration of additional models or processing modules.

3 Results

The results of the present study are structured to illustrate both the functional behavior of the proposed system and its ability to transform physiological data into interpretable decision outputs. The evaluation focuses on the operational performance of the AI-agent-based pipeline, as well as on the interpretability and traceability of the scoring process.

As an initial step, it is important to present the interface through which the system integrates signal processing, neural inference, and decision outputs. The developed application, named NeuroPolygraph, provides a unified environment for data input, signal visualization, statistical analysis, and final decision generation, presented in Figure 1.

Figure 1. *Example of the NeuroPolygraph application interface used for the automated analysis of polygraph charts and visualization of the AI-based scoring system results.*

Figure 1 illustrates the main components of the NeuroPolygraph interface, designed to support the complete processing pipeline from raw data input to final decision output. The interface integrates multiple functional modules within a single environment, allowing both automated processing and human supervision.

The upper section of the interface displays the multi-channel physiological signals, including electrodermal activity, cardiovascular responses, and respiratory patterns, synchronized with event markers corresponding to question onset. This visualization enables the inspection of raw and processed signals, ensuring transparency in the signal transformation process.

In the left panel, the system allows the import of session data, including polygraph chart images or structured data formats, which are subsequently processed through the automated pipeline. This module ensures flexibility in data acquisition and compatibility with different recording formats.

The central component of the interface is the statistical reaction report, where physiological responses are quantified and organized per item. Parameters such as electrodermal amplitude, cardiovascular variation, and respiratory indices are displayed alongside item type (relevant, control), providing a structured representation of physiological reactivity.

The system also includes a neural trace analysis module, which visualizes the contribution of different features within the neural network model. This component enhances interpretability by illustrating how input variables are propagated through the model and how they influence the final prediction.

A key output of the system is the probability of deception, expressed as a continuous value (e.g., 23.98%), accompanied by a categorical interpretation (e.g., truthful). This probabilistic output reflects the integration of expert-based scoring and neural inference within the decision fusion module.

Finally, the lower section presents the system engineering schema, highlighting the pipeline structure and the integration between rule-based logic and neural modeling. This visualization reinforces the modular and systemic nature of the proposed architecture.

The application of the NeuroPolygraph system on the experimental dataset consisting of 20 participants revealed a high level of classification performance. The evaluation of system accuracy was conducted using the ground truth principle, defined based on the actual behavior of each participant within the experimental scenario. Following the polygraph examination, all participants explicitly confirmed their role in the experiment, namely whether they had performed the simulated theft or not. This post-experimental validation ensured an unambiguous ground truth and allowed for a direct comparison between real behavior and the outputs generated by the automated scoring system.

The obtained results indicate that 19 out of the 20 participants were correctly classified, corresponding to an overall classification accuracy of 95%. The system successfully identified both categories of subjects: those who did not engage in the simulated theft (classified as truthful) and those who performed the action (classified as deceptive).

The only misclassification occurred in a specific experimental context, within a group in which two participants were designated as deceptive, each stealing one banknote. In this case, the system correctly identified one of the deceptive participants, while the second was classified as truthful, with a probability value located very close to the decision threshold.

This type of result suggests the presence of a decision uncertainty zone, a phenomenon commonly observed in classification systems based on physiological signals, particularly when the generated responses are of reduced intensity or exhibit high variability. Table 1 presents the detailed classification results generated by the NeuroPolygraph system for each participant.

Table 1. Classification results generated by the NeuroPolygraph system

Subject	*Group*	*Ground Truth*	*System Classification*	*Probability of Deception (%)*	*Result*
S1	G1	Truthful	Truthful	21.4	Correct
S2	G1	Deceptive	Deceptive	71.8	Correct
S3	G1	Truthful	Truthful	24.7	Correct
S4	G1	Deceptive	Deceptive	76.2	Correct
S5	G2	Truthful	Truthful	18.9	Correct
S6	G2	Deceptive	Deceptive	68.5	Correct
S7	G2	Truthful	Truthful	27.1	Correct
S8	G2	Deceptive	Deceptive	74.3	Correct
S9	G3	Truthful	Truthful	19.6	Correct
S10	G3	Deceptive	Deceptive	70.4	Correct
S11	G3	Truthful	Truthful	25.2	Correct
S12	G3	Deceptive	Deceptive	73.7	Correct
S13	G4	Truthful	Truthful	22.6	Correct
S14	G4	Deceptive	Deceptive	69.1	Correct
S15	G4	Truthful	Truthful	20.3	Correct
S16	G4	Deceptive	Deceptive	75.5	Correct
S17	G5	Truthful	Truthful	23.9	Correct
S18	G5	Deceptive	Deceptive	72.6	Correct
S19	G5	Deceptive	Truthful	49.8	Error
S20	G5	Truthful	Truthful	26.4	Correct

The results demonstrate that the NeuroPolygraph system exhibits a strong capability to identify deceptive behavior under controlled experimental conditions. The high classification rate confirms that the integration of expert-based scoring and neural inference represents a promising approach for automated polygraph data analysis.

Within the experimental validation, only a single participant was misclassified by the system. This case provides valuable insight into the current limitations of the model and the behavior of the system near decision boundaries. The misclassified subject belonged to group G5, where both designated participants engaged in the simulated theft, each stealing one banknote. According to the ground truth, this participant should have been classified as deceptive. However, the system assigned a probability value close to the decision threshold, resulting in a final classification of truthful.

A detailed analysis of the physiological signals recorded for this subject indicates that the responses to relevant questions were comparable in amplitude to those associated with control questions. Under such conditions, the difference in the Reaction Strength (RS) indicator between relevant and control items becomes minimal, leading the neural model to estimate an intermediate probability between the two decision classes. Another contributing factor may be inter-individual variability in physiological reactivity. In polygraph examinations, some individuals exhibit attenuated physiological responses even when engaging in deceptive behavior. This phenomenon is well documented and may be associated with differences in autonomic nervous system reactivity or with cognitive strategies aimed at regulating emotional responses.

It is also important to note that this case falls within the uncertainty region of the model, where the estimated probability is close to the classification threshold. In practical polygraph applications, such cases are often interpreted as inconclusive and require additional expert analysis. Therefore, the observed misclassification does not necessarily indicate a fundamental limitation of the model, but rather highlights the existence of a decision uncertainty zone, inherent to systems based on psychophysiological signal analysis. The integration of a human-in-the-loop mechanism allows expert intervention in such cases, supporting contextual interpretation and reducing the risk of incorrect conclusions.

To evaluate the performance of the NeuroPolygraph system, standard classification metrics were computed based on the comparison between system outputs and ground truth labels.

The experimental dataset included 20 participants, equally distributed between truthful and deceptive conditions. The system correctly classified 19 cases and produced a single misclassification. The main performance indicators are summarized in Table 2.

Table 2. Performance metrics of the NeuroPolygraph system

Indicator	***Value***
Total number of subjects	20
Correct classifications	19
Misclassifications	1
Accuracy	95%

Sensitivity (deception detection)	90%
Specificity (truth detection)	100%

The overall accuracy of 95% indicates a high level of system performance under the experimental conditions. The system successfully identified all truthful subjects and nearly all deceptive subjects.

The single classification error occurred in a case where the estimated probability was very close to the decision threshold, further confirming the presence of a decision uncertainty zone. Such situations are characteristic of systems based on physiological signal analysis and are also encountered in expert-based polygraph examinations.

These results should be interpreted in the context of the exploratory nature of the experiment and the relatively small sample size. Nevertheless, the performance obtained suggests that the proposed NeuroPolygraph architecture, combining expert scoring with neural inference, represents a viable approach for automated polygraph analysis.

At the same time, the results reinforce the importance of integrating a human-in-the-loop mechanism, allowing expert intervention in uncertain cases and contributing to the overall reliability of the decision-making process.

4 Discussion

The results obtained in this study demonstrate that the proposed NeuroPolygraph system achieves a high level of classification performance under controlled experimental conditions, with an overall accuracy of 95%. This level of performance supports the viability of integrating expert-based scoring, neural inference, and AI-agent orchestration into a unified system architecture.

From a systems engineering perspective, the proposed pipeline can be interpreted as a controlled flow system, in which physiological signals are progressively transformed into decision outputs. Similar principles are observed in complex engineered systems such as pipeline control and flow modeling architectures, where system stability depends on the integrity of each transformation stage and on the coordination between subsystems. In this context, the performance of the NeuroPolygraph system is not solely determined by the neural model, but by the robustness of the entire pipeline, including preprocessing, feature extraction and the decision fusion.

The integration of an AI agent as an orchestration mechanism represents a key contribution of this study. Unlike traditional monolithic models, the proposed system adopts a modular and agent-based architecture, enabling dynamic coordination between processing components. This approach is consistent with the evolution of multi-agent autonomic systems, where agents manage distributed processes and ensure system-level adaptability and control [14]. More recent developments in agentic AI architectures further emphasize the role of agents in structuring workflows and enabling scalable, adaptive intelligent systems [15].

An important aspect highlighted by the results is the presence of a decision uncertainty zone, reflected in the single misclassified case. This phenomenon aligns with broader findings in AI-assisted decision-making, where classification boundaries are often probabilistic rather than deterministic. In such contexts, system performance must be evaluated not only in terms of accuracy, but also in terms of how uncertainty is managed and communicated [22].

The incorporation of a human-in-the-loop (HITL) mechanism addresses this limitation by allowing expert intervention in borderline cases. This hybrid approach reflects current trends in intelligent system design, where full automation is complemented by human oversight to ensure safety, interpretability, and ethical compliance [23-25]. The results of the present study confirm that such integration is particularly relevant in domains involving physiological signals and high variability, where purely automated decisions may be insufficient in certain conditions.

Another important implication of this work relates to the balance between interpretability and predictive performance. While neural networks provide the capacity to model complex nonlinear relationships, their outputs must be contextualized within a transparent decision framework. The use of a rule-based subsystem (SKILL), combined with probabilistic outputs and visualization modules, contributes to the justification of the system, which is essential for practical adoption.

Furthermore, the experimental results highlight the role of individual variability in physiological responses, which remains a fundamental challenge in polygraph analysis. Even in controlled conditions, subjects may exhibit attenuated or atypical responses, leading to intermediate probability estimates. From an application perspective, the NeuroPolygraph system demonstrates potential for use in controlled forensic and investigative environments. However, the relatively small sample size and the experimental nature of the setup limit the generalizability of the results. Future research should focus on validating the system on larger and more diverse datasets, as well as on exploring adaptive learning mechanisms capable of personalizing the model to individual response patterns.

In addition, further work is needed to refine the decision thresholds and to formalize the handling of uncertainty zones, potentially through confidence calibration or multi-stage decision strategies. The integration of additional physiological or behavioral signals may also enhance system performance and robustness.

Overall, the findings of this study support the conclusion that AI-agent-based architectures represent a promising direction for automated polygraph scoring, enabling a shift from isolated algorithmic solutions toward integrated, system-level approaches. By combining signal processing, expert knowledge, neural modeling, and intelligent orchestration, the proposed framework contributes to the development of more reliable, transparent, and adaptable decision systems.

5 Conclusions

The present study introduced and validated an AI-agent-based, neural system for automated polygraph scoring, designed as a modular and system-oriented architecture integrating physiological signal processing, expert knowledge formalization, and machine learning techniques. The results obtained in a controlled experimental setting, demonstrate that such an approach is both feasible and effective, achieving a classification accuracy of 95% and successfully identifying the majority of deceptive and truthful behaviors.

From an engineering perspective, one of the main contributions of this work lies in the reframing of polygraph scoring as a structured processing pipeline, rather than a standalone analytical task. By explicitly modeling the transformation from raw signals to decision outputs through well-defined stages—preprocessing, feature extraction, modeling, and decision fusion—the system ensures reproducibility, traceability, and modularity. This design allows each component to be independently analyzed, optimized, and extended, which is essential for building robust real-world systems.

A second major contribution is the integration of an AI agent as a supervisory orchestration layer, which coordinates the entire processing workflow. This element shifts the paradigm from static models to adaptive and controlled systems, where data validation, process monitoring, and execution logic are explicitly managed. In practice, this enables better handling of variability in input data and ensures consistent system behavior, under different operational conditions.

The results also highlight the importance of combining expert-based scoring with neural inference. While neural networks provide the ability to capture complex nonlinear relationships in physiological data, the inclusion of a rule-based subsystem ensures interpretability and alignment with domain knowledge. This hybrid approach proves to be particularly effective in maintaining both predictive performance and transparency, which are critical requirements in forensic and decision-support applications.

An important finding of this study is the identification of a decision uncertainty zone, illustrated by the single misclassified case. Rather than representing a failure of the system, this result reflects a fundamental characteristic of classification processes based on physiological signals, where responses may be ambiguous or weakly differentiated. The presence of such a zone reinforces the necessity of incorporating human-in-the-loop mechanisms, allowing expert intervention in borderline cases and ensuring that final decisions remain contextually grounded.

From a practical standpoint, the NeuroPolygraph system demonstrates strong potential for application in controlled investigative environments, training scenarios, and decision-support systems. The integrated interface, probabilistic outputs, and visualization capabilities facilitate both automated processing and expert interpretation, bridging the gap between algorithmic analysis and operational use.

However, several limitations must be acknowledged. The experimental validation was conducted on a relatively small sample size, and the controlled conditions may not fully capture the variability encountered in real-world settings.

Future research should focus on scaling the system to larger and more diverse datasets, as well as on exploring adaptive learning strategies capable of accounting for individual differences in physiological reactivity. Further development directions include the refinement of decision thresholds, the formalization of uncertainty handling mechanisms, and the integration of additional data modalities. In addition, the evolution of the AI agent toward more autonomous and context-aware behavior may further enhance system performance and adaptability.

Thus, this study supports the view that automated polygraph scoring should be approached as a system engineering problem, rather than solely as a classification task. The proposed architecture demonstrates that the integration of AI, expert knowledge and structured processing pipelines, can lead to more reliable, interpretable and scalable solutions. Such approaches are likely to play a key role in the future development of intelligent decision systems, operating at the intersection of technology and human behavior.

References

[1]. Ayoub, A., Rizvi, F., Akram, S., & Tahir, M. A. (2018). The polygraph and lie detection: A case study. Arab Journal of Forensic Sciences & Forensic Medicine, 1(7), 902–908.

[2]. Winkler-Galicki, J., Bartkowiak-Wieczorek, J., Synowiec, D., Dąbrowska, R., & Mądry, E. (2022). Polygraph analyses: Technical and practical background. Journal of Medical Science, 91(1), e590.

[3]. Kleiner, M. (2018). The Psychophysiology of deception and the orienting response. In Interviewing and Deception (pp. 183-208). Routledge.

[4]. Krapohl, D. J., Halford, K., Benson, T., Mayston, A., & Dutton, D. W. (2021). Response Onset Latencies of Electrodermal, Cardiovascular and Vasomotor Responses During Field Polygraph Testing. Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice, 50(1), 15-27.

[5]. Kiss, C. (2015). Differences in psychophysiological reactivity related to the seriousness of offence and the relapse/revealed by the polygraph technique-ethical issues. Revista Romana de Bioetica, 13(2), 321–330.

[6]. Mao, Y. Y., Liang, Y. D., & Hu, Z. Q. (2015). Accuracy rate of lie-detection in China: Estimate the validity of CQT on field cases. Physiology & Behavior, 140, 104-110.

[7]. Honts, C. R., Thurber, S., & Handler, M. (2021). A comprehensive meta-analysis of the comparison question polygraph test. Applied cognitive psychology, 35(2), 411-427.

[8]. Meijer, E. H., & van Koppen, P. J. (2017). Lie detectors and the law: The use of the polygraph in Europe. In Psychology and law (pp. 45-64). Routledge.

[9]. Han, Y. (2016). Deception detection techniques using polygraph in trials: Current status and social scientific evidence. Contemporary Readings in Law and Social Justice, 8(2), 115-147.

[10]. Rad, D., Paraschiv, N., & Kiss, C. (2023). Neural network applications in polygraph scoring—A scoping review. Information, 14(10), 564.

[11]. Blikhar, V., Zaiats, O., Pavliuk, N., & Kalka, N. (2022). Psychological and legal aspects of verification and detection of lies during polygraph examination. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 13(1), 276-291.

[12]. Feng, J., Yu, T., Zhang, K., & Cheng, L. (2025). Integration of multi-agent systems and artificial intelligence in self-healing subway power supply systems: Advancements in fault diagnosis, isolation, and recovery. Processes, 13(4), 1144.

[13]. Giri, D. R., Kandula, C. S. R., Srikanth, M., Kotipalli, S. S., & Kumar, J. R. (2025). Hybrid System framework for AI pipeline and AI Agent. In Algorithms in Advanced Artificial Intelligence (pp. 738-743). CRC Press.

[14]. Bigus, J. P., Schlosnagle, D. A., Pilgrim, J. R., Mills, W. N., III, & Diao, Y. (2002). ABLE: A toolkit for building multiagent autonomic systems. IBM Systems Journal, 41(3), 350–371. https://doi.org/10.1147/sj.413.0350

[15]. Krishnan, N. (2025). Ai agents: Evolution, architecture, and real-world applications. arXiv preprint arXiv:2503.12687.

[16]. Jaggavarapu, M. K. R. (2025). The evolution of agentic AI: architecture and workflows for autonomous systems. Journal Of Multidisciplinary, 5(7), 418-427.

[17]. Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic ai: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. Future Internet, 17(9), 404.

[18]. Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S. K., & Sharma, R. (2024). Applications, challenges, and future directions of human-in-the-loop learning. IEEE Access, 12, 75735-75760.

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

[20]. Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., ... & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. ACM computing surveys, 55(9), 1-33.

[21]. dos Santos Filho, E. B. (2026). ESAA: Event Sourcing for Autonomous Agents in LLM-Based Software Engineering. arXiv preprint arXiv:2602.23193.

[22]. Steyvers, M., & Kumar, A. (2024). Three challenges for AI-assisted decision-making. Perspectives on Psychological Science, 19(5), 722-734.

[23]. Bensch, L., Bensch, O., Nilsson, T., Paradiso, J. A., Davidoff, S., Blanchard, L., ... & Flemisch, F. (2025). Human-in/on-the-Loop AI: Enabling Human Controllability and Decision-Making in Spaceflight. In Handbook of Human-Centered Artificial Intelligence (pp. 1-46). Singapore: Springer Nature Singapore.

[24]. Dobeš, M., & Sabolová, N. (2023). Emotion recognition using pretrained deep neural networks. *Acta Polytechnica Hungarica*, *20*(4), 195-204.

[25]. Rad, D., Paraschiv, N., Kiss, C., Balas, V. E., & Barna, C. (2024). Physiological Reactions Profiling in Polygraph Testing: Insights from Fuzzy C-Means Clustering Analysis. *Acta Polytechnica Hungarica*, *21*(10), 365-377.

NO PAGE NUMBERS